university of groningen

faculty of science and engineering

Iterative Subspace Correction Procedure for Generalized Matrix Learning Vector Quantization in Biomedicine

Kaitlin Vos

university of groningen / faculty of science and engineering

University of Groningen

Iterative Subspace Correction Procedure for Generalized Matrix Learning Vector Quantization in Biomedicine

Bachelor's Thesis

To fulfill the requirements for the degree of
Bachelor of Science in Computing Science
at University of Groningen under the supervision of
Prof. dr. M. Biehl (Intelligent Systems, University of Groningen)
and
S. Lövdal (Intelligent Systems, University of Groningen)

Kaitlin Vos (s4266811)

July 13, 2023

Contents

Abstract

Neurodegenerative diseases are challenging to diagnose accurately due to their incurable nature and overlapping symptoms. One potential solution is to use a trained Generalized Matrix Learning Vector Quantization (GMLVQ) system on patients' FDG-PET scans to differentiate between different neurodegenerative diseases.
In this thesis, we explore a modified method incorporating iterative subspace corrections. The GMLVQ system is trained on homogeneous healthy control cohorts, allowing extraction of vectors highlighting centre differences. By ignoring these in our feature space, we aim to eliminate the centre-specific variance present in the data. The method's performance is assessed by using artificial data sets of increasing complexity, generated as Gaussian clusters, and evaluating it based on the known characteristics used to generate them.
The results showed that the iterative method successfully aligned the subspace with the ground truth centre direction in simple cases, maintaining a high accuracy in disease classification and arbitrary classification in centre identification. In complex cases, the iterative method struggled to accurately capture the underlying differences. Larger data sets generally improved performance, but complexity posed challenges in classification accuracy. The centre-wise z-scored data approach demonstrated better adaptability, achieving high disease classification accuracy and arbitrary centre classification across different complexities.

1 Introduction

Neurodegenerative diseases are currently incurable and their decline can partially or completely cause a loss of function of the nerve cells in specific regions of the brain, which could eventually be fatal [1]. Due to the lack of an effective cure, an accurate diagnosis in the early stages is needed in order to attempt to alleviate the symptoms. Pathological anatomical changes of the brain are generally seen in the later stages of the development of these neurodegenerative diseases [2]. Thus, methodologies that are able to show functional changes at these early stages are necessary.

Recent publications study the machine learning-based analyses of fluorodeoxyglucose-positron emission tomography (FDG-PET) scan images. A Generalized Matrix Learning Vector Quantization (GMLVQ) system can be trained on the pre-processed FDG-PET data with the objective to be able to distinguish between different neurodegenerative diseases. Initially, several GMLVQ systems are trained and based on the average balanced accuracy [3, 4], the final GMLVQ system is chosen.
The current application has proven to be able to result in a reliable classification in centre-specific settings. However, due to the lack of large quantities of data, combining data from different neuroimaging centres is necessary.

However, cross-centre classification is difficult due to the potential centre-specific characteristics of the available FDG-PET data [2], which might lead to the GMLVQ system to confuse the differentiation between centres and the differentiation between neurodegenerative diseases.

A universal classifier that can differentiate between the different neurodegenerative diseases regardless of the origin of the data presented is the main objective. Simultaneously, we aim for the classification to rely on the subspace that is most relevant to the disease classes. This is necessary because incorporating centre-specific variations in the data may potentially lead to misleadingly improved performance and introduce biases when interpreting the model's outcomes. This would allow for the classifier to be used within a wide range of centres and produce a reliable classifications for the patient cohorts of the respective centres.

In order to build this classifier, we would like to remove the centre-specific variance in the data, so that the classification solely relies on the intrinsic variation found in the disease information. By training a GMLVQ to classify between healthy control cohorts, which are assumed to be homogeneous, from different centres, we plan to extract the vectors describing the main difference between the centres. These vectors are to be ignored in the feature space for the disease classification. This procedure is to be completed in an iterative manner as to allow for subsequent removal of these vectors.

In this thesis, we implement a method for removing the centre-specific bias from the FDG-PET data by means of an iterative subspace correction, and perform experiments on artificial data to examine the behaviour of this method. We aim to demonstrate the functioning of this method and seek to gain valuable insights that can contribute significantly to enhancing the analysis of real-world neuroimaging data.

1.1 Research Questions

To summarise, this thesis focuses on the following problems:

Q1. How does the iterative subspace correction perform on artificial data?

Q2. How much unwanted variation due to source-specific characteristics can be removed from the artificial data?

Q3. Will the iterative subspace correction remove too much information at some point?

1.2 Thesis Outline

The following is the outline for the next chapters of this thesis: A literature study and background information on the topic of this thesis can be found under Section 2. In Section 3, we introduce the iterative subspace correction procedure based on the Iterated Relevance Matrix Analysis. A description of the implementation and tools used to develop this procedure, an explanation of the construction of our artificial data set, and the performance criteria are included in Section 4. Based on this information, we present the results and evaluation of the outcome in Section 5. Last but not least, we wrap this thesis up in Section 6 and present our future work in Section 7.

2 Related Work

To achieve the objective of universal classifier that can differentiate between the different neurodegenerative diseases regardless of the origin of the data presented, earlier attempts have been made that are relevant for the development of the approach taken in this project. In this chapter, we lay out the biomedical context of this paper and discuss the different approaches and, ultimately, the findings of the previous attempts.

2.1 Background Information

In this section, we give a brief description of the assumptions and pre-processing steps that allow for machine learning algorithms to be applied to the data that allow for the classification of diseases. In this section, we also go into detail about the Learning Vector Quantization algorithm and how the algorithm of interest, General Matrix Learning Vector Quantization, was derived.

2.1.1 Neurodegenerative Diseases

The brain's nerve cells require glucose as their primary source of cellular energy in order to perform and maintain their function. Cells typically break down native glucose via the glycolysis and mitochondrial oxidative systems, thereby generalising the energy-rich universal biological substance adenosine triphosphate (ATP). If nerve cells are lost or diseased, they require less energy and thus consume less glucose. Therefore, the total glucose consumption in any particular brain region reflects the locally averaged functional state of the brain. For each neurodegenerative disease, there exist patterns of altered glucose use, which could be used to differentiate between them [2]. These properties can be taken advantage of, in order to distinguish between the different neurodegenerative diseases.

2.1.2 Neuroimaging

Positron emmission tomography (PET) scanning in combination with $\left[^{18}F\right]$ fluorodeoxyglucose (FDG) tracer is used to reconstruct the activity distribution in the brain. FDG, when injected in a small amount for a detectable signal, behaves like native glucose but is radiolabeled with positron-emitting radio-isotope $\left[^{18}F\right]$. Therefore, with the help of FDG, the quantification and localisation of the glucose uptake can be determined. [5] The FDG-PET data can be used to train a machine learning algorithm with the objective of differentiating between neurodegenerative diseases.

2.1.3 Notation

A data set is referred to as

$$\mathcal{D} = \{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \in \mathbb{R}^N, y_i \in \{1, ..., C\}\}_{i=1}^{P} \tag{1}$$

where P represents the number of samples $\mathbf{x}_i$, with labels y_i that represent which one of the C unique classes it belongs to.
The set of prototypes is described by

$$\mathcal{W} = \{(\mathbf{w}_j, z_j) \in \mathbb{R}^N \times \{1, ..., C\}\}_{j=1}^{Q} \tag{2}$$

with the constraint $Q \geq C$.

2.1.4 Learning Vector Quantization

Learning Vector Quantization (LVQ) has been introduced by Kohonen [6] and is widely used for classification problems. The prototypes represent classes and they are determined during the training phase of the algorithm using labeled samples. In the training phase, the prototypes are initially distributed in an initial configuration, e.g. randomly or by the class-conditional mean. In each iteration, the closest prototype to the training sample is updated according to two rules. The closest prototype is the determined by the distance measure function $d(\cdot,\cdot)$ where

$$d(\mathbf{w}_J, \mathbf{x}_i) = \min_{\mathbf{w}_j}(d(\mathbf{w}_j, \mathbf{x}_i)), \tag{3}$$

which could be any distance function that provides a positive number. If the label of the nearest prototype is the same as that of the training sample, the prototype is moved towards that sample. If the labels are not the same, the prototype is moved away from the sample data point. This can also be described by the following:

$$\begin{aligned} \mathbf{w}_J &= \mathbf{w}_J + \alpha(t) \cdot (\mathbf{x}_i - \mathbf{w}_J), \text{ if } c(\mathbf{w}_J) = y_i \\ \mathbf{w}_J &= \mathbf{w}_J - \alpha(t) \cdot (\mathbf{x}_i - \mathbf{w}_J), \text{ if } c(\mathbf{w}_J) \neq y_i \end{aligned} \tag{4}$$

where $0 < \alpha(t) < 1$ can stay constant or decrease monotonically. New samples, novel data, can then be classified based on the nearest prototype principle, meaning the sample gets the label of the closest prototype assigned.

Due to the possible divergence problems of LVQ and consequently suboptimal performance, where the prototypes' positions keep moving away from the class distributions [6, 7], another version of the algorithm has been introduced that overcomes this problem.

2.1.5 Generalized Learning Vector Quantization

Generalized Learning Vector Quantization (GLVQ) is an extension of LVQ, introduced by Sato and Yamada [7], which introduces the following objective function:

$$E(\theta, \mathcal{D}) = \sum_{i=1}^{P} e(\theta, \mathbf{x}_i), \text{ with } e(\theta, \mathbf{x}_i) = f[\mu(d_L(\mathbf{x}_i), d_K(\mathbf{x}_i))] \tag{5}$$

where current model parameters θ and data set $\mathcal{D}$ are the input to the function, $f(\cdot)$ is the action function, $\mu(\cdot,\cdot)$ is the discriminant function and $d_L(\cdot)$ and $d_K(\cdot)$ are the distances to the closest prototype with the same label and different label, respectively. μ describes the relative difference distance:

$$\mu(d_L(\mathbf{x}_i), d_K(\mathbf{x}_i)) = \frac{d_L(\mathbf{x}_i) - d_K(\mathbf{x}_i)}{d_L(\mathbf{x}_i) + d_K(\mathbf{x}_i)} \tag{6}$$

For correct classification, $\mu(\cdot)$ is negative and for incorrect classification, $\mu(\cdot)$ is positive. Thus, GLVQ aims to minimise this value during training, to increase the amount of correct classifications. This objective function can be optimised using the gradient descent methods. The update gradient can be described by:

$$\frac{\partial e(\theta, \mathbf{x}_i)}{\partial \mathbf{w}_j(\mathbf{x}_i)} = \frac{\partial f}{\partial \mu} \frac{\partial \mu}{\partial d_j(\mathbf{x}_i)} \frac{\partial d_j(\mathbf{x}_i)}{\partial \mathbf{w}_j(\mathbf{x}_i)} \text{for} j \in \{L, K\} \tag{7}$$

In this extension, the closest prototype with the same label is moved towards the training sample and the closest prototype with a different label is moved away from the training sample. The aim is to minimise the objective value by moving the prototypes in a certain way by means of using the first order derivatives.

2.1.6 Generalized Matrix Learning Vector Quantization

Generalized Matrix Learning Vector Quantization (GMLVQ) is an extension of GLVQ and was introduced by P. Schneider et al [8]. GMLVQ makes use of a parameterised distance measure and a $N \times N$ relevance matrix Λ of adaptive relevances, described by:

$$d^{\Lambda}(\mathbf{w}_j, \mathbf{x}_i) = (\mathbf{x}_i - \mathbf{w}_j)^{\mathsf{T}} \Lambda (\mathbf{x}_i - \mathbf{w}_j) \tag{8}$$

The relevance matrix Λ needs to be a positive semi-definite and symmetric matrix for the distance measure to be a non-negative measure, achieved by substituting $\Lambda = \Omega^{\mathsf{T}}\Omega$ and optimising Ω during training, where Ω is a real $M \times N$ matrix, with $M \leq N$, and is not unique. If $M < N$, it is of limited rank. To ensure numerical stability and interpretability after training, the relevance matrix is normalised by enforcing $Tr(\Lambda) = 1$. The parameters of the matrix are changed so that the distance from the closest correct prototype decreases and the distance to the closest incorrect prototype increases. The update gradient described in the previous section remains the same. The gradient with respect to Ω is described by the following equation:

$$\frac{\partial e(\theta, \mathbf{x}_i)}{\partial \Omega} = \frac{\partial f}{\partial \mu} \frac{\partial \mu}{\partial d_{L,K}(\mathbf{x}_i)} \left(\frac{\partial d_L(\mathbf{x}_i)}{\partial \Omega} + \frac{\partial d_K(\mathbf{x}_i)}{\partial \Omega} \right) \tag{9}$$

The previous equation changes Ω, so that the distance to the closest prototype with the same label is decreased and the distance to the closest prototype with a different label is increased. A GMLVQ system can be trained so that it is able to distinguish between neurodegenerative diseases based on the given processed FDG-PET data, using the scaled subprofile model (SSM)/principal component analysis method (PCA). [2]

2.1.7 Orthogonal Learning Correction

Orthogonal learning correction is an extension of GMLVQ, introduced in [2]. The relevance matrix lambda with the previously described conditions can be written as

$$\Lambda = \sum_{j=1}^{N} \lambda_j \mathbf{v}_j \mathbf{v}_j^{\mathsf{T}}, \tag{10}$$

with N orthonormal eigenvectors $\mathbf{v}_j$ and ordered eigenvalues λ_j in a descending manner such that

$$\lambda_1 \geq \lambda_2 \cdot \cdot > 0 = \lambda_{J+1} = \lambda_{J+1} = \cdot \cdot \lambda_N, \tag{11}$$

which means that the relevance matrix has J eigenvectors with non-zero eigenvalues. The eigenvectors contain the directions that describe the learned label-dependent discriminativeness in the data. They represent the relevance space of the system.

In general, GMLVQ will produce zero-eigenvalue-rich singular matrices Λ with low rank. Since Λ is symmetric, the eigenvectors are orthonormal, allowing for a low-dimensional visualisation of the

data to be constructed by simply projecting the data onto the eigenvectors. [8]

This correction procedure needs directions that describe the subspace that contains the most unwanted variation. Healthy control cohorts are expected to be homogeneous due to their similar activity profiles, a GMLVQ system can be trained on them to discriminate data from different centres. Any useful routes discovered with a GMLVQ system can map to centre differences, allowing for the extraction of the relevance matrix Λ_c, along with its eigenvectors $\mathbf{v}_j$ and eigenvalues. This is due to the data points being mapped in such a way that two centres are as far away from each other as possible. The leading eigenvectors of Λ_c then describe the main directions or linear subspace of the difference between the classes and can be used to construct the correction matrix that limits the training of the second GMLVQ system that is supposed to differentiate between neurodegenerative diseases.

The correction matrix can be constructed like the following:

$$\Psi_c = \left[I - \sum_{j=1}^{J} \mathbf{v}_j \mathbf{v}_j^{\mathsf{T}} \right] = \left[\sum_{j=J+1}^{N} \mathbf{v}_j \mathbf{v}_j^{\mathsf{T}} \right], \tag{12}$$

which describes the nullspace of the centre problem's relevance space, where no relevant directions from the centre classification are present.

This correction is then used in the second GMLVQ system that will attempt to differentiate between neurodegenerative diseases. Since projecting out the relevance space of the system that was trained to distinguish between the centres, the system is not able use any of the centre differences to distinguish between diseases. This results in a relevance space more faithful to the intrinsic differences between the diseases.

To achieve this, we apply the following correction each time Ω_d, omega of the disease problem, is updated:

$$\widetilde{\Omega}_d = \Omega_d \Psi_c \tag{13}$$

As a result, the classifier for the disease problem will ignore these directions when computing distances from the prototypes and will not be misled by the data's centre-specific features. [2]

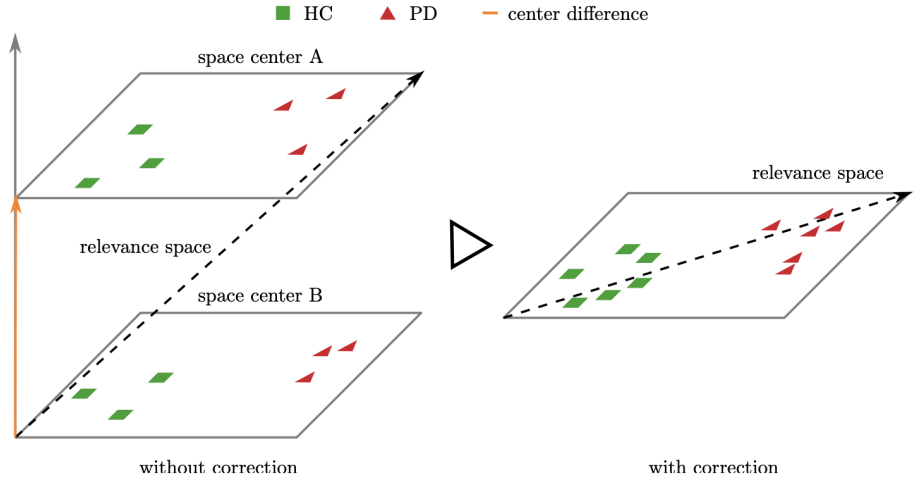

Figure 1: A graphical representation of the problem and the outcome of the orthogonal learning correction technique. Both illustrations have two planes that depict the two separate centres, ”A” and ”B.” The orange line indicates the distinction between them. The systems' learnt relevance space is shown as a dashed arrow. The relevance space of an uncorrected GMLVQ system will be attributed to the centre differences on the left. Any contribution aimed at centre A is projected out to the right. This might be regarded as the data being projected onto the space of centre B in this simplified (and ideal) scenario. Figure reprinted from [2].

2.2 Background Literature

To address the challenge of accurate diagnosis of neurodegenerative diseases, particularly considering the overlapping symptoms and data originating from different centres, several relevant studies have been conducted. In this section, we discuss key findings related to the elimination of source-specific variance and the development of a universal classifier for neurodegenerative disease classification using FDG-PET data.

With transfer learning, one set of data is used to train a classification model, and then the mapped data from another set of sources is used to classify the new set of data [9]. A Taylor expansion assumption allows for an affine transformation to serve as a close approximation to the transfer mapping function [10]. However, if there are not enough training data from a single source to sufficiently prepare the classifier model, this method will fail in a multi-source environment. Moreover, in this setup, the entire data set is used to train the model after initial calibration [11]. In most cases, the meticulous pre-processing is carried out in isolation from the subsequent classification task, leading to the loss of valuable information and hence, subpar classification results [12].

To overcome these difficulties, T. Villmann et al. proposed a null-space transfer classification learning model based on limited rank Generalized Matrix Learning Vector Quantization (GMLVQ) [13]. This siamese-like architecture consists of class prototypes for discrimination between different disease classes and source prototypes for discrimination between different sources. By training a GMLVQ model to determine the source prototypes, the dissimilarity measures based on the model's Ω can be used to detect incorrect classifications of the source and misclassifications of the disease class. This approach takes both source and class information into account in the cost function of the transfer

learning GMLVQ.

Another approach explored in the literature focuses on cross-centre classification of , specifically, FDG-PET data, where the model is trained on data from one centre and tested on data from another centre [2]. However, this approach often results in decreased performance, similarly in [14], suggesting the presence of different patterns in data from different centres. The variation may arise from differences in patient cohorts, disease progression stages, technical equipment, and pre-processing pipelines among centres [2]. It has been observed that the data contains centre-dependent variations, which can be predicted with high accuracy using classifiers trained to assign data samples to their centre of origin [2]. This further demonstrates the existence of centre-specific variances in the data, and failing to take this into account could lead to misleadingly higher performance and biases in the interpretation of the model's conclusions.

In an attempt to eliminate centre-specific variation, a two-step approach has been proposed [2]. First, a GMLVQ system is trained on centre data to identify a subspace containing centre-dependent variation. Then, this subspace is used to correct another GMLVQ system trained on disease data. This method, called the orthogonal learning correction procedure, see Section 2.1.7 for detailed explanation, involves the use of a correction matrix to restrict the classifier from using features that contain unwanted source-dependent variation. Consequently, the GMLVQ system ignores the directions with centre-specific properties when computing distances from the prototypes. While this approach can partially remove centre-dependent variation, it may not identify all possible directions that contain centre differences [2].

This latter method is able to partially remove centre-dependent variation but not all, since the GMLVQ system might not have been able to find all possible directions that contain centre differences [2].
To further improve the performance and develop a universal classifier for neurodegenerative disease classification that relies on the subspace that is most relevant to the disease classes, it is crucial to identify and eliminate all directions with centre differences. This will enable the construction of a classifier capable of distinguishing between neurodegenerative diseases regardless of the centre from which the data originated.

3 Concept

The concept of the thesis is based on the Iterated Relevance Matrix analysis (IRMA) that enables us to remove the centre-specific variance from the data. This method, therefore, allows us to define the iterative subspace correction algorithm described in this section. With this method, we aim to construct a universal classifier that should be able to classify neurodegenerative disease from FDG-PET images regardless of the neuroimaging centre of origin.

3.1 Algorithm

The Iterated Relevance Matrix Analysis has been developed by Sofie Lövdal and Michael Biehl [15], which seeks to define the most class-relevant subspace by repeatedly discarding dominantly relevant directions and retraining the GMLVQ system with the remaining data.

The initial uncorrected GMLVQ system trained on a two-class problem often finds a single most discriminative direction $\mathbf{v}_j^{(0)}$ where $\lambda_1^{(0)} \approx 1$ and $\Lambda^{(0)} \approx \mathbf{v}_1^{(0)}\mathbf{v}_1^{(0)\mathsf{T}}$. As a result, the following is a restriction placed on the second round of GMLVQ system training:

$$\Omega^{(1)}\mathbf{v}_1^{(0)} = 0, \tag{14}$$

on the distance measure:

$$\Lambda^{(0)} = \Omega^{(1)\mathsf{T}}\Omega^{(1)} \tag{15}$$

which can be accomplished by performing the following correction after each update step:

$$\Omega^{(1)} \rightarrow \Omega^{(1)}\left[I - \mathbf{v}_1^{(0)}\mathbf{v}_1^{(0)\mathsf{T}}\right] \tag{16}$$

As a result, the feature space will disregard the direction $\mathbf{v}_1^{(0)}$ for the classification, and the resulting Λ has the leading eigenvector $\mathbf{v}_1^{(1)}$, this characterises the orthogonal direction with the highest discriminatory power relative to the eigenvector we disregarded. In an iterative process, we obtain a sequence of eigenvectors $\mathbf{v}_1^j$ that is orthogonal to the previous eigenvectors $\mathbf{v}_1^{(i)}$ with $i = 1, 2, .., j-1$. At each step for $j \geq 1$, after the update step, the correction

$$\Omega^{(j)} \rightarrow \Omega^{(j)}\left[I - \sum_{i=0}^{j-1} \mathbf{v}_1^{(i)}\mathbf{v}_1^{(i)\mathsf{T}}\right] \tag{17}$$

is applied. The uncorrected GMLVQ system is referred to as the 0-*th iteration*.

When the classifier in iteration $k+1$ reaches merely random or near-random classification performance, the operation can be stopped.

The following algorithm provides a rough description of this approach within the scope of this paper:

Algorithm 1: Iterative subspace correction

Data: Correction matrix = Identity matrix, Saved centre eigenvectors = empty
while *stopping criterion not met* **do**
- a. Train GMLVQ to distinguish between centre 1 and centre 2. The correction matrix is incorporated in the training such that the relevance matrix is multiplied by the correction matrix for each update step.
- b. Extract the leading eigenvector of the relevance matrix and add it to the set of saved centre eigenvectors.
- c. Construct the new correction matrix based on all eigenvectors in the set of saved centre eigenvectors.

end

Several stopping criteria could be considered, but a reasonable approach that we use in this thesis is to halt when validation accuracy with regard to discrimination of centres matches to random accuracy.

4 Experiments

In this section, we provide the technical details of the implemented algorithm as well as the construction and parameter settings of the artificial data sets in terms of Gaussian clusters that we use in this thesis. We consider three cases of data sets with increasing complexity to observe the method's behavior and compare the results of the cases.

4.1 Experimental Setup

4.1.1 Tools

To implement the algorithm and complete the experiments, Python (version 3.10.11) in combination with the scikit-learning vector quantization (sklvq) package [16], which is an open-source Python implementation of a set of Learning Vector Quantization algorithms and compatible with the Python machine learning package, scikit-learn, is used.

4.1.2 Settings

In our experiments, we use the identity function as the activation function, denoted as $f(\mathbf{x}) = \mathbf{x}$. However, alternative choices such as sigmoid, soft+, and swish are also available.

For the GMLVQ algorithm, the options for discriminant functions are quite limited, only allowing for the relative distance function. This function is defined by Equation 6.

In our implementation, we employ the adaptive squared Euclidean distance function, which is specifically compatible with the GMLVQ algorithm. The formulation of this distance function can be found in Equation 8. Alternatively, other distance function options include Euclidean distance, squared Euclidean distance, and local adaptive squared Euclidean distance.

To minimise the objective function, we employ solvers. In our experiments, the selected solver is the waypoint gradient descent, which is a modification of batch gradient descent. This solver incorporates the waypoint average optimization algorithm introduced by G. Papari et al. [17]. Additional implementation details can be found in [16]. Other available solver options include steepest gradient descent, adaptive moment estimation, Broyden Fletcher Goldfarb Shanno (BFGS), and limited memory variant of BFGS.

4.2 Artificial Data Set

The implementation will undergo several experiments and elaborate testing on artificial data before it can be applied to neuroimaging data. This artificial data allows us to approach this in a controlled way, instead of using neuroimaging data.

The artificial data we consider in this thesis, produced by Sofie Lövdal, is constructed in terms of Gaussian clusters with 200 data points per cluster with known parameters that represent the classes in a simplified way, see Figure 2. We consider three cases of increasing complexity.

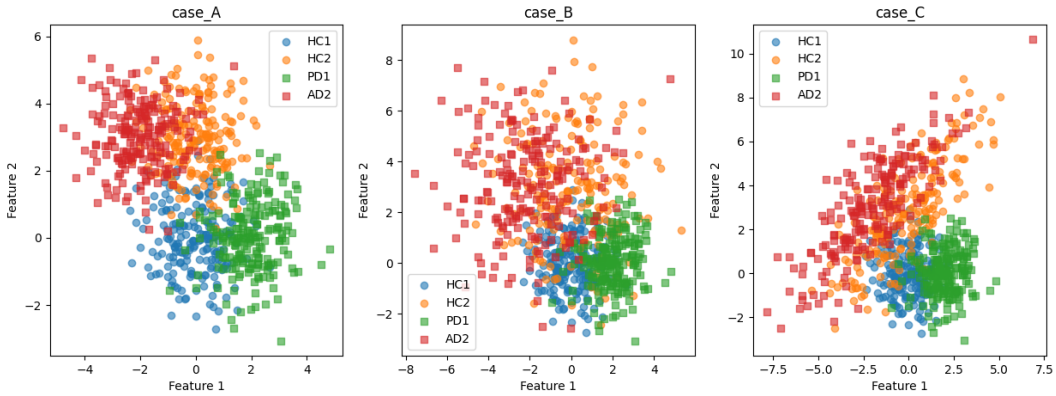

Figure 2: Artificial data sets of increasing complexity that correspond to two Healthy Control (HC) cohorts from centre 1 and centre 2, as well as a cohort of Parkinson's disease (PD) that is only available from centre 1 and a cohort of Alzheimer's disease (AD) that is only available from centre 2.

The information has been created using Gaussian distributions with fixed parameters. We also provide a transformation description of the centre effect as $x' = Ax + b$. Only centre 2 is modified; the difference between the two is expressed as a linear transformation with respect to centre 1.

The offset between the means of the HC clusters (before any centre effect would have been applied) is indicated by $b = [0,2]$. For the PD1 cluster vs HC1, we assume an offset of $d1 = [0,2]$, and for the AD2 cluster vs HC2, we assume $d2 = [0,-2]$. After any conceivable transformation by Matrix A, the disease offset is added.

- Case A: $A = [1,0;0,1]$ (The only true effect is the offset from the centre.)
- Case B: $A = [2,0;0,2]$ (The centre effect, in addition to the offset, causes greater variance for centre 2.)
- Case C: $A = [2,1;1,2]$ (In addition to the offset, the centre effect causes a larger variance, but it also creates a covariance effect between features 1 and 2.)

The normalised ground truth for the centre difference and disease difference is $[0,1]$ and $[1,0]$, respectively. Furthermore, the artificial data contains three noise features.

An overview of the parameters of the artificial data sets, shown in Figure 4.2, can be found in Appendix Section A.

4.2.1 Experiment setup

For our purposes, we have split the artificial data in a training set and a testing set of the same size, namely 100 data points per cluster, see Figure 3.

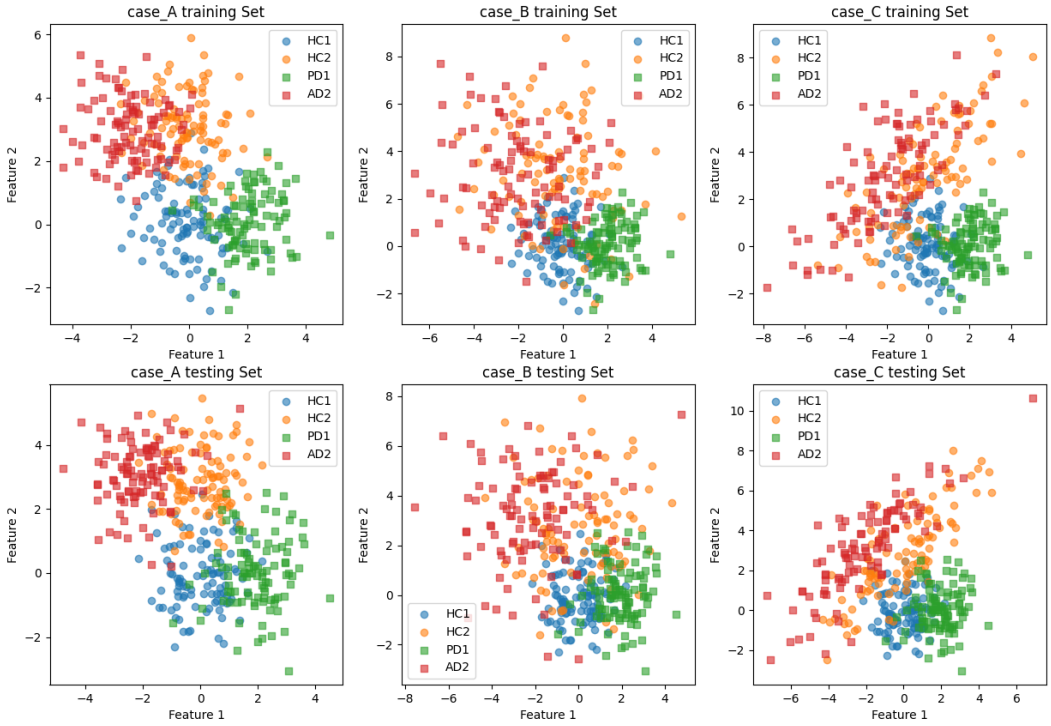

Figure 3: Training set for each case can be found in the top row and the corresponding testing sets are below.

For our experiments, we randomly sample the training and testing set with $n = [10, 20, 100]$ number of data points per cluster, in order to see the differences in performances.

We also experiment on transformed artificial data in terms of the centre-wise z-score:

$$x' = \frac{x - \mu_c}{\sigma_c},$$

where the mean and standard deviation are of the respective training centre data corresponding to where the data point stems from. In this setting, we only need to train the uncorrected GMLVQ on the disease problem, once. We assume that the centre-wise z-score takes care of removing all centre-specific information, and therefore this transformed data is not in need of any corrections.

For more reliable results, we execute the experiments several times and average the results, as well as, extract the standard deviation between the runs. The training sets and testing sets are drawn from the initial split of the complete set, see Figure 3, and remain the same throughout the same run but are random for each run separately.

4.3 Performance Criteria

To evaluate the performance of our modified method, we divided the artificial data into training and testing sets of equal size, comprising 100 data points per cluster, as depicted in Figure 3.

It is important to note that evaluating the results in this context is a complex task due to the inherent difficulty of the problem and the various factors that can influence the outcomes.

In case A and case B, where the ground truth is described by vector d, the evaluation can be relatively straightforward. However, in case C, the evaluation becomes more intricate due to the effect of covariance. We acknowledge that our chosen evaluation methods represent one approach among many possible alternatives.

The performance is evaluated in the following ways:

We calculate the accuracy of the trained model in each iteration by using a test set of the same size and same amount of noise as the original set used for training. We also calculate the accuracy of the training set itself after GMLVQ has been trained. We do this for both the centre problem and the disease problem. There should be a drastic drop in performance for centre discrimination attempts following the adjustment. We hope to see the accuracy for the centre problem to degrade to 0.5 accuracy, which indicates that the classification is arbitrary. In fact, as the process is repeated, performance should degrade further.

After correcting the training on the disease cohorts, we measure how well the similar cohorts match. We calculate the squared distance of the means of the two centre clusters and of the two disease clusters as a function of the number of iterations. We also calculate the Frobenius norm between the covariance matrices of the two centre clusters and between the two disease clusters as a function over the iterations. With this information, we can interpret the results better, knowing the separation and behavior of the clusters.

Additionally, we calculate the angle between the leading eigenvector of the disease problem and the ground truth disease direction as a function over the iterations. This allows us to see how close the leading eigenvector of the disease problem is to the ground truth and in what way this angle changes as the process is repeated. For example, if the angle rises after a removal, we can conclude that disease information has been removed as well.

Furthermore, we calculate the subspace angle between the set of accumulated eigenvectors and the ground truth direction of the centre difference. The eigenvectors are removed from the feature space of the disease classification in an iterative manner. With this calculation, we can see how much of the subspace covers the ground truth and know how much centre-specific variation we have removed from the feature space of the disease classification. The closer the subspace angle is to 0, the better, since this allows us to conclude that we have removed all centre variation.

5 Results & Evaluation

In this section, we present and evaluate the results from our experiments with respect to our performance criteria, described in Section 4.3, for each case introduced in Section 4.2.

5.1 Case A

This section focuses on the simplest case, namely case A. The construction of this case is explained in Section 4.2.

In Figure 4, we present the accuracy trends observed during our analysis. The analysis focuses on the behavior of accuracy in relation to data size, the uncorrected GMLVQ and the removal of $\mathbf{v}_1^{(i)}$ for $i < 4$.

Due to the simplicity of this data, we can observe that the accuracy behaves in similar patterns regardless of the size.

Regarding the centre classification, we observe that the accuracy drops to a near arbitrary level after the second removal. This suggests that even with limited data, the model can effectively classify the centre information once $\mathbf{v}_1^{(0)}$ is removed.

For the disease classification, the accuracy remains high but experiences decreases with each removal. We hypothesise that this decline may be attributed to the removal of disease-related information during the iterative process. These decreases become bigger after the removal of $\mathbf{v}_1^{(1)}$.

Notably, after the third and fourth removals, the disease classification accuracy demonstrates significant reductions. This indicates that the removals may have eliminated additional disease-related information, causing centres to become indistinguishable and leading to a lack of informative leading eigenvectors.

In the smaller data sets, the GMLVQ systems face challenges in identifying the directions that describe centre variation. This is evident from the clear downward distances observed between the centre classification curves of the smaller data sets compared to that of the largest data set with $n = 100$.

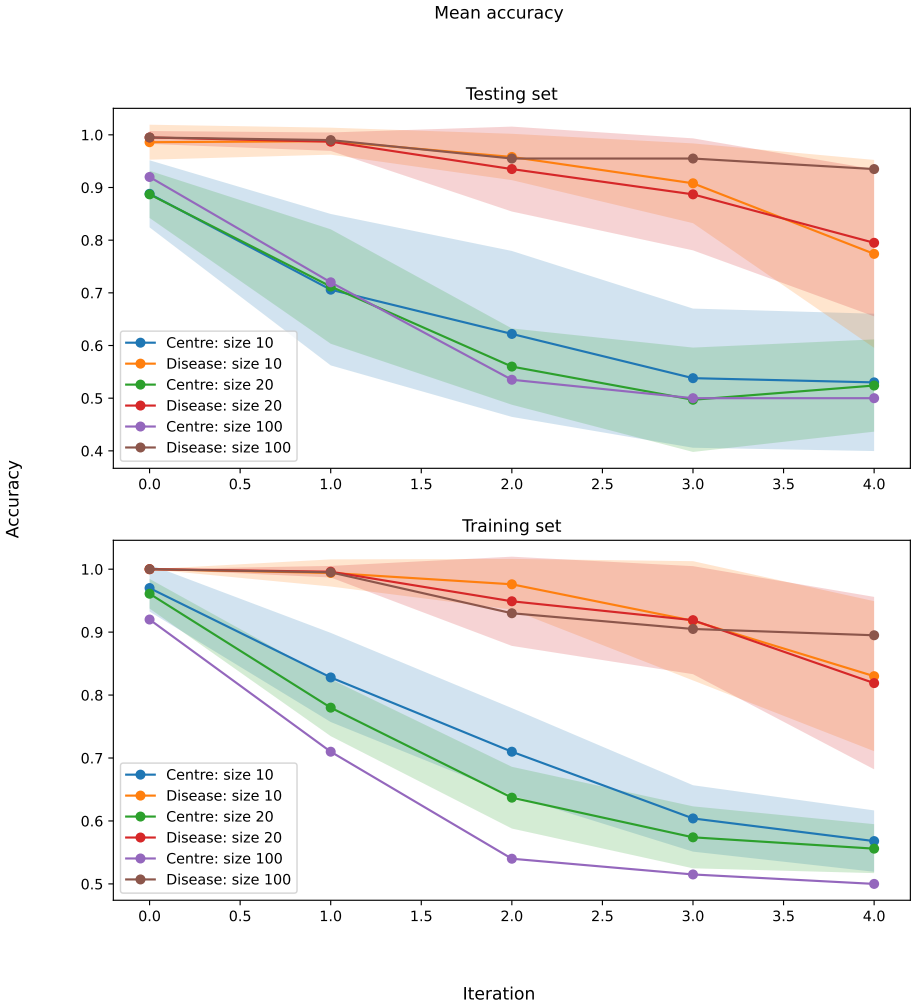

Figure 4: Mean accuracy and standard deviation as a function of the number of iterations of the testing set and training set for case A. The 0-*th iteration* is the uncorrected GMLVQ system.

In Figure 5, we present the accuracy results obtained from the centre-wise z-scored data, using only the uncorrected training processes.

The accuracy analysis reveals that the centre classification accuracy for the testing set remains close to 0.5 for all sizes, accompanied by high accuracy (≥ 0.9) for the disease classification. However, when employing the iterative method, we achieve this level of performance only around the second and third removal stages for the testing set. By the third removal, we consider that too much information has been eliminated, resulting in a reduced acceptable subspace comprising only $\{\mathbf{v}_1^{(0)}, \mathbf{v}_1^{(1)}\}$.

Furthermore, the metrics obtained from the training sets for the centre and disease data demonstrate a significant alignment with the metrics observed after the removal of $\mathbf{v}_1^{(0)}$.

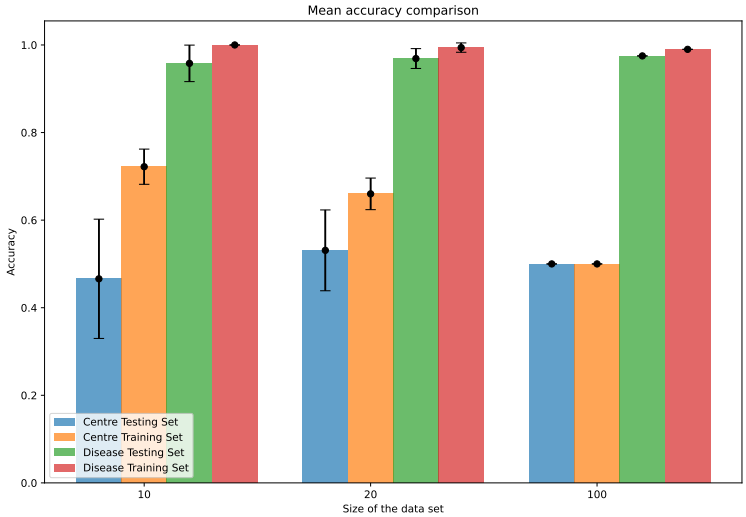

Figure 5: Mean accuracy and standard deviation of the centre-wise z-scored training set and testing set for case A.

With Figure 6, we illustrate the mean squared distance between the means of the centre clusters and the means of the disease clusters. Additionally, it includes the Frobenius norm between the covariance matrices of the clusters.

Upon analysing the results, we observe that the removal of $\mathbf{v}_1^{(0)}$ already significantly reduces the distance between the centre clusters, bringing them closer together. Following the removal of $\mathbf{v}_1^{(1)}$, the distance between the clusters becomes close to zero, indicating overlapping clusters. Notably, there is a noticeable abrupt decline in distance from the first removal to the second removal of the distance between the disease cluster means, suggesting that the second removal likely involved the elimination of disease-related information.

The Frobenius norm between the covariance matrices remains consistently low throughout the iterations, with only minimal increases observed. This indicates that the removal of $\mathbf{v}_1^{(i)}$, for $i < 4$, has a limited impact on the covariance structures of the clusters. Additionally, it is worth noting that the clusters have identical covariance to begin with in this case, so it is natural that the Frobenius norm is approximately zero for all cases. The stability of the Frobenius norm, despite the iterative process, further suggests that the removal of $\mathbf{v}_1^{(i)}$ does not significantly alter the covariance relationships between the data points within each cluster.

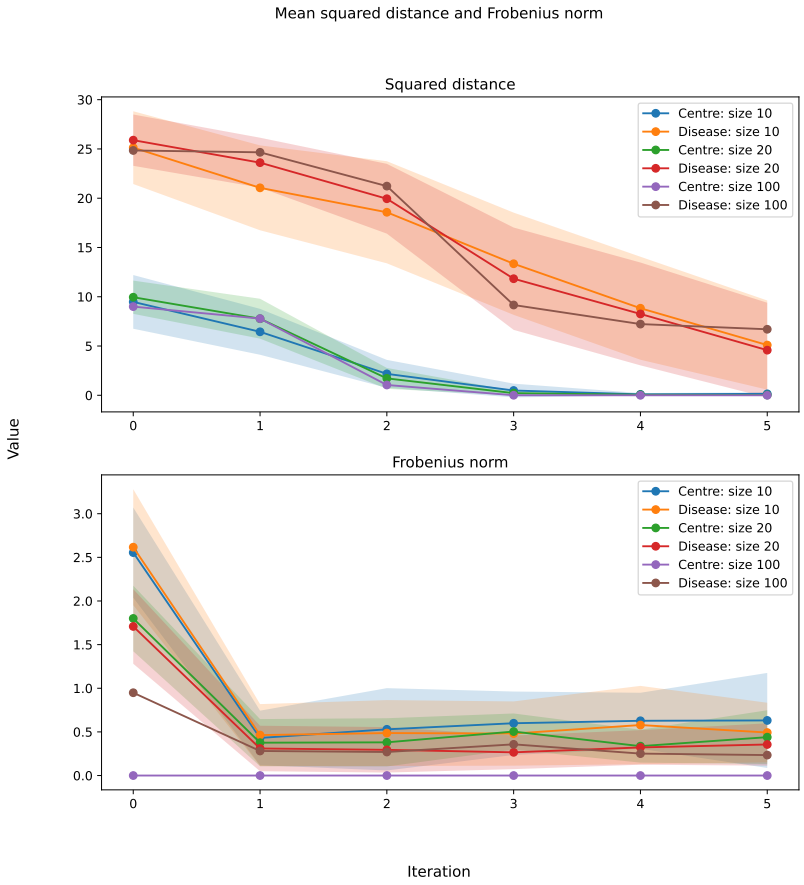

Figure 6: Mean squared distance and Frobenius norm as a function of the number of iterations of the training set. Iteration 0 is the original training set, Iteration 1 correlates to the 0-*th iteration* (the uncorrected GMLVQ) and the remaining iterations represent the metrics after the removal(s) of the leading eigenvector(s).

In Figure 7, we present the same metrics for the centre-wise z-scored data.

The distance between the means of the centres remains consistently low across the different values of n, primarily due to the transformation applied. The metrics for the Frobenius norm between the covariance matrices show a slightly lower range, but their patterns align with those observed in the iterative approach in Iteration 0 and Iteration 1.

Regarding disease classification, the previously mentioned alignment between the metrics persists. The distance between the means of the disease clusters remains within the range of 20 to 25, indicating a clear separation between the disease clusters.

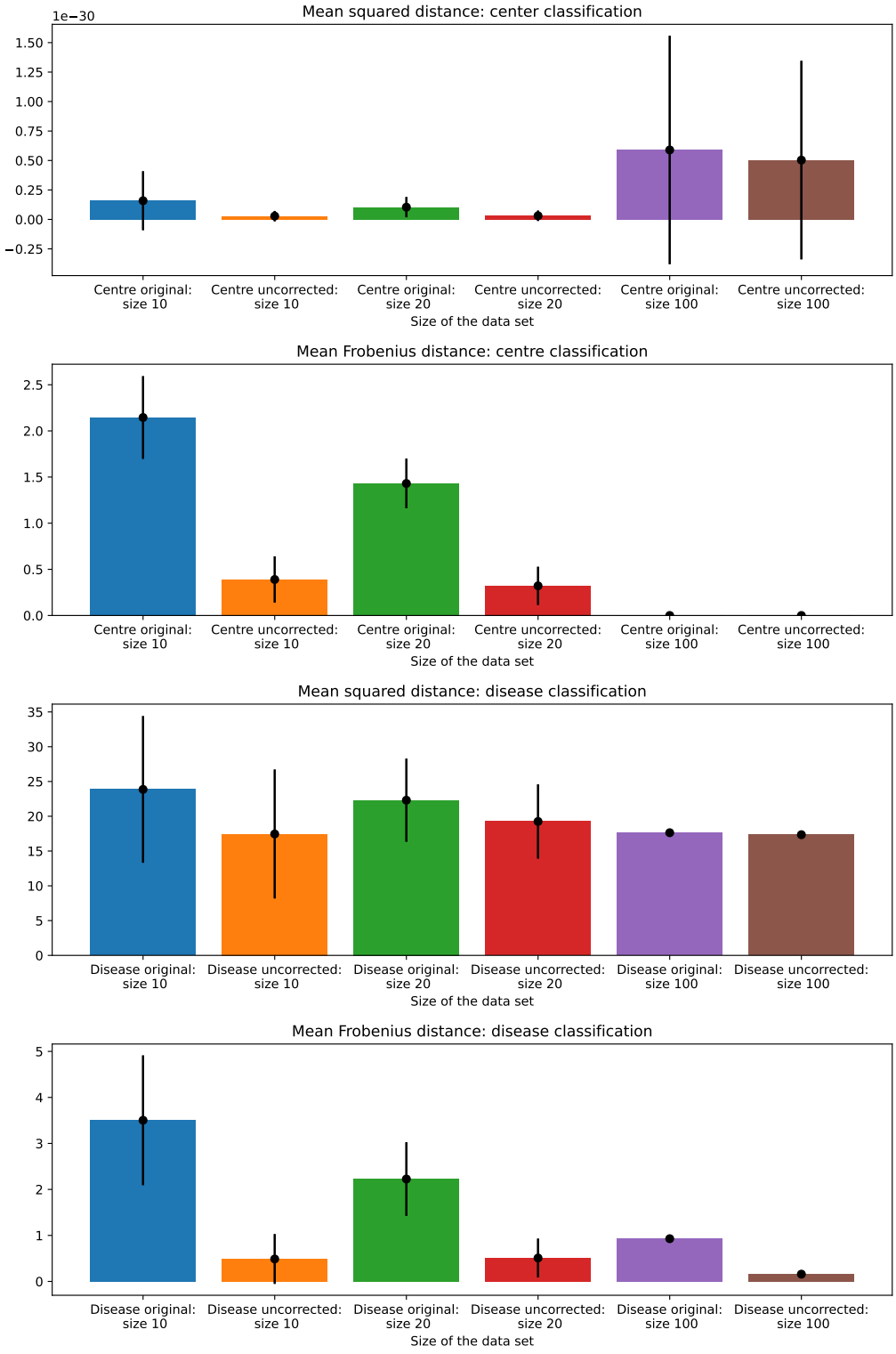

Figure 7: Mean squared distance and Frobenius norm for various sample sizes.

In Figure 8, we examine the angle between the leading eigenvectors of the disease classification and the ground truth disease direction described in Section 4.2. The results confirm our assumptions.

The correction of $\mathbf{v}_1^{(0)}$ noticeably reduces the influence of centre information on the leading eigenvector of the disease classification, aligning it more closely with the ground truth disease direction that captures the disease variation. Interestingly, for the smaller data sets, this angle tends to be larger. However, as observed in Figure 4, the accuracy is only minimally impacted by this discrepancy. Additionally, the second removal appears to eliminate disease-related information, which aligns with the findings from Figures 4 and 6. It is also noteworthy that while the angle after the second removal is similar to that of the 0-*th iteration*, the accuracy is slightly lower compared to the 0-*th iteration*.

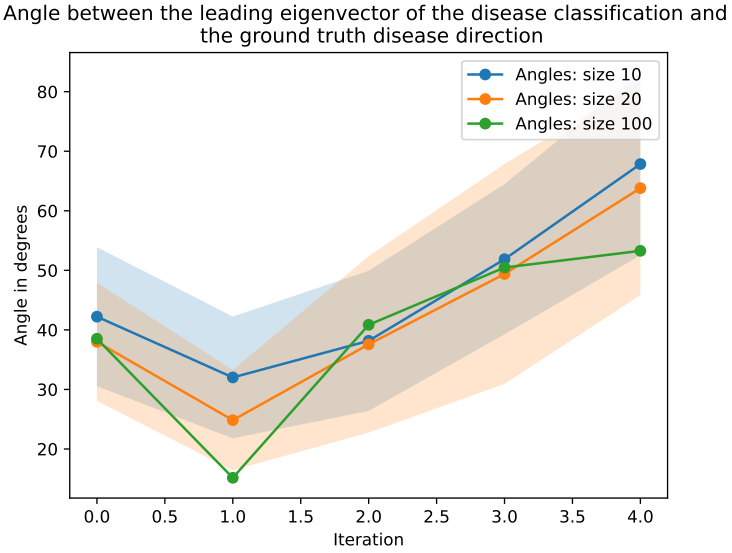

Figure 8: Angle between the leading eigenvector of the disease classification and the ground truth disease direction as a function of the number of iterations. The 0-*th iteration* is the uncorrected GMLVQ system.

In Figure 9, we present the same metrics for the centre-wise z-scored data.

The results demonstrate that as we include more data points, the classifier's performance improves, resulting in a leading eigenvector that closely aligns with the ground truth.

Comparing Figure 8 and Figure 9, we observe significant differences in the angles when we map the accuracy results of the iterative approach to the accuracy results of the centre-wise z-scored data. Namely, focusing on the second removal since the accuracies for the centre classification are both at an arbitrary level. In Figure 8, the angles after the removal of $\mathbf{v}_1^{(1)}$ are around 40 degrees, whereas in Figure 9, they remain below 30 degrees and decrease further as the number of data points per cluster (n) increases. Notably, for $n = 100$, the angle is even below 10 degrees. These findings suggest that the iterative approach may remove more disease-related information, resulting in a worse alignment between the leading eigenvectors and the ground truth.

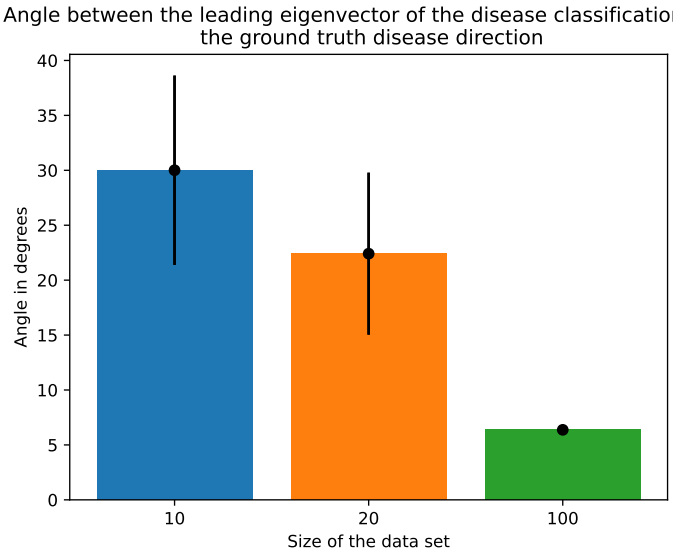

Figure 9: Angle between the leading eigenvector of the disease classification and the ground truth disease direction.

Figure 10 shows the subspace angle between the set of accumulated eigenvectors of the centre classification and the ground truth centre direction. It shows us that it takes the method two removals to achieve a subspace angle of close to zero for the biggest data set and four removals for the smaller sets. It is important to note that the removal of $\mathbf{v}_1^{(1)}$ not only reduces the subspace angle but also removes disease information. This highlights the trade-off between achieving a more aligned subspace and potential loss of disease-related information.

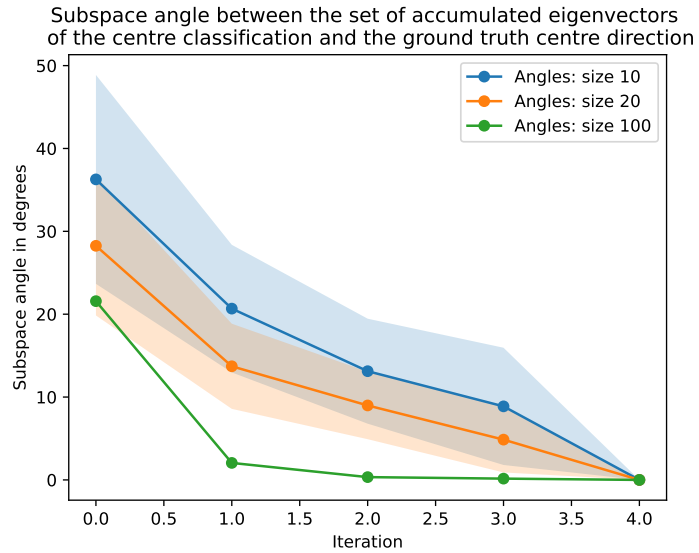

Figure 10: Subspace angle between the set of accumulated eigenvectors of the centre classification and the ground truth centre direction as a function of the number of iterations.

All in all, it seems like if we were to stop the iterative method after the removal of $\mathbf{v}_1^{(0)}$, we already have a subspace that defines the ground truth centre direction with an angle close to zero. We also

remain at a high accuracy for the disease classification and the main direction GMLVQ finds deviates by ≈ 35 degrees from the ground truth disease direction. As the number of data points per cluster increases, this angle decreases and approaches zero. However, we would have to take into account that the centre classification does not achieve an arbitrary level and there are still traces of centre specific information in the disease data, confirmed by the subspace angle.

To remove a subspace that defines the ground truth centre direction better, we would have to take into account a lowered accuracy of the disease classification by ≈ 0.05 compared to when we only remove $\mathbf{v}_1^{(0)}$. The main direction of the disease classification shows an angle of ≈ 40 degrees to the ground truth disease direction. However, the subspace angle can be considered zero. This indicates that if we want to remove more centre-specific variation, we would also be removing a certain amount of disease information.

On the other hand, the centre-wise z-scored data exhibits a slightly improved performance. For this data as well, including more data points leads to better classifier performance and a leading eigenvector that closely aligns with the ground truth. Comparing the angles between the leading eigenvectors and the ground truth, the centre-wise z-scored data shows significantly smaller angles, even approaching zero for larger data set sizes while maintaining a level of arbitrary classification for the centre problem.

Overall, these findings highlight the effectiveness of the iterative method in aligning the subspace with the ground truth, while emphasizing the challenges in completely removing centre-specific information from the disease data.

5.2 Case B

This section focuses on the more complex case B. The construction of this case is explained in Section 4.2.

In Figure 11, we present the accuracy trends observed during our analysis. The analysis focuses on the behavior of accuracy in relation to data size, the uncorrected GMLVQ and the removal of $\mathbf{v}_1^{(i)}$ for $i < 4$. Due to the increased complexity, we can observe that the accuracy curves show more noticeable variations, compared to case A.

Regarding the centre classification, we observe that the accuracy drops to a near arbitrary level at a later stage, specifically after the third removal. This suggests that the GMLVQ systems face challenges in accurately capturing the main directions that differentiate the centres. The removals have minimal impact on the centre accuracy in each iteration.

However, for the largest data set with $n = 100$, when $\mathbf{v}_1^{(0)}$ is removed, we can observe a similar decrease in accuracy as in case A, even though the uncorrected GMLVQ system already exhibits lower accuracy due to the increased complexity. This indicates that the main direction that distinguishes the centres has been identified quite accurately while still maintaining a high accuracy for the disease classification.

For the disease classification, the accuracy remains high but experiences more significant decreases with each removal compared to case A. There is notable decrease between the removal of $\mathbf{v}_1^{(1)}$ and $\mathbf{v}_1^{(2)}$ for $n = 100$, which indicates that a larger amount of disease information might have been removed with $\mathbf{v}_1^{(2)}$.

Similar to scenario A, the performance on the training set is poorer for the centre classification, but not for the disease, when compared to that on the testing set.

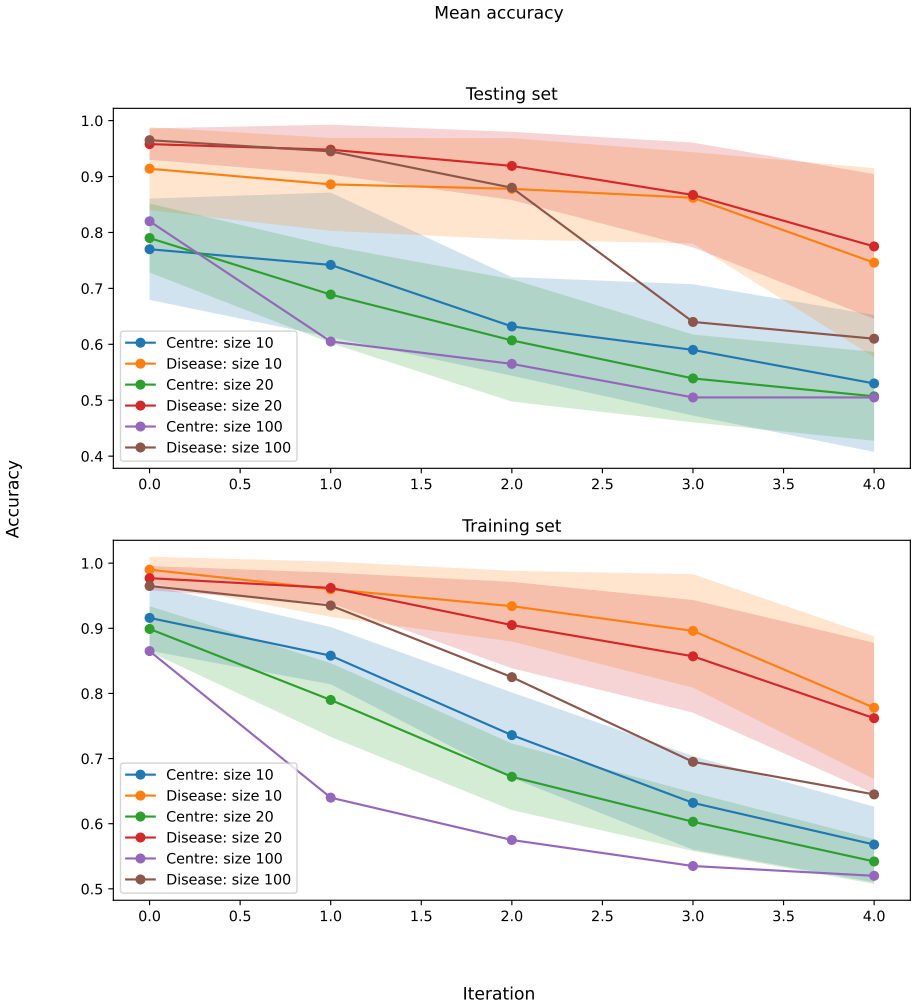

Figure 11: Mean accuracy and standard deviation as a function of the number of iterations of the testing set and training set for case B. The 0-*th iteration* is the uncorrected GMLVQ system.

In Figure 12, we present the accuracy results obtained from the centre-wise z-scored data, focusing on the uncorrected training processes.

The analysis of accuracy reveals that the centre classification accuracy for the testing set remains consistently around 0.5 across all data sizes, indicating a near arbitrary level of classification. However, the disease classification accuracy remains high, approximately 0.9. It is worth noting that the disease accuracy is slightly lower compared to case A.

When applying the iterative method, we observe similar performance levels for the testing set in the centre problem during the last two iterations. However, this decline in accuracy for the disease classification suggests that disease-related information has also been unintentionally removed. This highlights the challenge of balancing the removal of irrelevant information while preserving the relevant information.

Furthermore, the metrics obtained from the training sets for the centre classification align with the removal of $\mathbf{v}_1^{(2)}$. The disease classification shows very high accuracies for all n.

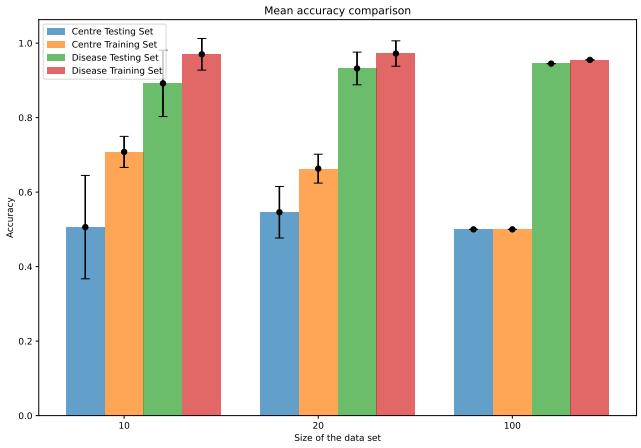

Figure 12: Mean accuracy and standard deviation of the centre-wise z-scored training set and testing set for case B.

With Figure 13, we illustrate the mean squared distance between the means of the centre clusters and the means of the disease clusters. Additionally, it includes the Frobenius norm between the covariance matrices of the clusters.

Upon analysing the results, several observations can be made. First, the removal of $\mathbf{v}_1^{(0)}$ already leads to a significant reduction in the distance between the centre clusters, bringing them closer together. Subsequently, after the removal of $\mathbf{v}_1^{(1)}$, the distance between the clusters becomes close to zero, indicating overlapping clusters.

Furthermore, when considering the distance between the disease cluster means for $n = 100$, there is a noticeable abrupt decline from the first removal to the second removal. This sharp decrease suggests that the second removal likely involves the elimination of disease-related information, resulting in the clusters being brought much closer together. This finding aligns with the observations made in Figure 11.

Additionally, the distance between the means of the smaller data sets, where $n = [10, 20]$ gradually decreases in each iteration. This indicates that with each removal, more disease-specific information is being eliminated, leading to clusters that are closer together. This can also be seen in the gradual decline of the accuracy in Figure 11.

We can also see that the Frobenius norm between the covariance clusters is not as low as for case A, which is due to the way we constructed case B with centre 2 containing a greater variance. The decrease of the Frobenius norm in each iteration is due to the removal of the leading eigenvectors since these directions capture the variations specific to each class, and their removal reduces the dissimilarity between the classes. We can see that for the smaller sets, the curves behave the same, indicating that disease information has also been removed. For $n = 100$, however, we see the Frobenius norm for the centre clusters decrease to a value below 2 but the Frobenius norm for the disease clusters remains at the same value as in the uncorrected GMLVQ iteration, further supporting the assumption that the direction of the main difference of the centre classification has been estimated quite accurately.

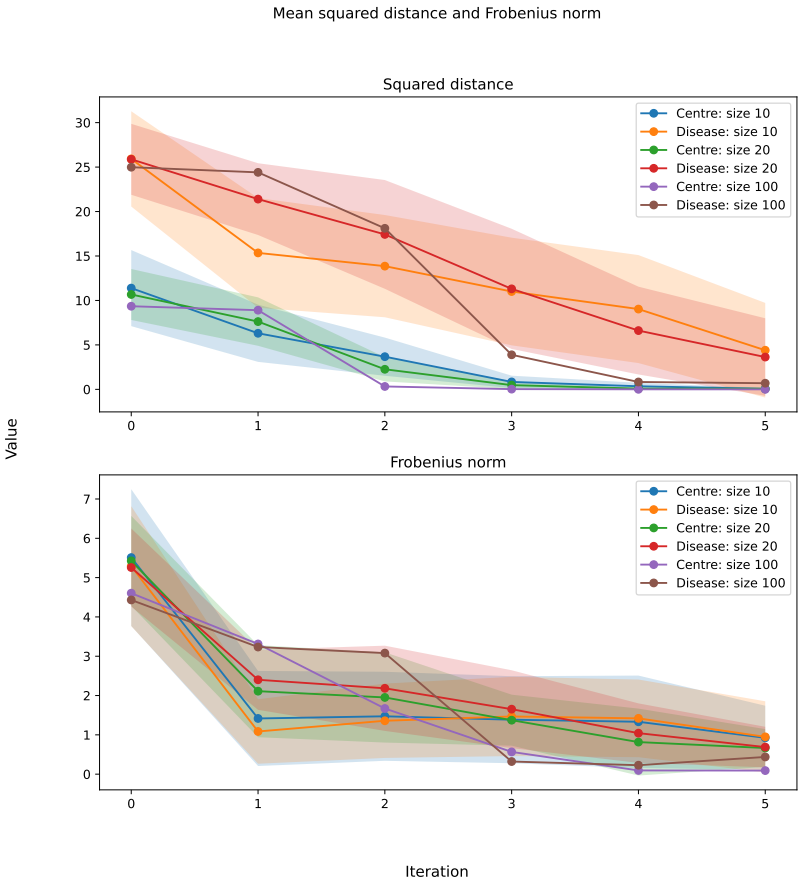

Figure 13: Mean squared distance and Frobenius norm as a function of the number of iterations of the training set. Iteration 0 is the original training set, Iteration 1 correlates to the 0-*th iteration* (the uncorrected GMLVQ) and the remaining iterations represent the metrics after the removal(s) of the leading eigenvector(s).

In Figure 14, we present the same metrics for the centre-wise z-scored data.

The distance between the means of the centres remains consistently low across the different values of n, primarily due to the transformation applied. The squared distances between the means of the disease clusters are slightly lower than those observed in case A. The metrics for the Frobenius norm between the covariance matrices show a lower range, but their patterns align with those observed in the iterative approach in Iteration 0 and Iteration 1.

These results are very similar to that of Figures 7 and 14.

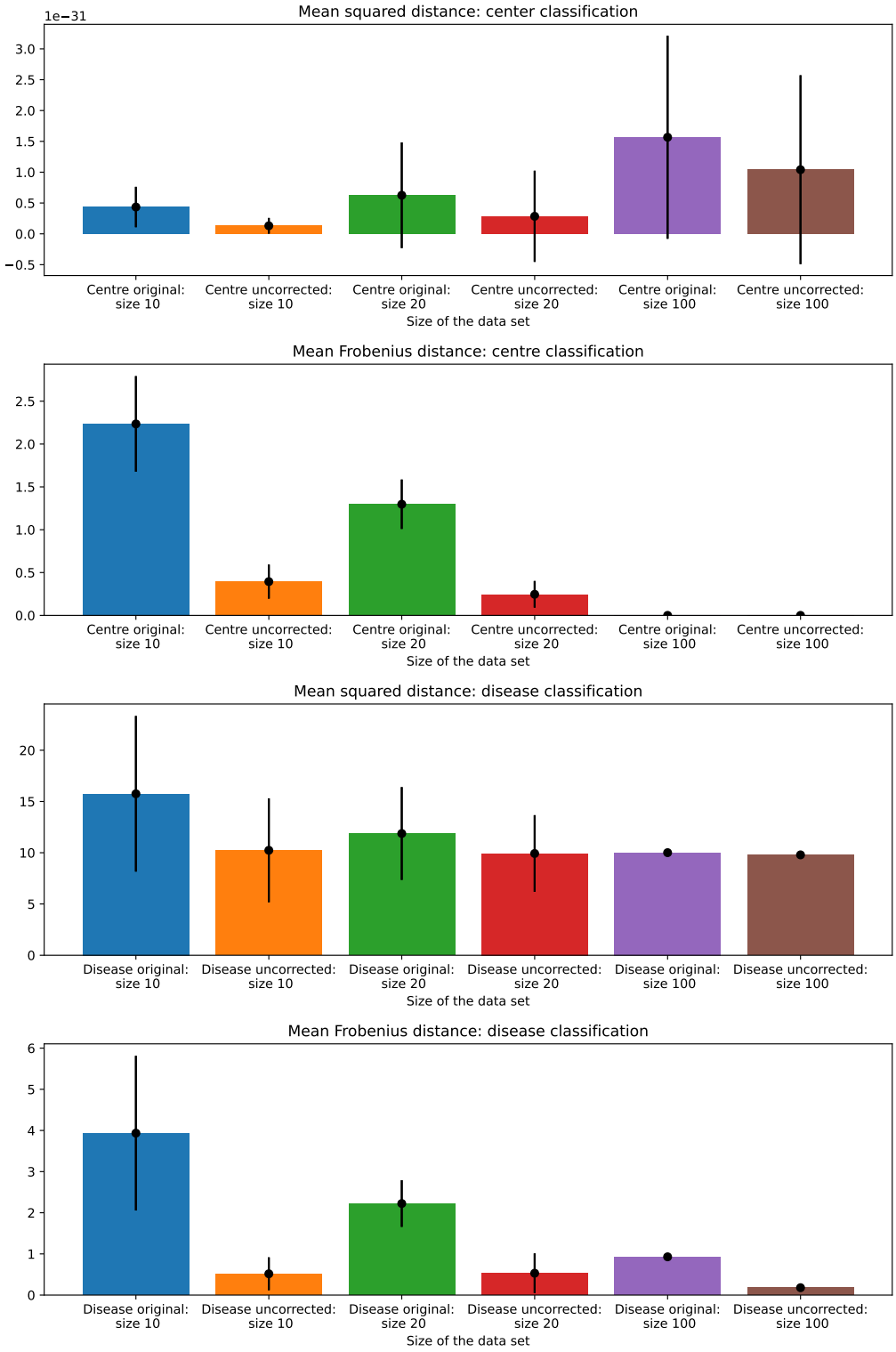

Figure 14: Mean squared distance and Frobenius norm for various sample sizes.

In Figure 15, we examine the angle between the leading eigenvectors of the disease classification and the ground truth disease direction described in Section 4.2.

The results reveal several noteworthy findings. Firstly, the correction of $\mathbf{v}_1^{(0)}$ significantly reduces the influence of centre information on the leading eigenvector for $n = 100$, aligning it more closely with the ground truth disease direction that captures the disease variation. This finding supports the observations made in Figures 11 and 13, indicating that the removal of $\mathbf{v}_1^{(0)}$ improves the accuracy of capturing disease-related information.

For $n = 20$, a slight drop in the angle is observed at the same stage of removal, indicating a similar trend, albeit to a lesser extent. However, for $n = 10$, a smaller decrease is observed when $\mathbf{v}_1^{(1)}$ is also removed, suggesting that GMLVQ struggles to effectively recognise the underlying difference with a limited number of data points.

Furthermore, the relatively large angles for $n = [10, 20]$ throughout all iterations indicate the difficulty faced by GMLVQ in recognising the underlying difference when an insufficient number of data points are provided.

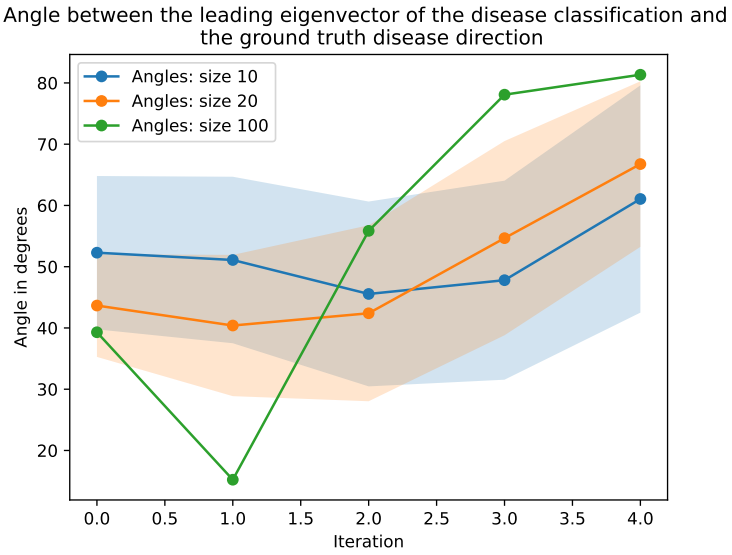

Figure 15: Angle between the leading eigenvector of the disease classification and the ground truth disease direction as a function of the number of iterations for case B. The 0-*th iteration* is the uncorrected GMLVQ system.

In Figure 16, we present the same metrics for the centre-wise z-scored data.

A comparison between Figure 15 and Figure 16 reveals notable differences in the angles obtained. Even for the smallest data size ($n = 10$), the GMLVQ algorithm with the centre-wise z-scored data produces a leading eigenvector that has an angle of approximately 30 degrees with respect to the ground truth disease direction. This is in contrast to Figure 15, where none of the smaller data sets achieve such a significant angle in any of the iterations. Similar observations can be made for the data size of $n = 20$.

These findings indicate that using the centre-wise z-scored data enhances the alignment between the leading eigenvectors and the ground truth disease direction. The GMLVQ algorithm with this transformed data is able to more accurately capture the disease-specific information, leading to eigenvectors that better reflect the underlying disease variation.

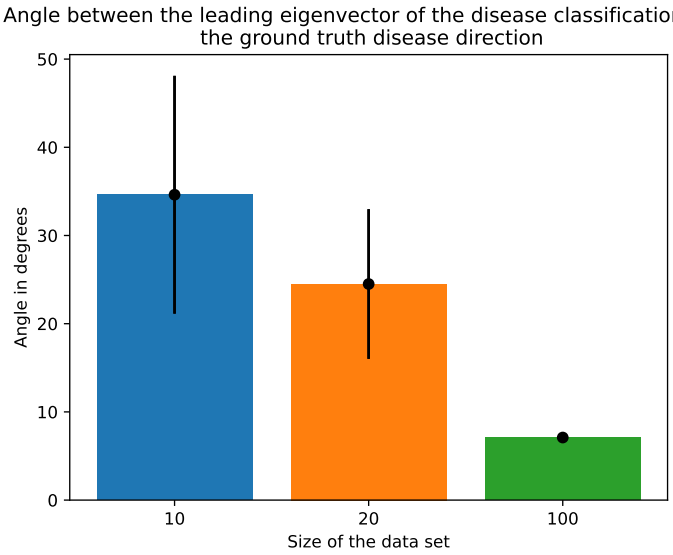

Figure 16: Angle between the leading eigenvector of the disease classification and the ground truth disease direction.

In Figure 17, we present the subspace angle between the set of accumulated eigenvectors of the centre classification and the ground truth centre direction.

Comparing the trajectory in Figure 17 with that in Figure 10, we observe a similar pattern for $n = 100$. However, for the smaller data sets, Figure 17 exhibits a similar pattern but with higher values in degrees.

This suggests that for $n = [10, 20]$, the subspace angle between the accumulated eigenvectors of the centre classification and the ground truth centre direction is generally larger. This indicates that the classification process has more difficulty in accurately capturing the true centre direction when the available data is limited. As a result, the accumulated eigenvectors may not align as closely with the ground truth centre direction compared to the larger data set.

These observations highlight the impact of data size on the accuracy of estimating the centre direction using the GMLVQ algorithm. While the trajectory remains consistent, the angle values in the smaller data sets indicate a reduced ability to capture the ground truth centre direction accurately.

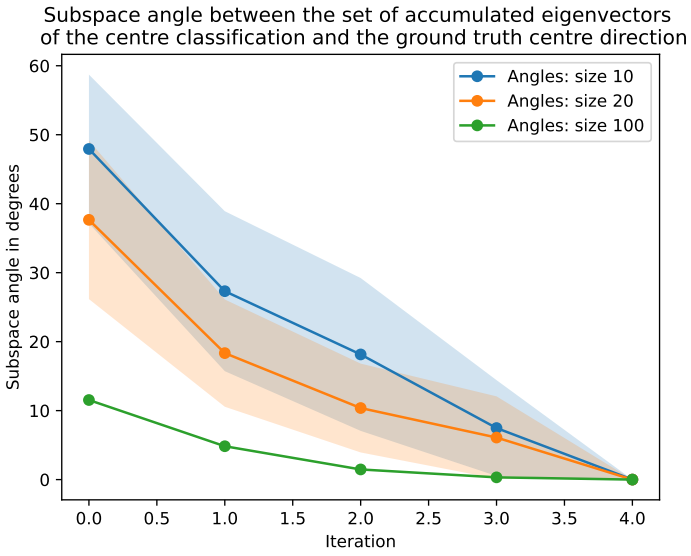

Figure 17: Subspace angle between the set of accumulated eigenvectors of the centre classification and the ground truth centre direction as a function of the number of iterations.

In conclusion, the performances of the smaller two data sets, where $n = [10, 20]$, seem to be notably worse than that in case A, due to the increased complexity. It becomes apparent that for case B, the size of the data set can significantly improve the results.

The removal of $\mathbf{v}_1^{(0)}$ is able to maintain a high accuracy for the disease classification and a close to arbitrary classification of ≈ 0.6 for the centre problem. It can align the leading eigenvector of the disease classification with the ground truth disease direction and the subspace removed contains most of the centre-specific variation with a subspace angle of ≈ 5 degrees. However, this is only achieved with $n = 100$.

For the smaller data sets, we would have to consider the removal up until $\mathbf{v}_1^{(1)}$ or $\mathbf{v}_1^{(2)}$, where $\mathbf{v}_1^{(2)}$ might already be considered as too much. The accuracy still lies above 0.9 but GMLVQ cannot find the direction that describes the main difference close to the ground truth and the subspace angle shows that there is still a significant amount of centre-specific variation in the data.

For this case, we are also able to retrieve a better performance from the centre-wise z-scored data. Regardless of the value of n, the centre classification can be considered arbitrary classification while still maintaining a classification accuracy of ≈ 0.9 for the disease problem. The GMLVQ is able to find the main direction that describes the difference between the diseases that mostly aligns with the ground truth with the biggest angle of ≈ 35 degrees for $n = 10$. This angle approaches an angle close to zero as n increases.

From this case, we can clearly see the difficulty of this method when dealing with smaller data sets. The performance for a big enough data set is more than satisfactory but this significantly degrades as n decreases.

5.3 Case C

This section focuses on the most complex case C. The construction of this case is explained in Section 4.2.

In Figure 18, we present the accuracy trends observed during our analysis. The analysis focuses on the behavior of accuracy in relation to data size, the uncorrected GMLVQ and the removal of $\mathbf{v}_1^{(i)}$ for $i < 4$. Due to the increased complexity, we can observe that the accuracy curves of the classifiers have moved even closer to each other.

Regarding the centre classification, we observe smaller reductions in accuracy with each iteration compared to the previous cases. This suggests that the estimations made by the GMLVQ system in case C are not as accurate in describing the main differences between the centre classes, resulting in the removal of a relatively smaller amount of centre-specific variation. It is also important to note that the level of arbitrary classification is not reached as well as in the previous cases. Close to arbitrary classification for the centre problem is only reached in the third iteration.

For the disease classification, we are observing increased reductions throughout the iterations. Due to the complexity, it seems like every removal of the leading eigenvector also removes disease information, more so than in the previous cases. Notably, in the largest data set with $n = 100$, the removal of $\mathbf{v}_1^{(0)}$ leads to a significant drop of 0.2 in accuracy, bringing the accuracy for the centre classification and the disease classification almost on par with each other. However, the subsequent removals of $\mathbf{v}_1^{(1)}$ and $\mathbf{v}_1^{(2)}$ have a pronounced impact on the centre classification accuracy, while the disease classification accuracy remains relatively stable.

Similar to scenarios A and B, the performance on the training set is poorer for the centre classification. In this case, the disease classification also performs worse, when compared to that on the testing set.

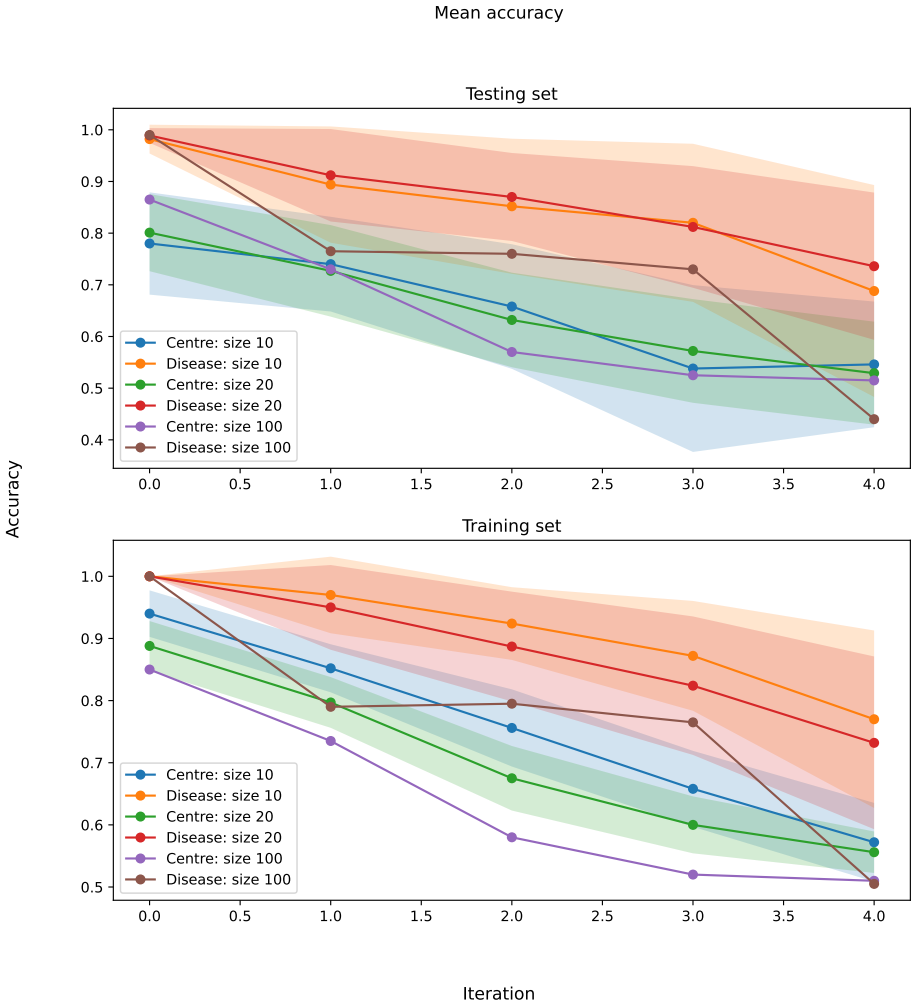

Figure 18: Mean accuracy and standard deviation as a function of the number of iterations of the training set and testing set for case C. The 0-*th iteration* is the uncorrected GMLVQ system.

In Figure 19, we present the accuracy results obtained from the centre-wise z-scored data, using only the uncorrected training processes.

The accuracy analysis reveals that the centre classification accuracy for the testing set remains close to 0.5 for all sizes, accompanied by high accuracy (≈ 0.9) for the disease classification. Even though the centre classification is close to arbitrary, it is slightly higher than in the other cases. However, when employing the iterative method, we achieve this level of performance at the last two iterations for the testing set, which we consider as removing too much information. The accuracy for $n = 100$ in these last two iteration is also less by 0.1.

Furthermore, the metrics obtained from the training sets for the centre classification align with the removal of $\mathbf{v}_1^{(2)}$. The disease classification shows very high accuracies for all n.

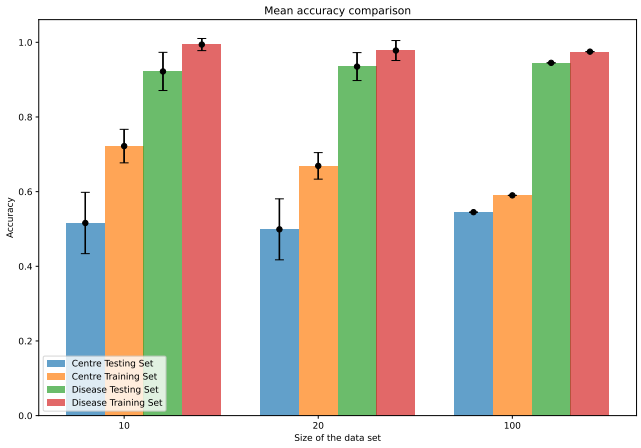

Figure 19: Mean accuracy and standard deviation of the centre-wise z-scored training set and testing set for case C.

With Figure 20, we illustrate the mean squared distance between the means of the centre clusters and the means of the disease clusters. Additionally, it includes the Frobenius norm between the covariance matrices of the clusters.

Upon analysing the results, we observe that after the removal of $\mathbf{v}_1^{(1)}$, the distance between the centre clusters becomes close to zero, indicating overlapping clusters. Furthermore, there is a notable and sudden drop in the squared distance between the means of the disease clusters for all n after the elimination of $\mathbf{v}_1^{(0)}$. This indicates that the elimination of this leading eigenvector has also removed a substantial amount of disease-related information. Due to how similar the differences between the clusters of the classifiers are (see Figure 3), it is possible. Subsequently, the squared distance between the disease cluster means only exhibits a gradual decline in the subsequent iterations.

The curves of the Frobenius norm experience a slight spike after the removal of $\mathbf{v}_1^{(0)}$, which is due to the way the data sets have been constructed with the covariance effect. In the later iterations, it returns to low values.

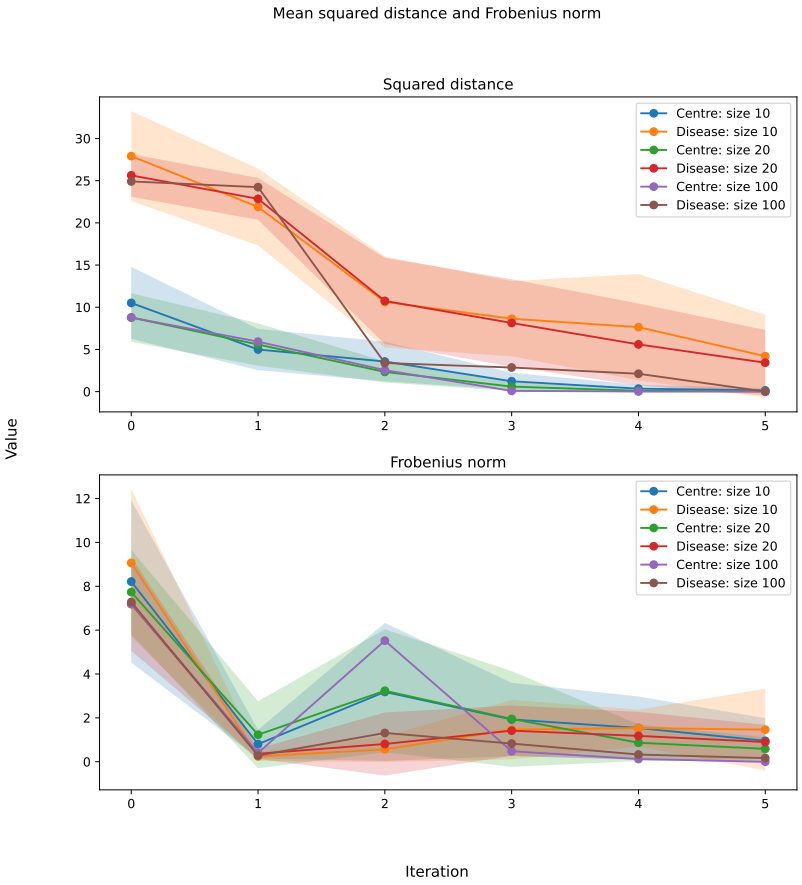

Figure 20: Mean squared distance and Frobenius norm as a function of the number of iterations of the training set. Iteration 0 is the original training set, Iteration 1 correlates to the 0-*th iteration* (the uncorrected GMLVQ) and the remaining iterations represent the metrics after the removal(s) of the leading eigenvector(s).

In Figure 21, we present the same metrics for the centre-wise z-scored data.

The distance between the means of the centres remains consistently low across the different values of n, primarily due to the transformation applied. The squared distances between the means of the disease clusters are decreased compared to Figure 14. The metrics for the Frobenius norm between the covariance matrices show a lower range, but their patterns align with those observed in the iterative approach in Iteration 0 and Iteration 1.

These results are very similar to that of Figure 7 and 14.

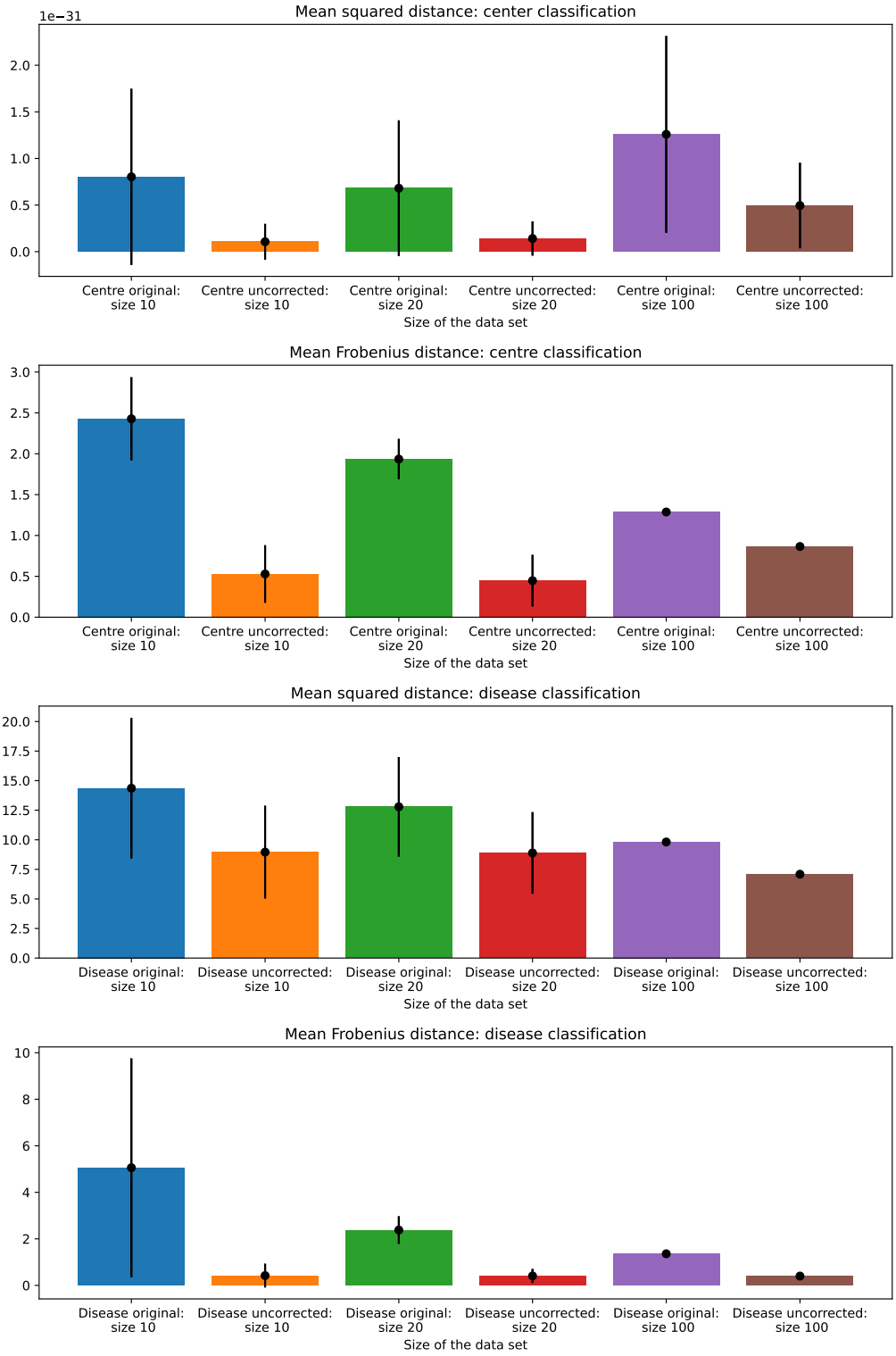

Figure 21: Mean squared distance and Frobenius norm for various sample sizes.

In Figure 22, we examine the angle between the leading eigenvectors of the disease classification and the disease mean offset vector described in Section 4.2.

Compared to 8 and 15, we notice that we do not have an iteration, where the found main direction of GMLVQ mostly aligns with the offset. All angles of each iteration for all n remain above 40 degrees. The removal of $\mathbf{v}_1^{(0)}$ caused a rise for all n, which indicates that this elimination also removed disease-related information. This is most likely caused by the overlap of the differences between the classes for the problems.

Although there is a slight drop in the angle after the removal of $\mathbf{v}_1^{(1)}$ in the smaller data sets compared to the previous iteration, the angle still remains around ≈ 50 degrees. Additionally, it is worth noting that the main directions found for $n = 100$ result in an increased angle to the disease mean offset vector over the iterations. These observations further confirm that case C does not yield satisfactory results, even with a large data set.

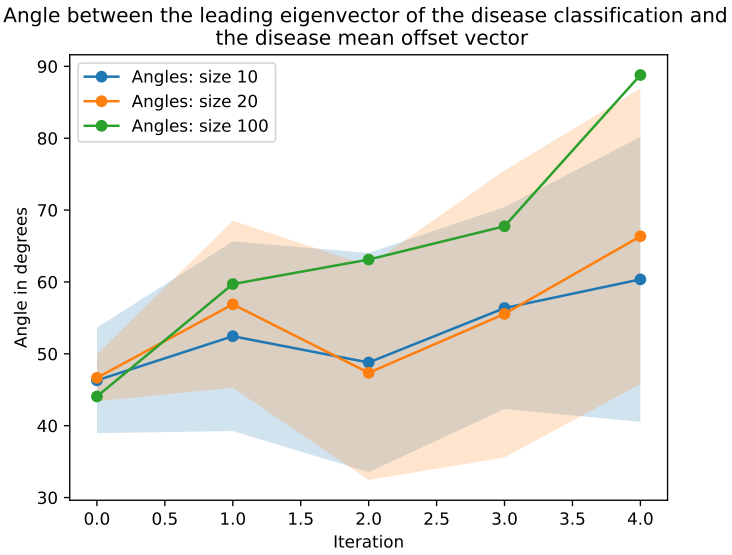

Figure 22: Angle between the leading eigenvector of the disease classification and the disease mean offset vector as a function of the number of iterations. The 0-*th iteration* is the uncorrected GMLVQ system.

In Figure 23, we present the same metrics for the centre-wise z-scored data.

Comparing the results in Figure 22 and Figure 23, we observe significant differences in the angles between the leading eigenvectors and the disease mean offset vector. For $n = 10$, the GMLVQ system using the centre-wise z-scored data finds a leading eigenvector with an angle of approximately 40 degrees the disease mean offset vector. In contrast, for the iterative method, all the angles remain above 40 degrees.

It is worth noting that the angles in case C are generally larger compared to the previous cases. For example, while the angles for $n = 100$ in the previous cases were below ten degrees, in case C, the angles are around 30 degrees.

These findings suggest that the use of the centre-wise z-scored data improves the alignment between the leading eigenvectors and the disease mean offset vector to some extent, particularly for simpler data sets. However, even with this improvement, the angles in case C remain relatively large, indicating the challenge in accurately capturing the disease-specific information for this data set.

Note that for case C, the interpretation of the ground truth is more complicated, which is why we explicitly state the previous in terms of the disease mean offset vector. The construction of this case is described in Section 4.2.

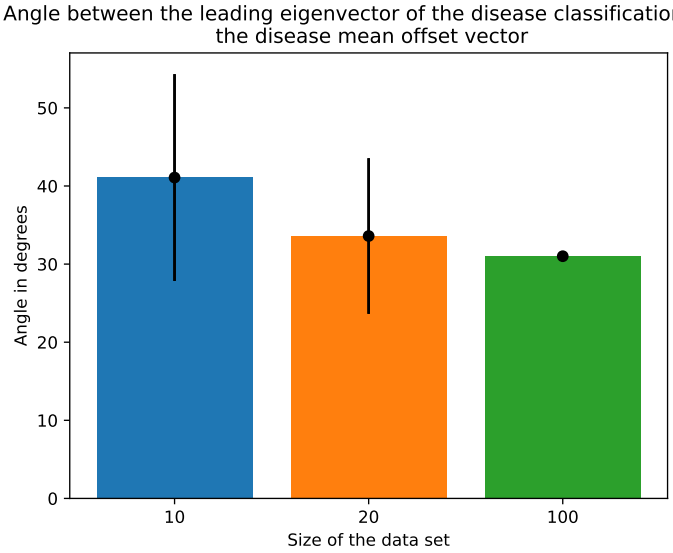

Figure 23: Angle between the leading eigenvector of the disease classification and the ground truth disease direction.

Figure 24 shows the subspace angle between the set of accumulated eigenvectors of the centre classification and the ground truth centre direction.

Comparing the trajectory in Figure 24 with those in Figures 10 and 17, we observe a similar pattern for the largest data set. However, for $n = [10, 20]$, Figure 24 demonstrates a similar pattern but with higher values in degrees. The smaller data sets seem to only reach an angle around or below 10 in the last two iterations, which are considered as removing too much. For $n = 100$, we reach an angle below 10 after disregarding $\mathbf{v}_i^{(0)}$, however, based on the previous figures, this subspace likely also defines the main direction of the disease difference. This indicates that removing this subspace may lead to a loss of disease-specific information.

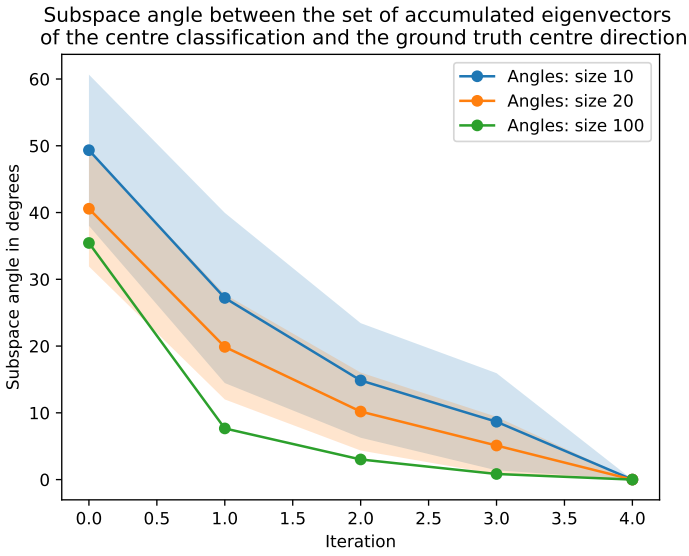

Figure 24: Subspace angle between the set of accumulated eigenvectors of the centre classification and the ground truth centre direction as a function of the number of iterations.

In conclusion, case C presents the most complex scenario among the three cases analysed. The analysis of accuracy trends reveals that the performance of the classifiers is impacted by the increased complexity. There are smaller decreases for the centre classification and bigger decreases for the disease classification as we remove the leading eigenvectors. The GMLVQ systems are unable to find a main difference between the diseases that align with the disease mean offset vector, however, the subspace does seem to be able to define the ground truth centre direction for $n = 100$ nonetheless. Finding a happy medium between eliminating centre-specific information and maintaining disease-specific information is difficult due to the similar nature of the variances between the classes.

Interestingly, the iterative method does not align the leading eigenvectors of the disease classification with the disease mean offset vector as effectively as in the previous cases. The ground truth is defined in a more complex manner for this case. Nonetheless, this could be attributed to the similarity of the differences between the centre and disease clusters, making it difficult to separate and identify the disease-specific patterns accurately.

In contrast, the centre-wise z-scored data approach demonstrates a more promising performance, even though the estimated main direction of the disease classification does not entirely align with the disease mean offset vector. This method shows better adaptability to the increased complexity, maintaining high classification accuracy for the disease problem and an arbitrary level of classification for the centre problem.

These findings highlight the challenges faced in case C, particularly in finding a balance between removing centre-specific information and preserving disease-specific information when the differences are seemingly overlapping.

6 Conclusion

To achieve the main objective of constructing a universal classifier that is able to distinguish between neurodegenerative diseases based on FDG-PET neuroimages from patients regardless of the neuroimaging centre of origin, we propose the iterative subspace correction procedure with this thesis.

In conclusion, our experiments investigated the application of the iterative method and centre-wise z-scored data in the context of GMLVQ for the classification of centre and disease problems. We explored three different cases: A, B, and C, of increasing complexity.

In case A, we observed that the iterative method successfully aligned the subspace with the ground truth centre direction, resulting in high accuracy for the disease classifications and a near arbitrary classification for the centre problem. The removal of leading eigenvectors effectively reduced centre-specific information, while maintaining disease classification performance.

In case B, which presented increased complexity, the iterative method demonstrated continued success in aligning the subspace with the ground truth centre direction, although with smaller reductions in accuracy for the centre classification. The disease classification accuracy remained high, albeit with larger decreases in accuracy after each removal. In this case, the importance of the number of data points we consider became apparent due to the improved performance of $n = 100$ compared to the smaller data sets.

Case C proved to be the most challenging scenario, with similar differences between the centre and disease classes. The iterative method faced difficulties in accurately identifying the main differences while not removing disease-related information, resulting in smaller reductions in centre classification accuracy and larger reductions in disease classification accuracy. The centre-wise z-scored data approach showed better performance in this complex case, maintaining high accuracy for the disease classifications while also achieving an arbitrary classification level for the centre problem. However, even with this data, the main direction estimated by the GMLVQ systems does not seem to align with the disease mean offset vector.

Overall, the experiments highlighted the importance of the amount of data points we consider and the complexity of the data set in GMLVQ classification. Larger data sets generally led to improved performance, while increased complexity posed challenges in accurately capturing the underlying differences. The iterative method demonstrated its ability to align the subspace with the ground truth direction, but its effectiveness varied across the different cases. The centre-wise z-scored data approach showed a higher adaptability to complexity, achieving high classification accuracy for the disease problem and ensuring arbitrary classification for the centre problem.

7 Future Work

Based on what has been done in this thesis, further experiments can be done to observe and analyse the behavior of this iterative method. For this purpose, more complex artificial data with different parameters could be constructed. It would also be interesting to see how this method performs with more than 5 features. The artificial data could also be defined with more relevant directions for either classification to investigate whether the method is able to find most, if not all of them.

Once there is a better understanding of this method, it could be applied to real neuroimaging data from different neuroimaging centres. As mentioned in the introduction of this thesis, the objective is to build a universal classifier that is able to distinguish between neurodegenerative disease regardless of the centre of origin. To achieve this, application to real neuroimaging data is inevitable. From this real-life application, more behavioral patterns can be discovered and possibly incite modifications or corrections.

In many other fields, data from different sources is used, which face similar issues [18, 19, 20]. Ideally, this method should not only be applicable to neuroimaging data but also to the data in these other fields that face the same problems. This way, we could enable the combination of data from different origins to further enhance the accuracy of the respective classifiers in these fields.

The construction of a cost function that takes both classifiers into account, similar to [13], in combination with the procedure described by IRMA could pose as a new research direction. Currently, the classifiers are completely disjoined and not aware of the other's existence. With this newly proposed structure, one could incorporate a certain degree of communication between the processes and try to prevent the removal of the disease information when removing the leading eigenvector(s) of the centre classification.

Bibliography

[1] H. Lei, Y. Zhao, Y. Wen, and B. Lei, "Adaptive sparse learning for neurodegenerative disease classification," in *2017 IEEE International Symposium on Multimedia (ISM)*, pp. 292–295, 2017.

[2] R. van Veen, *Learning Vector Quantization with Applications in Neuroimaging and Biomedicine*. PhD thesis, University of Groningen, 2022.

[3] L. Mosley, "A balanced approach to the multi-class imbalance problem," 2013.

[4] J. Kelleher, B. Namee, and A. D'Arcy, *Fundamentals of Machine Learning for Predictive Data Analytics, second edition: Algorithms, Worked Examples, and Case Studies*. MIT Press, 2020.

[5] F. Turkheimer, M. Veronese, and J. Dunn, *Experimental Design and Practical Data Analysis in Positron Emission Tomography*. 04 2014.

[6] T. Kohonen, "The self-organizing map," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464–1480, 1990.

[7] A. Sato and K. Yamada, "Generalized learning vector quantization," *Advances in neural information processing systems*, vol. 8, 1995.

[8] P. Schneider, M. Biehl, and B. Hammer, "Adaptive relevance matrices in learning vector quantization," *Neural Computation*, vol. 21, no. 12, pp. 3532–3561, 2009.

[9] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2021.

[10] S. Saralajew and T. Villmann, "Transfer learning in classification based on manifolc. models and its relation to tangent metric learning," in *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 1756–1765, 2017.

[11] K. Crammer, M. Kearns, and J. Wortman, "Learning from multiple sources," *Journal of Machine Learning Research*, vol. 9, no. 57, pp. 1757–1774, 2008.

[12] Z. Ding, M. Shao, and Y. Fu, "Transfer learning for image classification with incomplete multiple sources," in *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 2188–2195, 2016.

[13] T. Villmann, D. Staps, J. Ravichandran, S. Saralajew, M. Biehl, and M. Kaden, "A learning vector quantization architecture for transfer learning based classification in case of multiple sources by means of null-space evaluation," in *Advances in Intelligent Data Analysis XX - 20th International Symposium on Intelligent Data Analysis, IDA 2022, Proceedings*, pp. 354–364, Springer Science and Business Media Deutschland GmbH, 2022.

[14] D. Mudali, M. Biehl, S. Meles, R. Renken, D. García-García, P. Clavero, J. Arbizu, J. Obeso, M. Rodriguez-Oroz, K. Leenders, and J. Roerdink, "Differentiating early and late stage parkinson's disease patients from healthy controls," *Journal of Biomedical Engineering and Medical Imaging*, vol. 3, pp. 33–43, Dec. 2016. Accepted.

[15] S. Lövdal and M. Biehl, "Improved interpretation of feature relevances: Iterated relevance matrix analysis (irma)," Submitted 2023.

[16] R. van Veen, M. Biehl, and G.-J. de Vries, "sklvq: Scikit learning vector quantization," *Journal of Machine Learning Research*, vol. 22, no. 231, pp. 1–6, 2021.

[17] G. Papari, K. Bunte, and M. Biehl, "Waypoint averaging and step size control in learning by gradient descent (technical report)," in *MIWOCI 2011, Mittweida Workshop on Computational Intelligence* (F.-M. Schleif and T. Villmann, eds.), vol. MLR-2011-06 of *Machine Learning Reports*, pp. 16–26, Univ. of Bielefeld, 2011.

[18] K. Crammer, M. Kearns, and J. Wortman, "Learning from multiple sources," *Journal of Machine Learning Research*, vol. 9, no. 57, pp. 1757–1774, 2008.

[19] K. Bousmalis, G. Trigeorgis, N. Silberman, D. Krishnan, and D. Erhan, "Domain separation networks," *CoRR*, vol. abs/1608.06019, 2016.

[20] J.-C. Tsai and J.-T. Chien, "Adversarial domain separation and adaptation," in *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6, 2017.

Appendices

A Overview of cluster parameters

Table 1: Overview of cluster parameters

Cluster	Mean	Covariance matrix
Case A		
HC1	$[-0.04, -0.03]$	$[[0.87, -0.08], [-0.08, 1.09]]$
HC2	$[-0.04, 2.97]$	$[[0.87. - 0.08], [-0.08, 1.09]]$
PD1	$[1.98, 0.05]$	$[[0.83, 0.1], [0.1, 1.07]]$
AD2	$[-2.05, 3.03]$	$[[0.96, 0.01], [0.01, 0.95]]$
Case B		
HC1	$[-0.04, -0.03]$	$[[0.87, -0.08], [-0.08, 1.09]]$
HC2	$[-0.07, 2.94]$	$[[3.48, -0.31], [-0.31, 4.35]]$
PD1	$[1.98, 0.05]$	$[[0.83, 0.1], [0.1, 1.07]]$
AD2	$[-2.1, 3.07]$	$[[3.84, 0.05], [0.05, 3.8]]$
Case C		
HC1	$[-0.04, -0.03]$	$[[0.87, -0.08], [-0.08, 1.09]]$
HC2	$[-0.1, 2.9]$	$[[4.25, 3.52], [3.52, 4.9]]$
PD1	$[1.98, 0.05]$	$[[0.83, 0.1], [0.1, 1.07]]$
AD2	$[-2.07, 3.02]$	$[[4.85, 3.89], [3.89, 4.82]]$

B Case A

In this appendix several plots supplementing the results in Section 5.1 are presented, including the mean relevance profiles, eigenvalue spectrum of Lambda over the iterations and the difference between the data set mean and ground truth mean to show the stability of the data sets used for the experiments.

B.1 Results for the centre classification

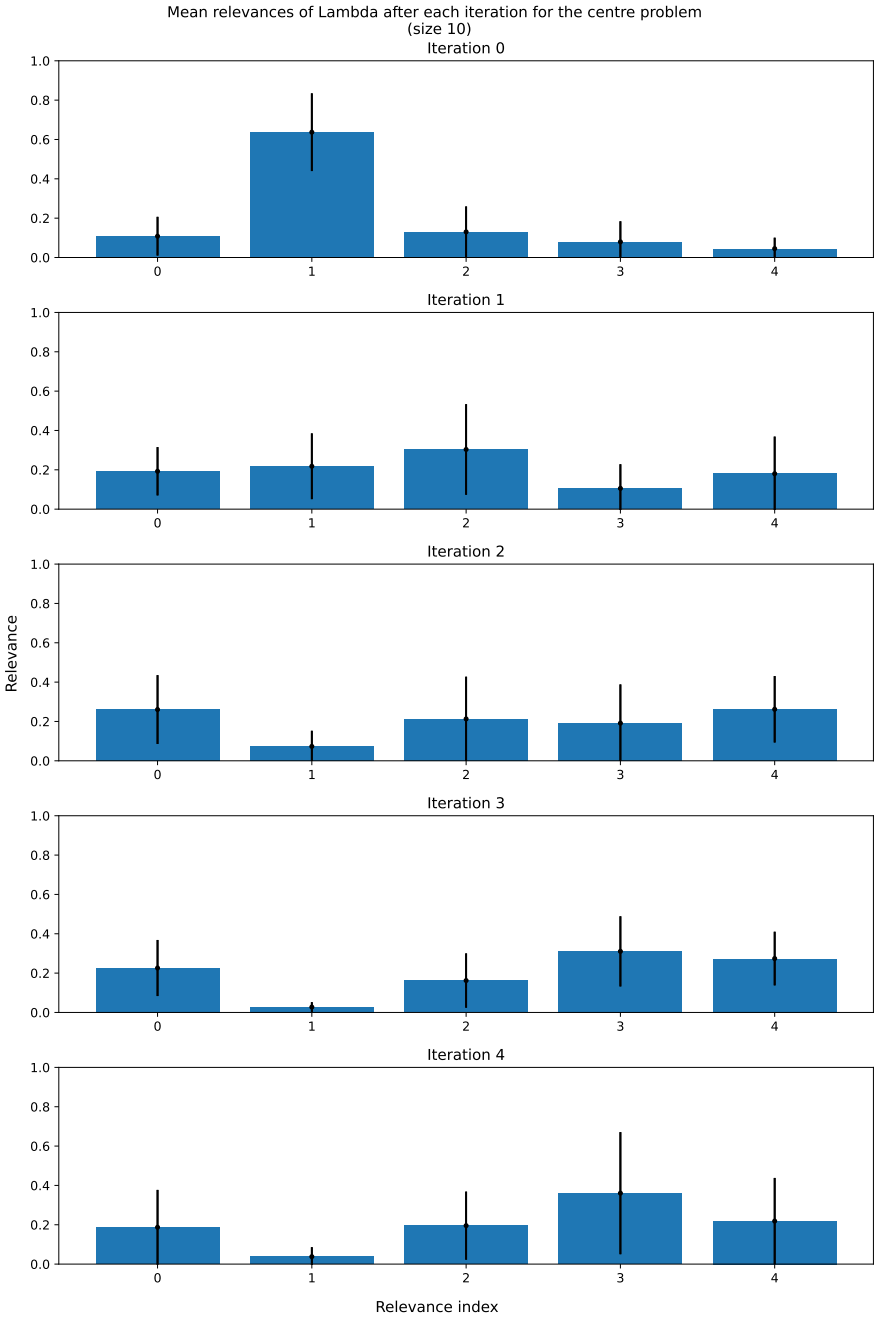

Figure 25: Mean relevance profile of case A with $n = 10$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

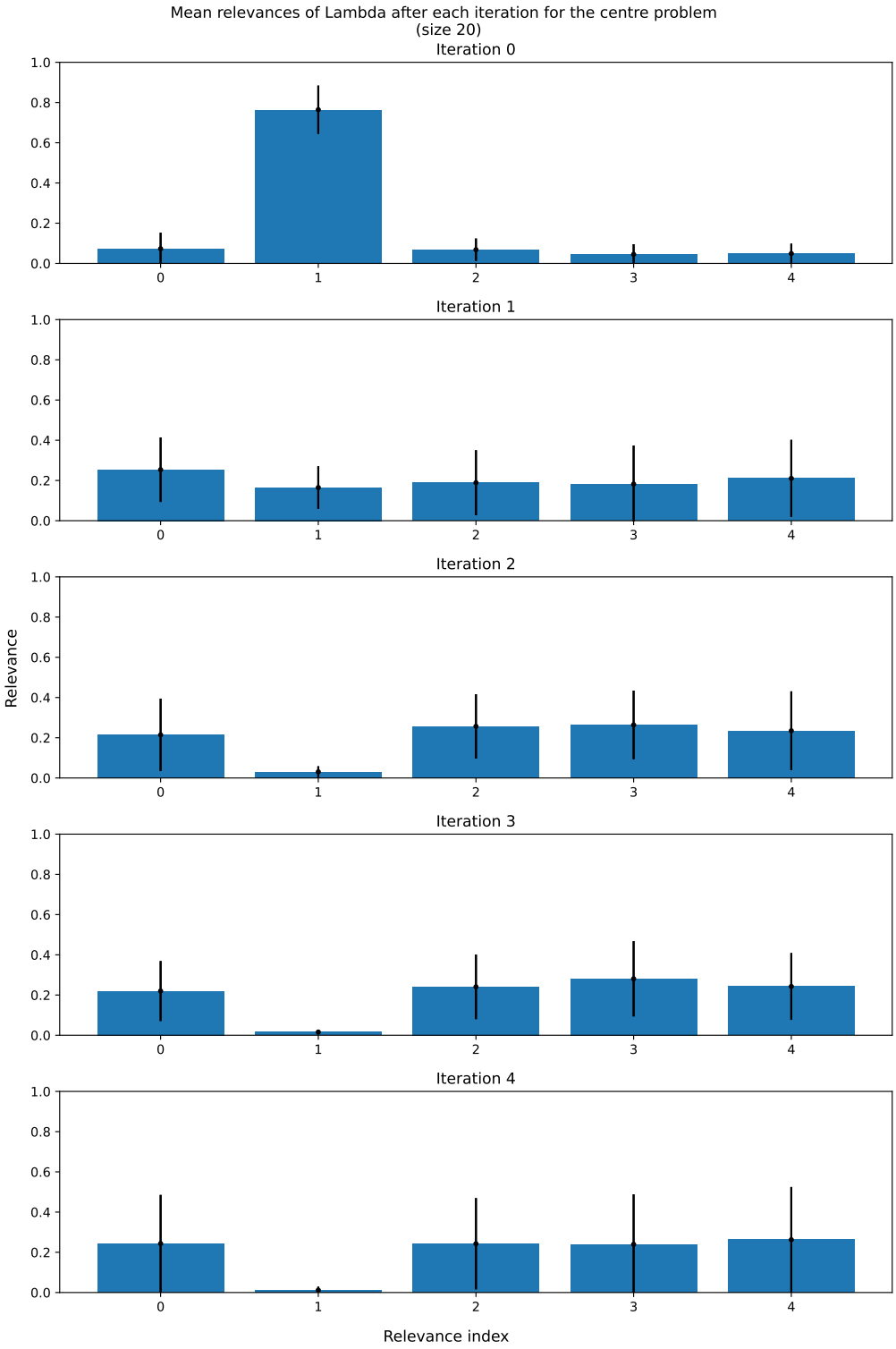

Figure 26: Mean relevance profile of case A with $n = 20$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

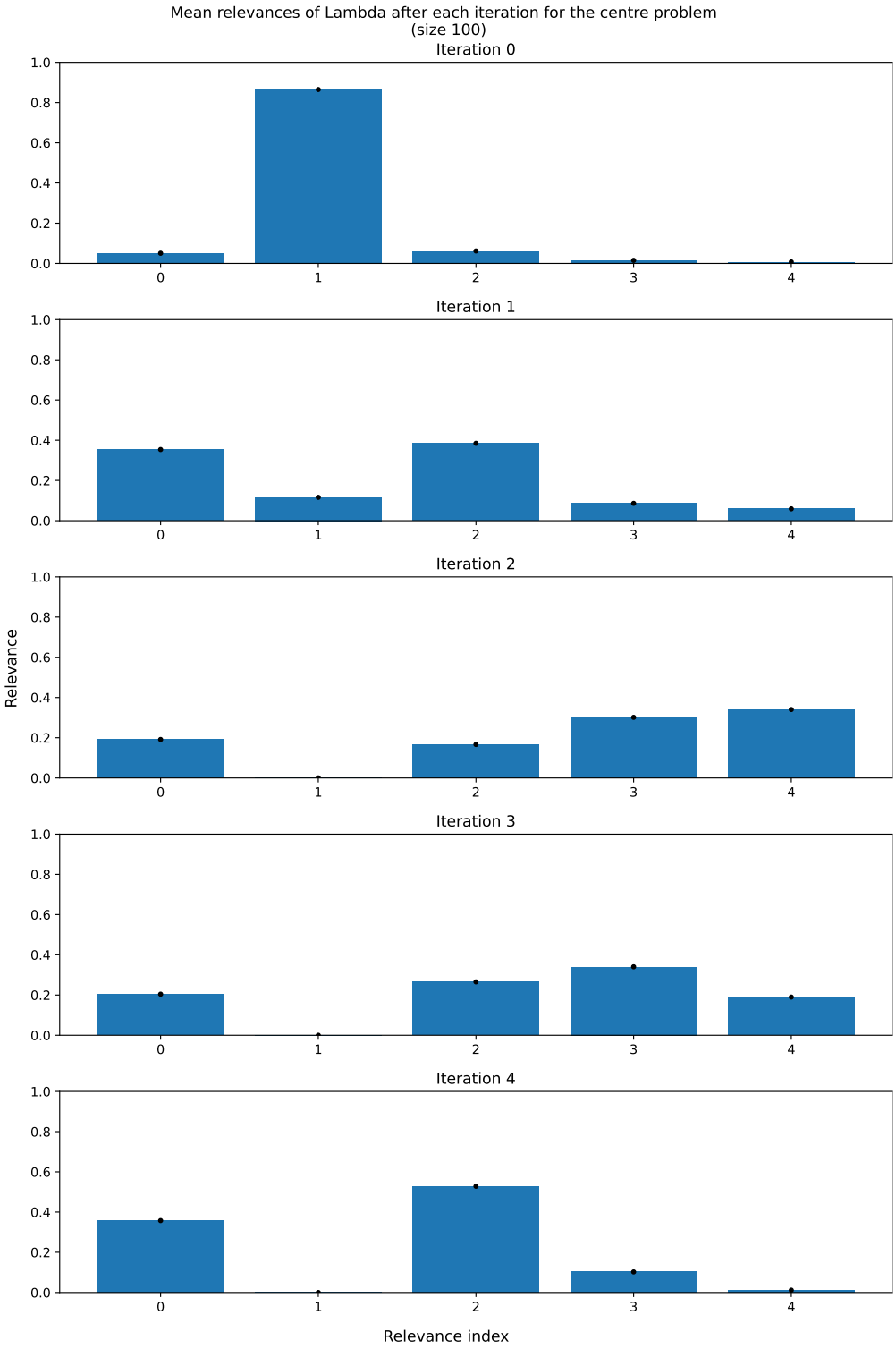

Figure 27: Mean relevance profile of case A with $n = 100$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

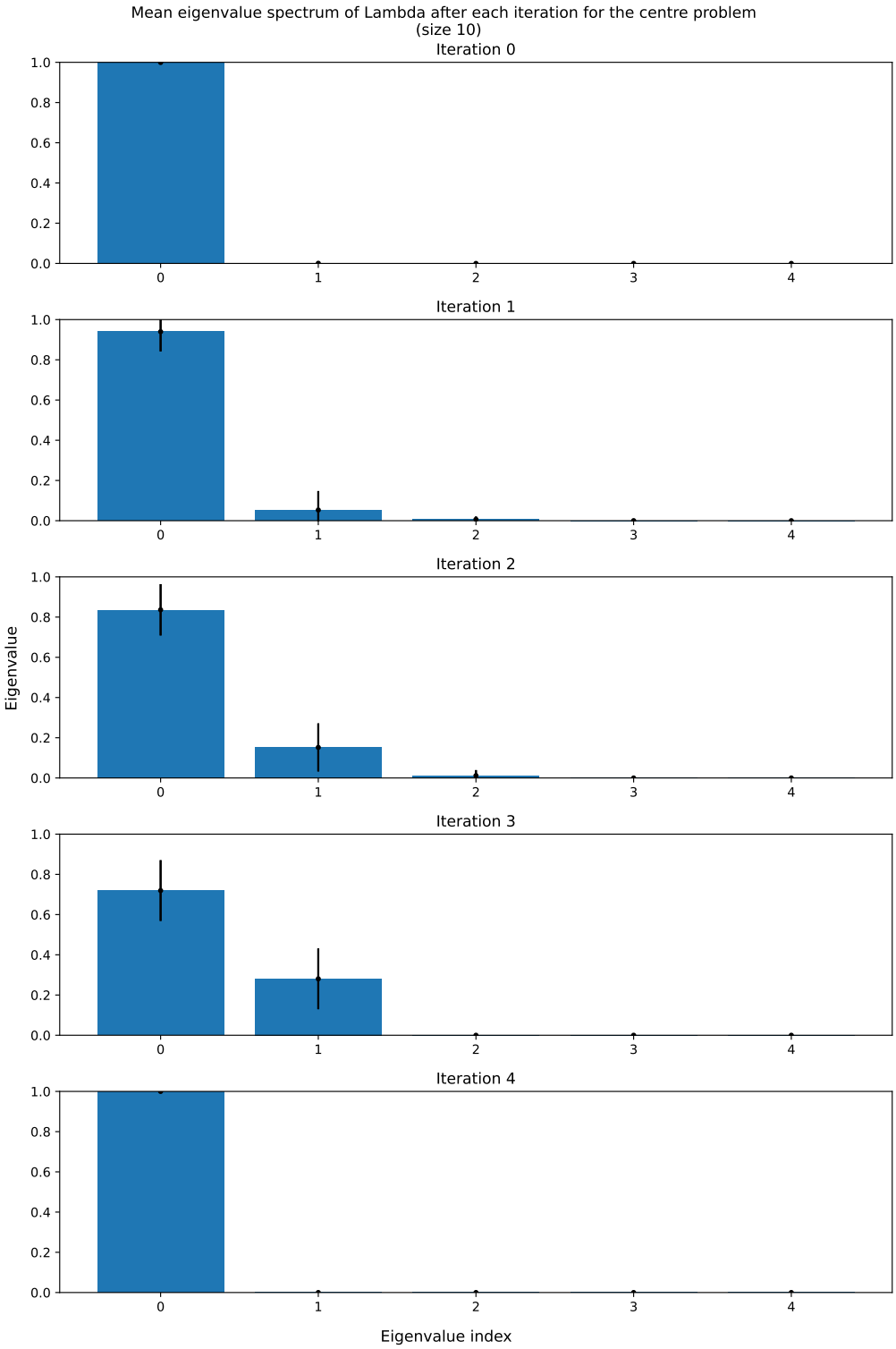

Figure 28: Mean eigenvalue spectrum of Lambda of case A with $n = 10$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

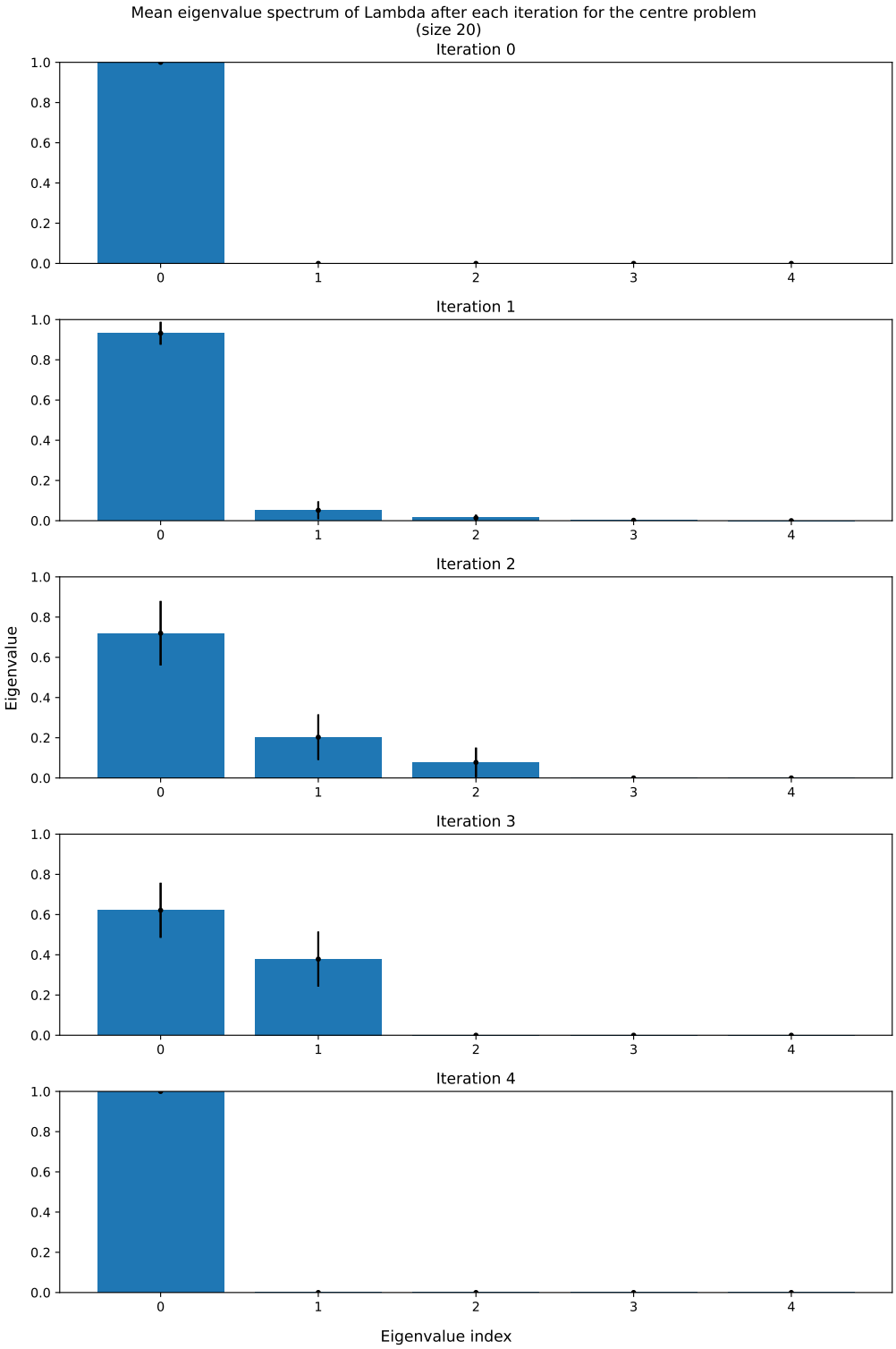

Figure 29: Mean eigenvalue spectrum of Lambda of case A with $n = 20$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

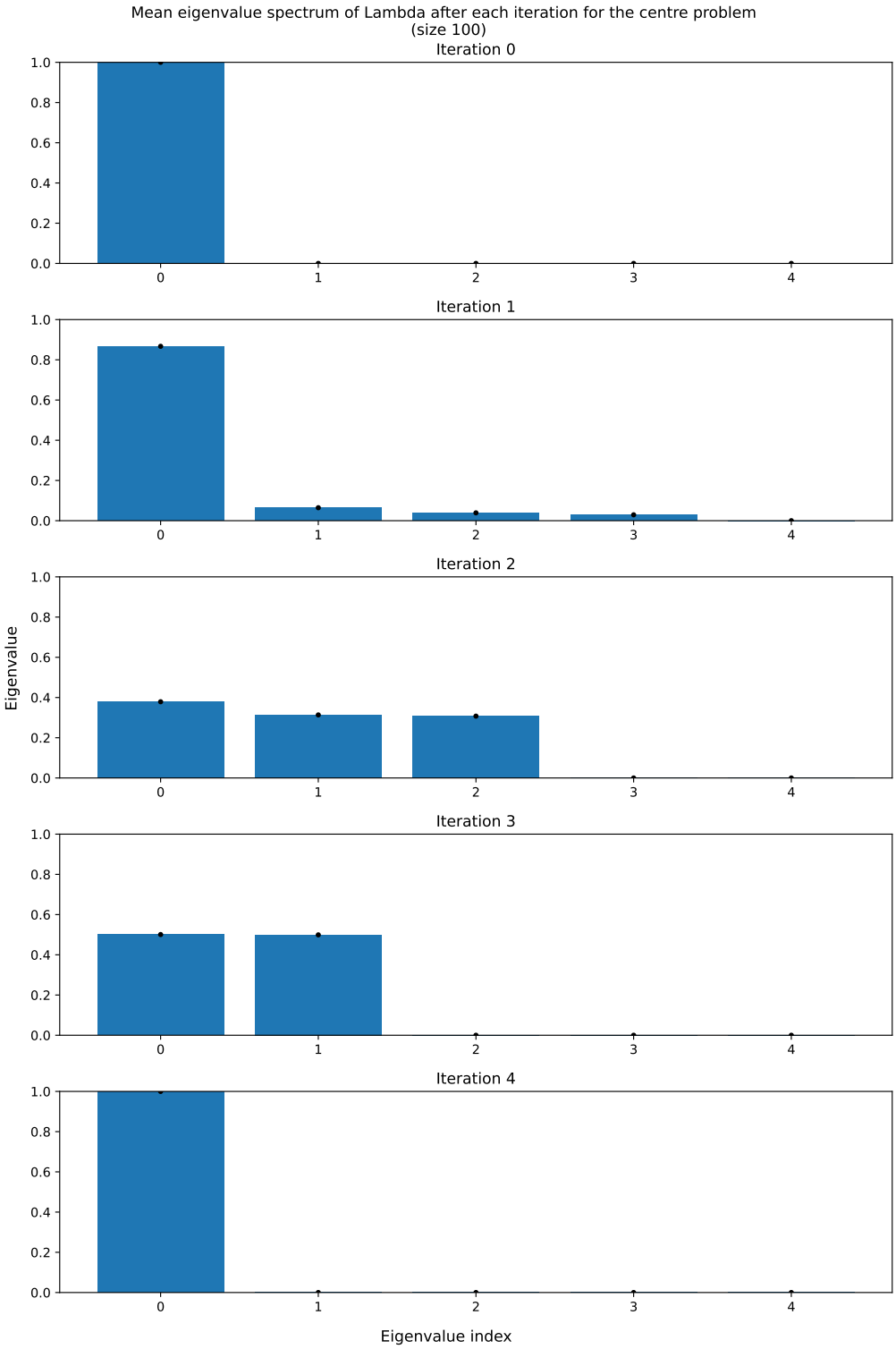

Figure 30: Mean eigenvalue spectrum of Lambda of case A with $n = 100$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

B.2 Results for the disease classification

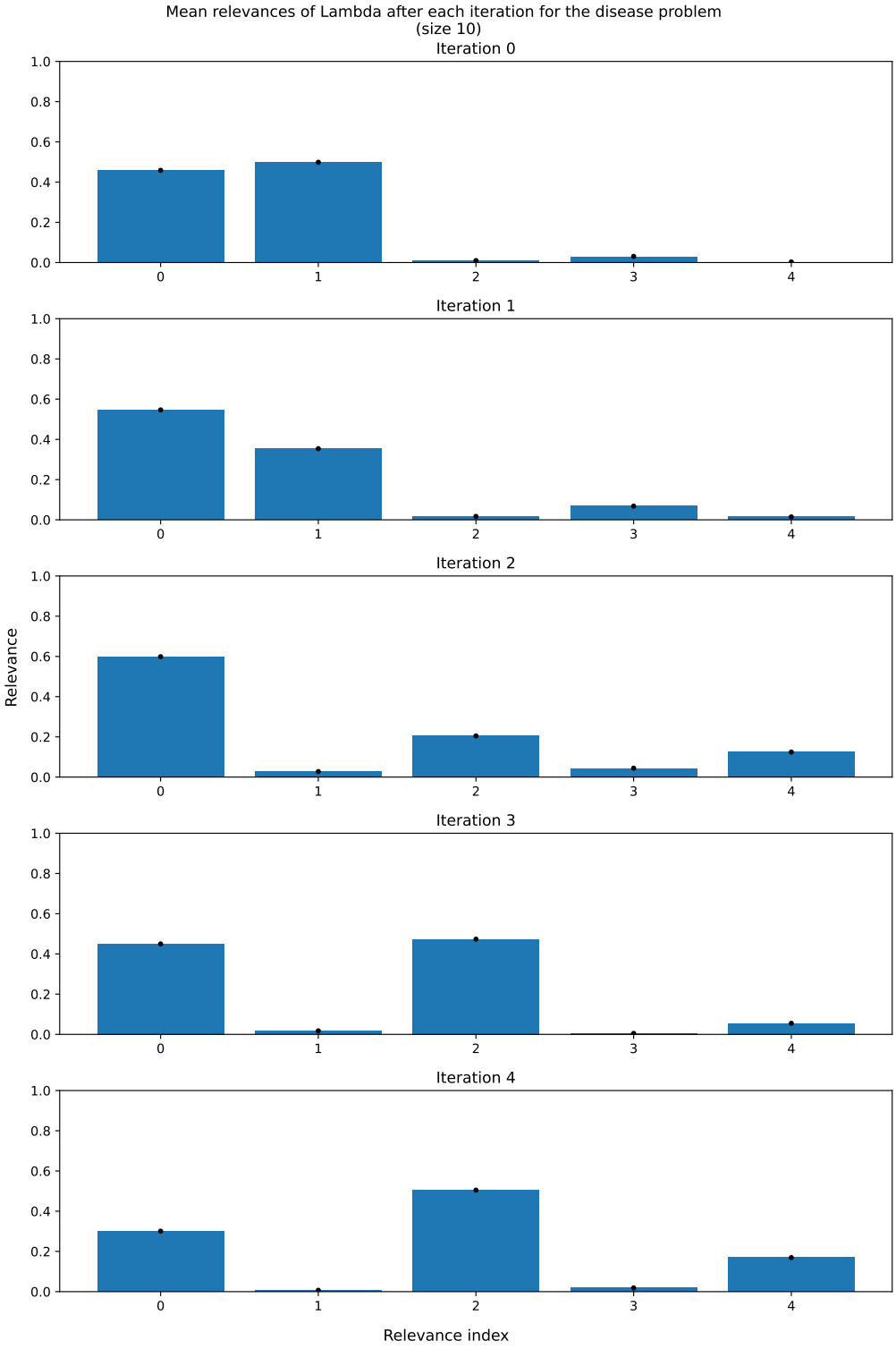

Figure 31: Mean relevance profile of case A with $n = 10$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

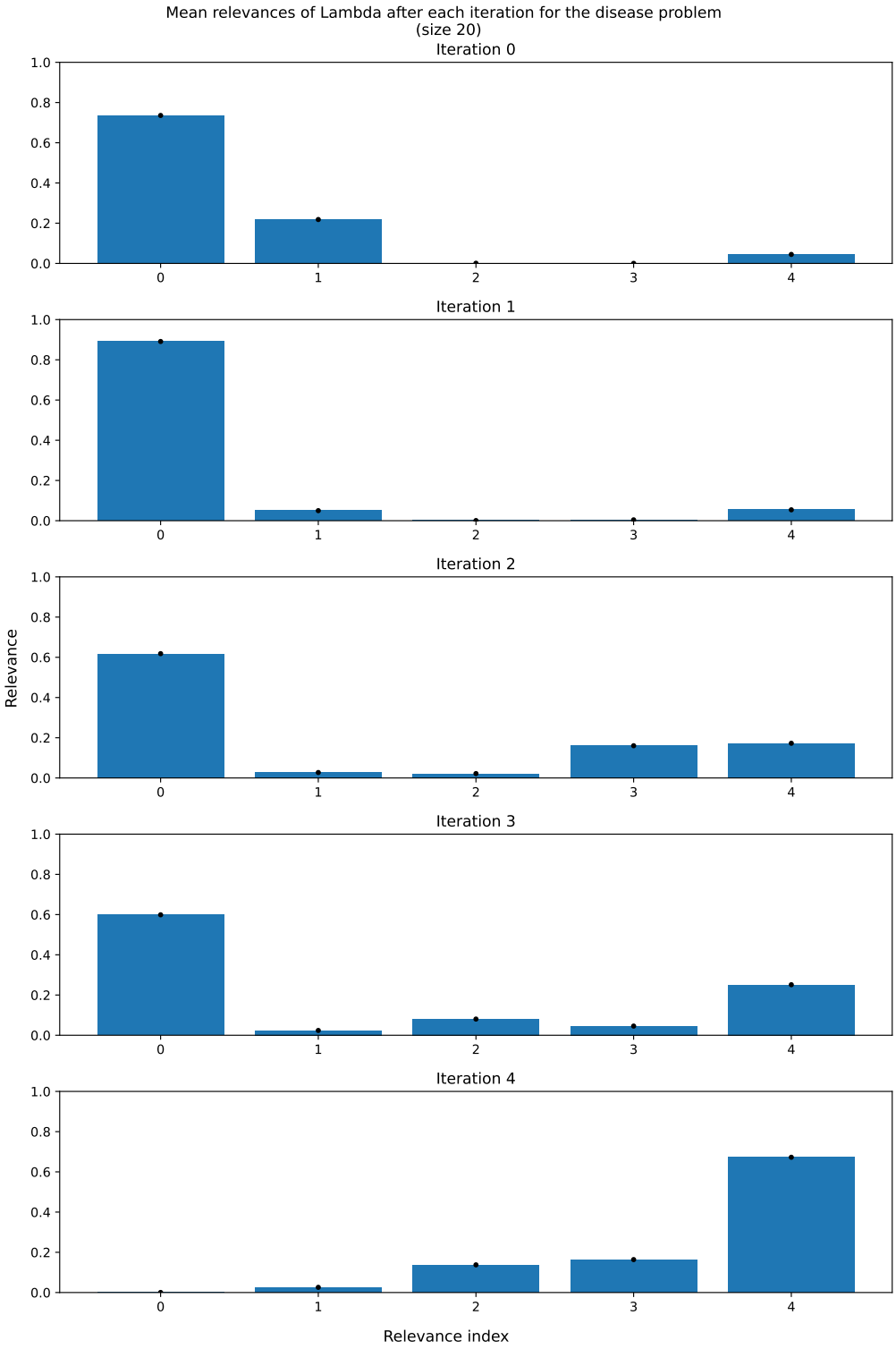

Figure 32: Mean relevance profile of case A with $n = 20$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

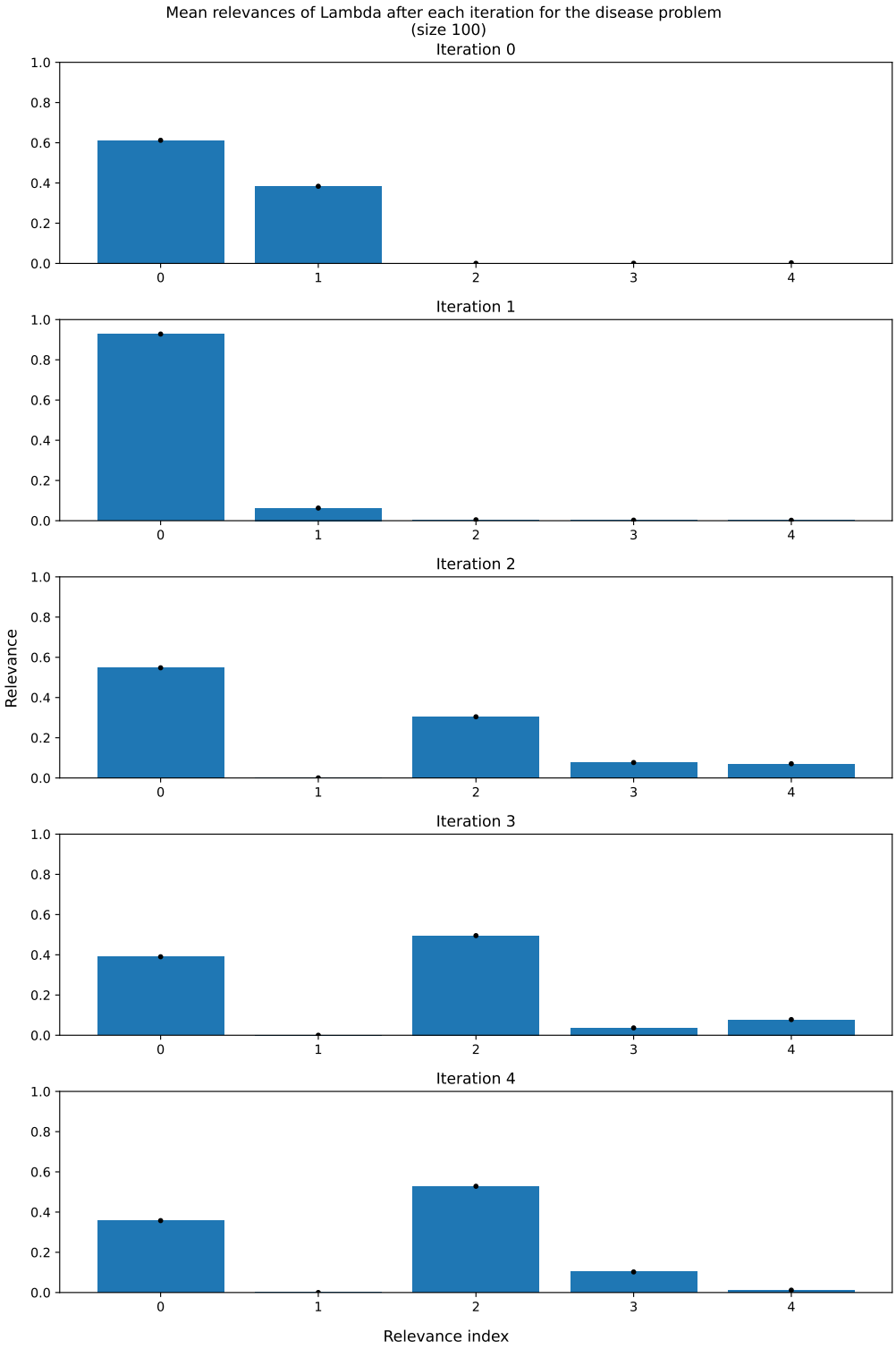

Figure 33: Mean relevance profile of case A with $n = 100$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

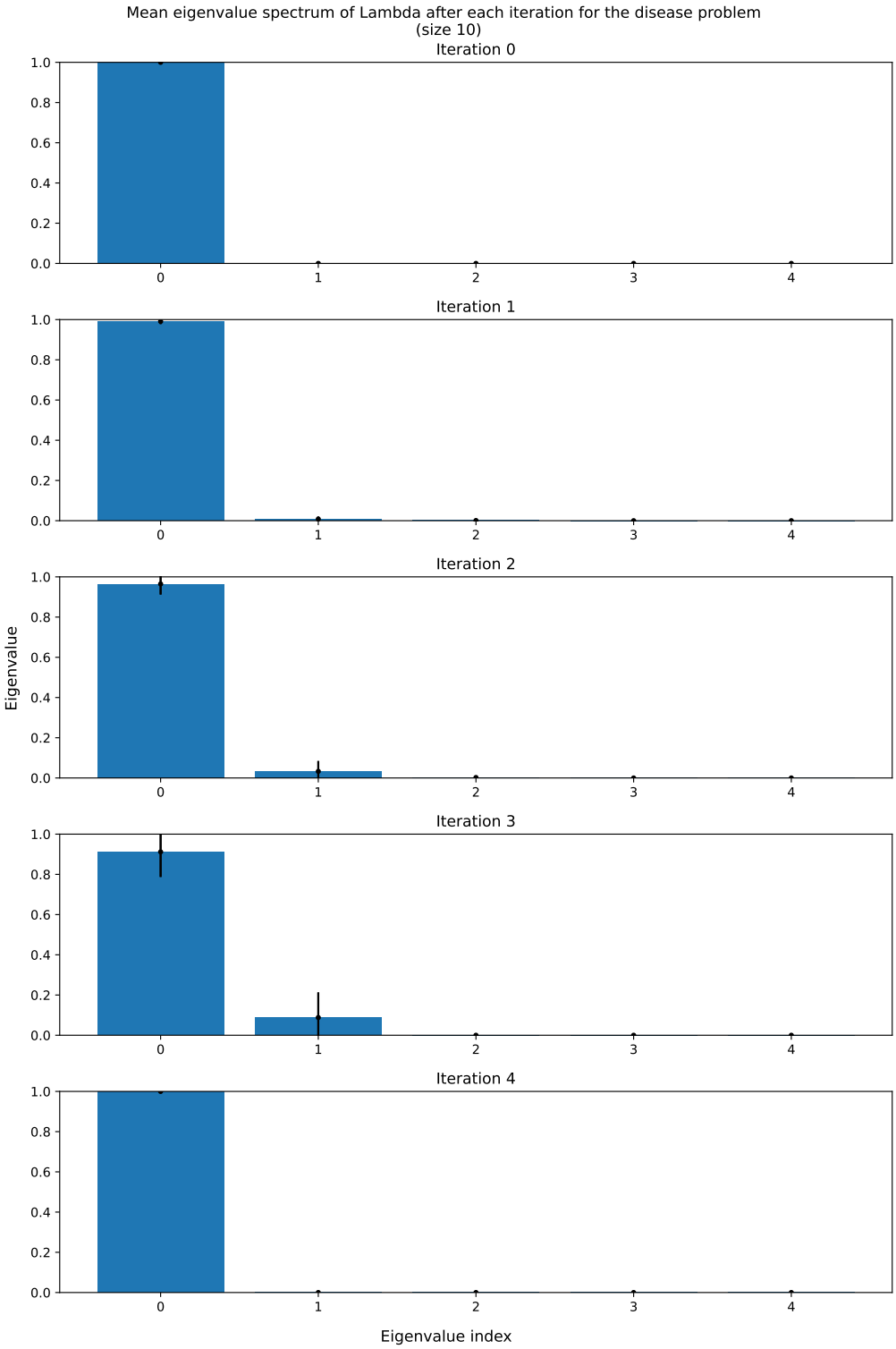

Figure 34: Mean eigenvalue spectrum of Lambda of case A with $n = 10$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

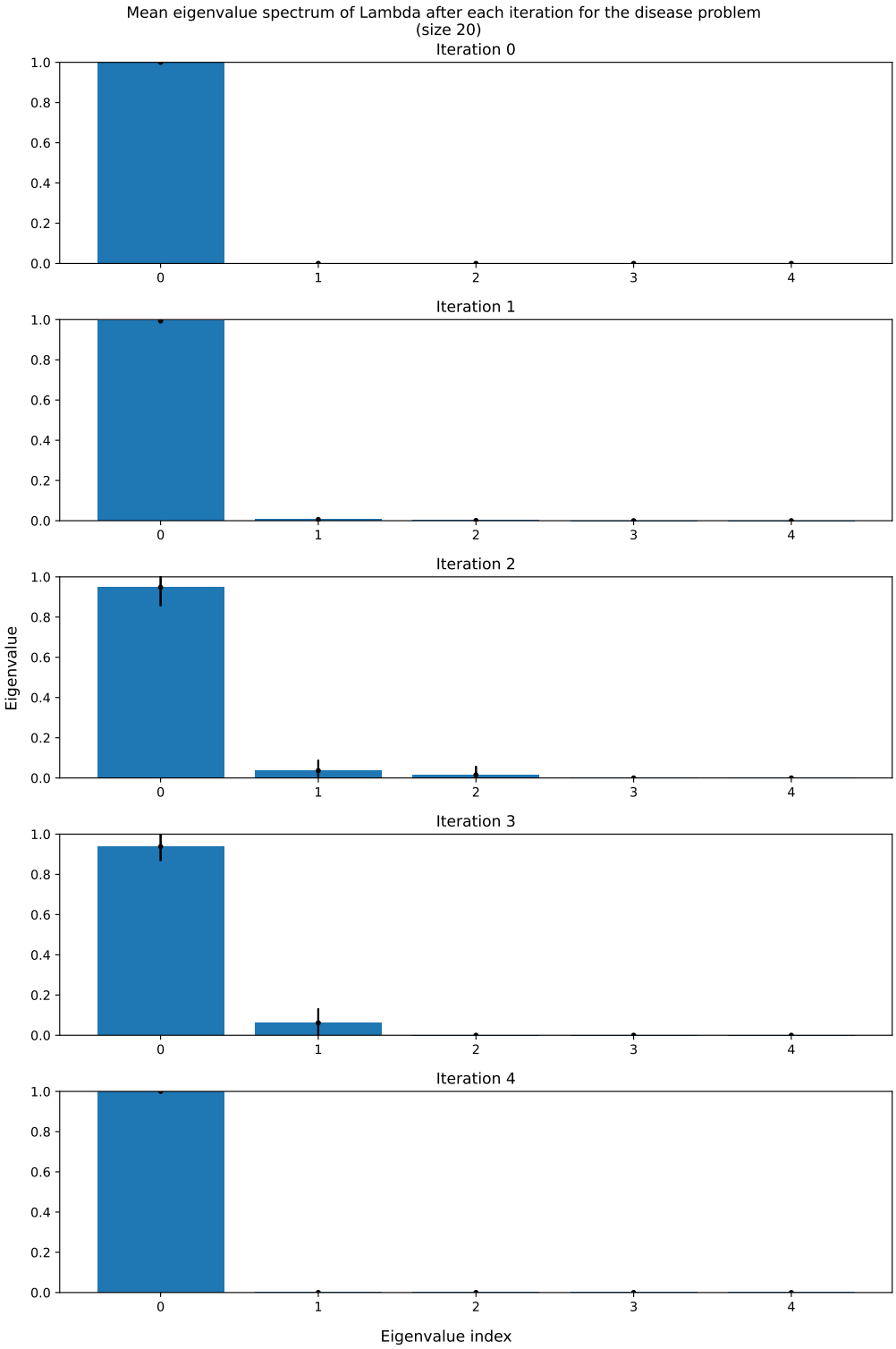

Figure 35: Mean eigenvalue spectrum of Lambda of case A with $n = 20$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

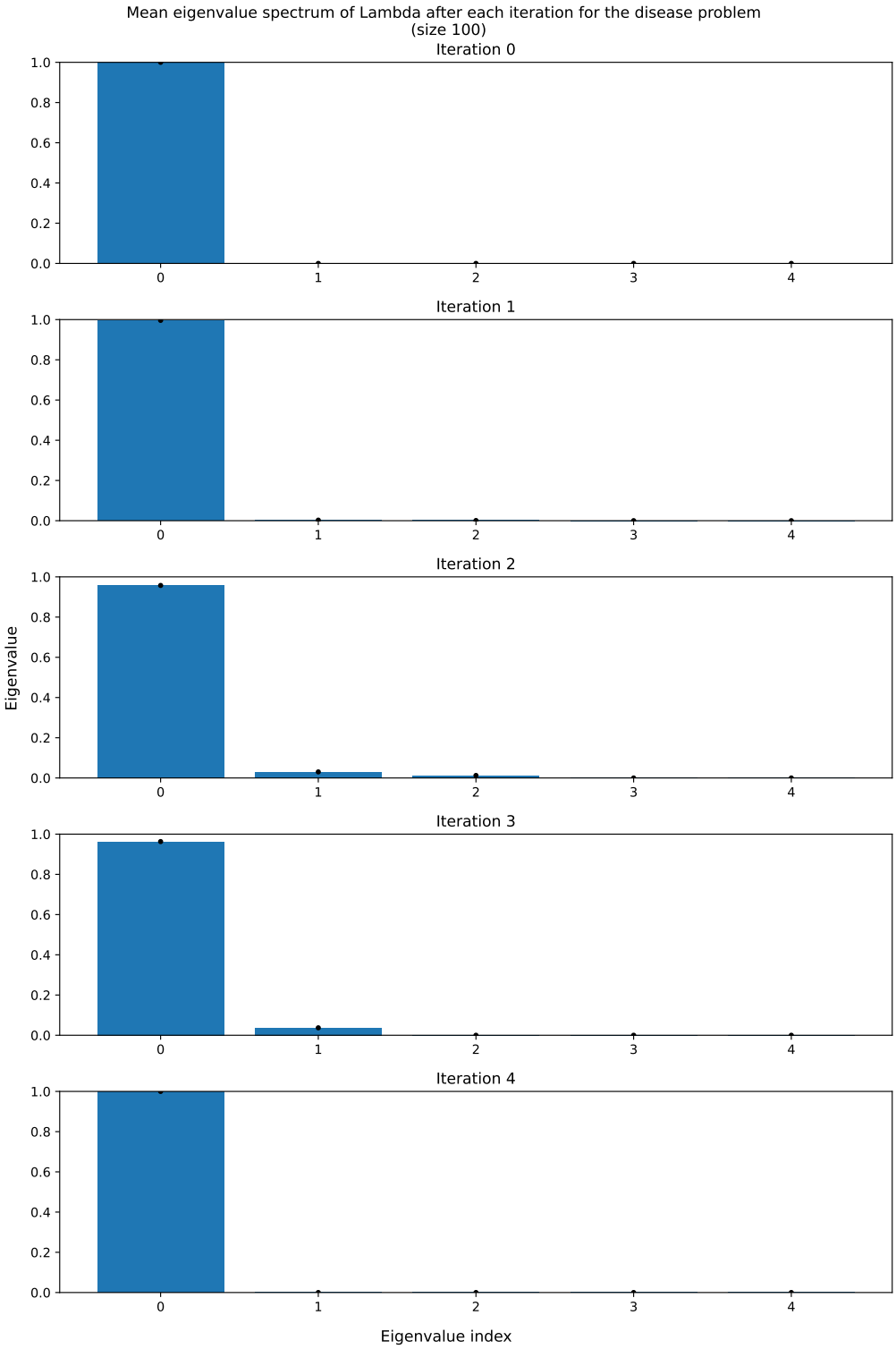

Figure 36: Mean eigenvalue spectrum of Lambda of case A with $n = 100$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

B.3 Centre-wise z-score transformed data: results for the centre classification

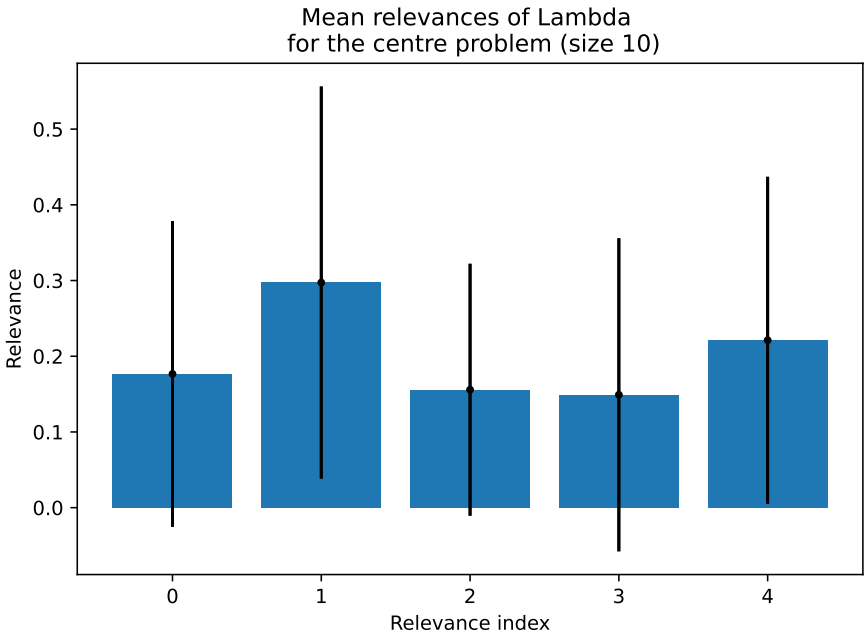

Figure 37: Mean relevance profile of case A with $n = 10$.

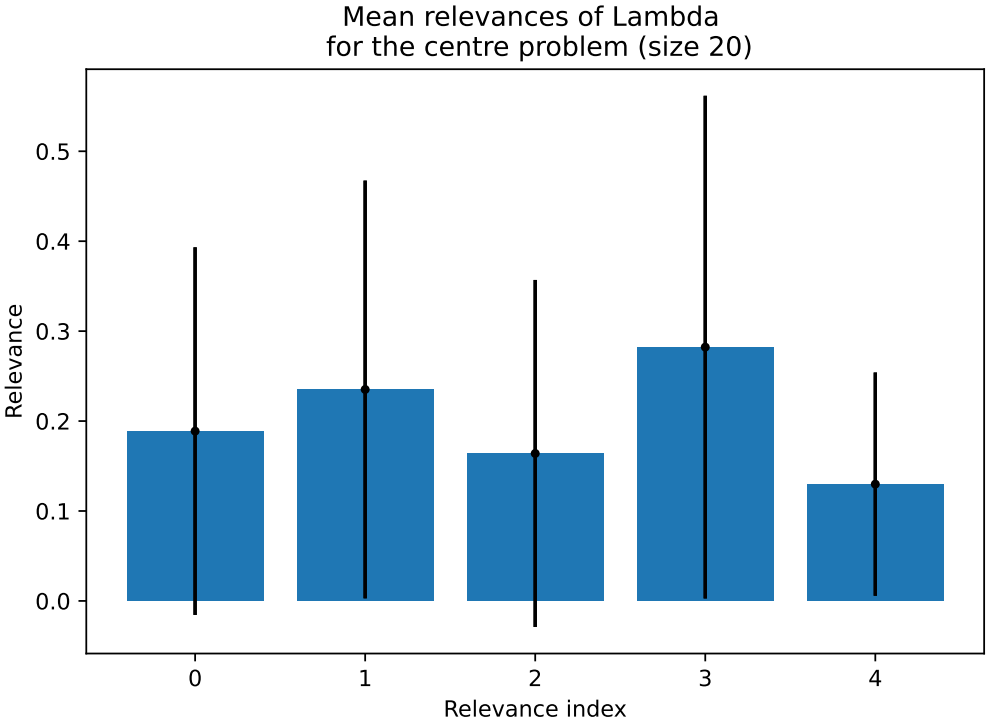

Figure 38: Mean relevance profile of case A with $n = 20$.

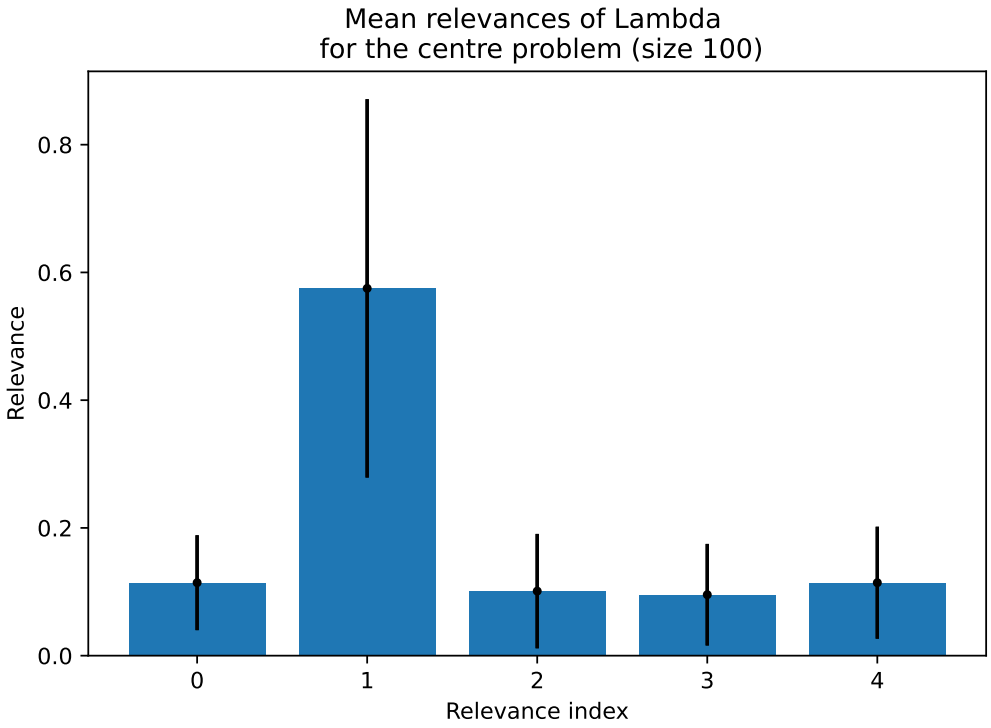

Figure 39: Mean relevance profile of case A with $n = 100$.

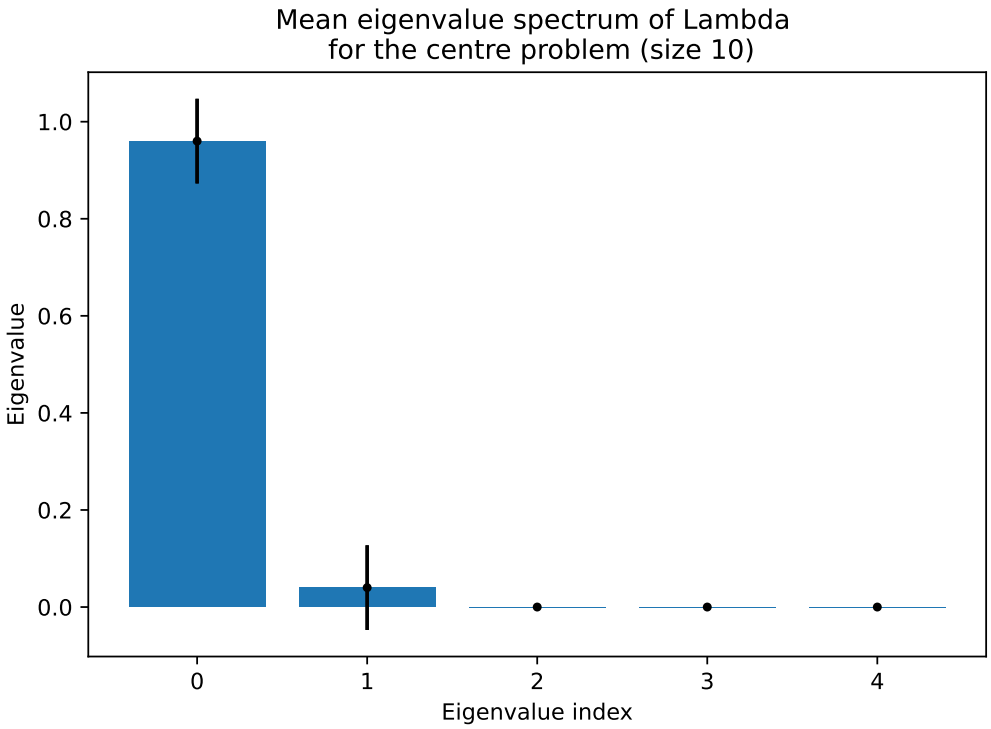

Figure 40: Mean eigenvalue spectrum of Lambda of case A with $n = 10$.

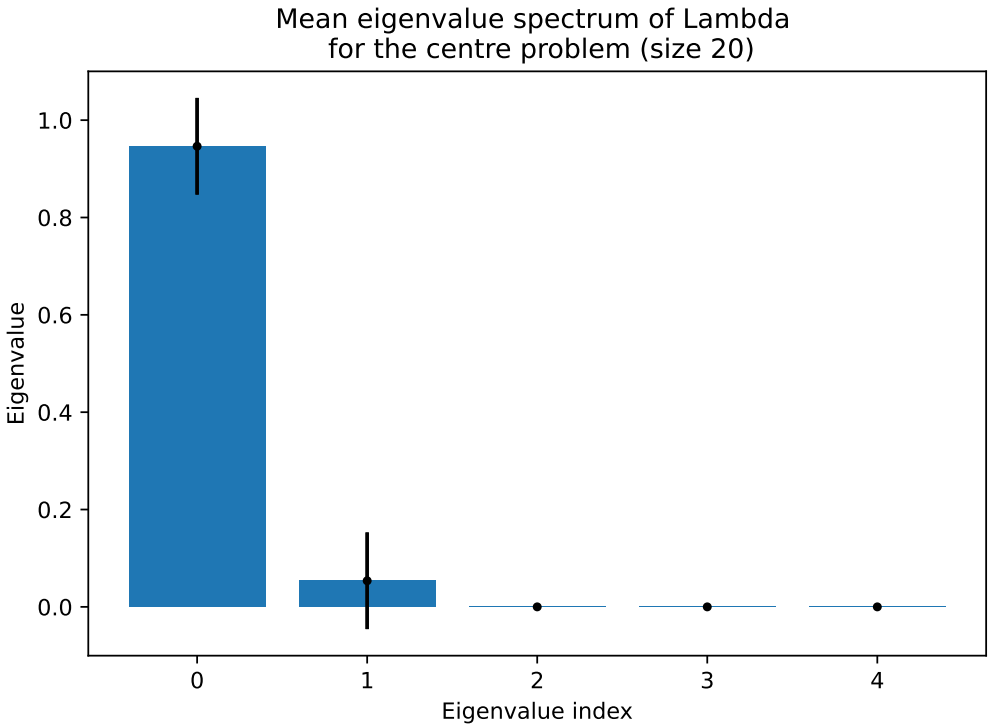

Figure 41: Mean eigenvalue spectrum of Lambda of case A with $n = 20$.

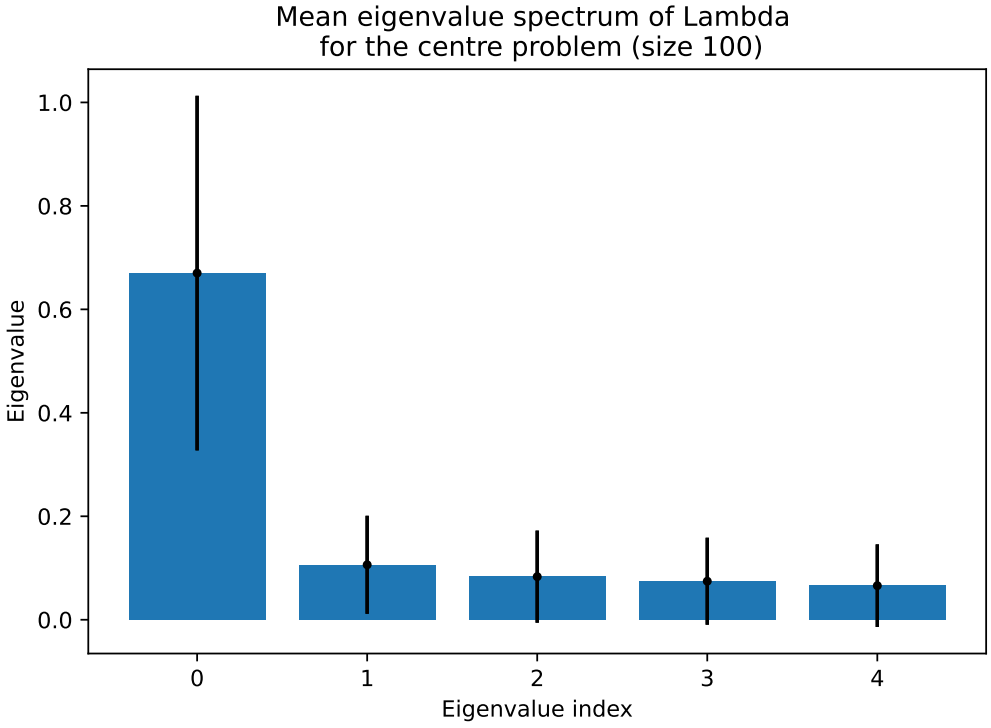

Figure 42: Mean eigenvalue spectrum of Lambda of case A with $n = 100$.

B.4 Centre-wise z-score transformed data: results for the disease classification

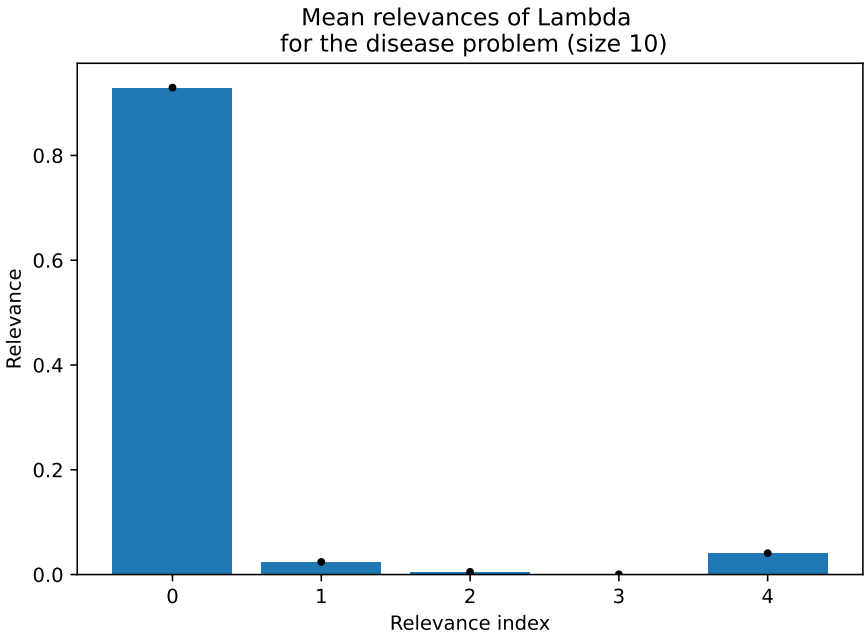

Figure 43: Mean relevance profile of case A with $n = 10$.

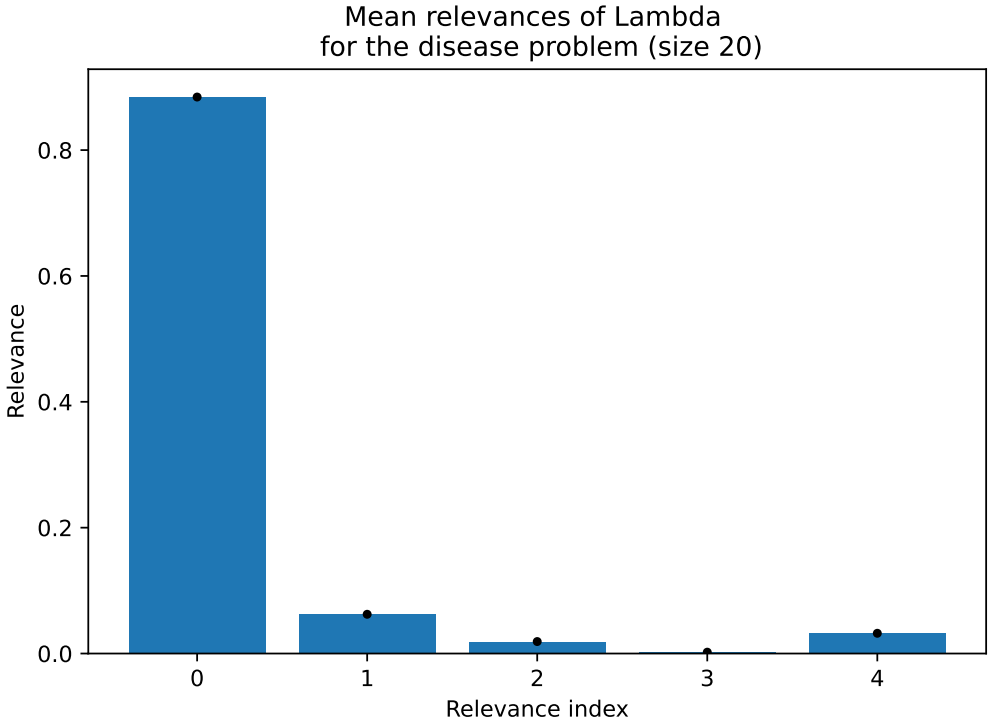

Figure 44: Mean relevance profile of case A with $n = 20$.

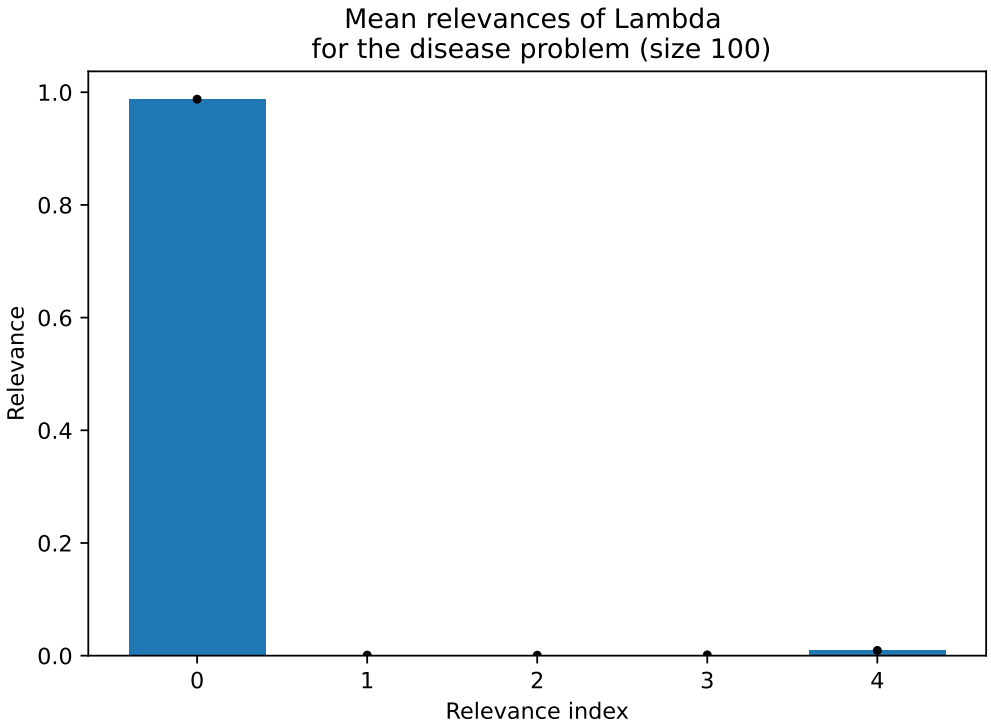

Figure 45: Mean relevance profile of case A with $n = 100$.

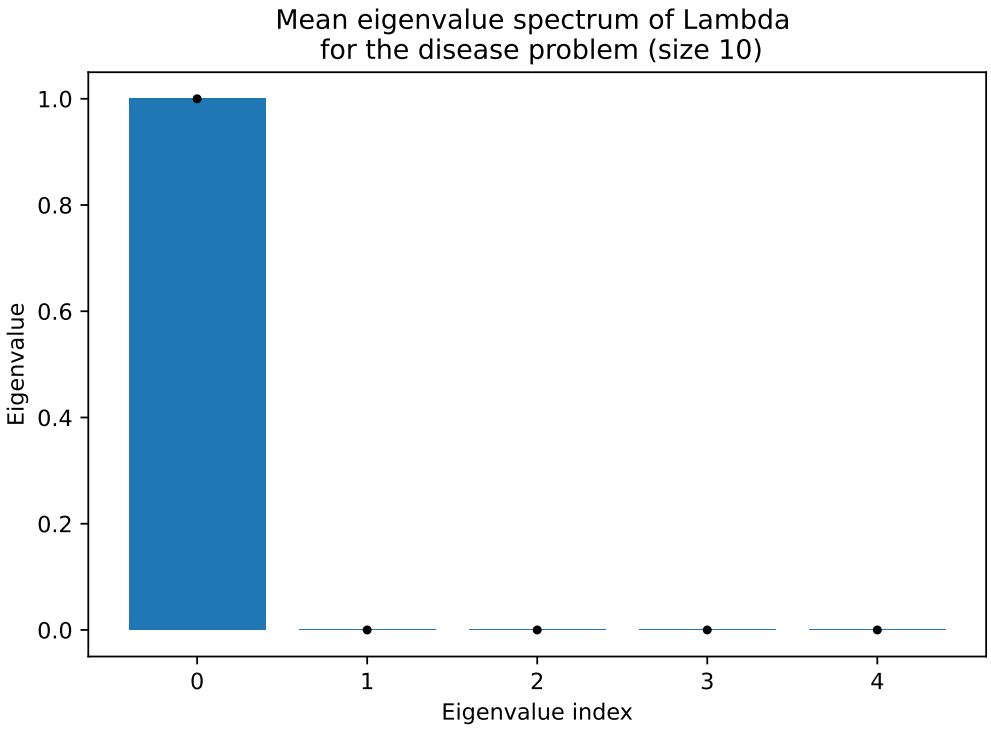

Figure 46: Mean eigenvalue spectrum of Lambda of case A with $n = 10$.

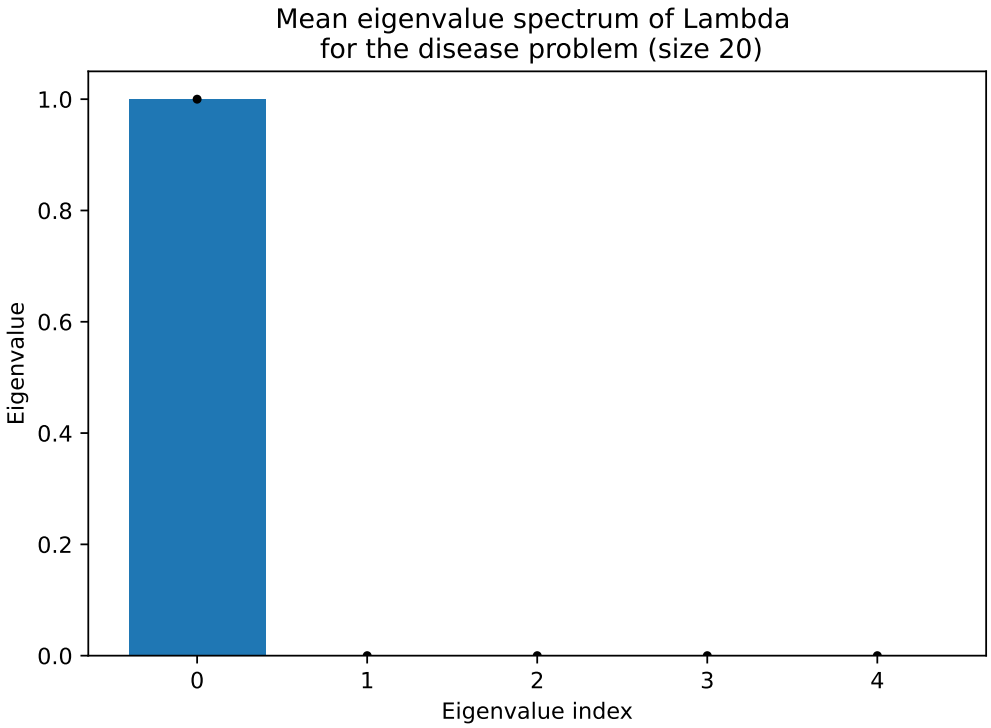

Figure 47: Mean eigenvalue spectrum of Lambda of case A with $n = 20$.

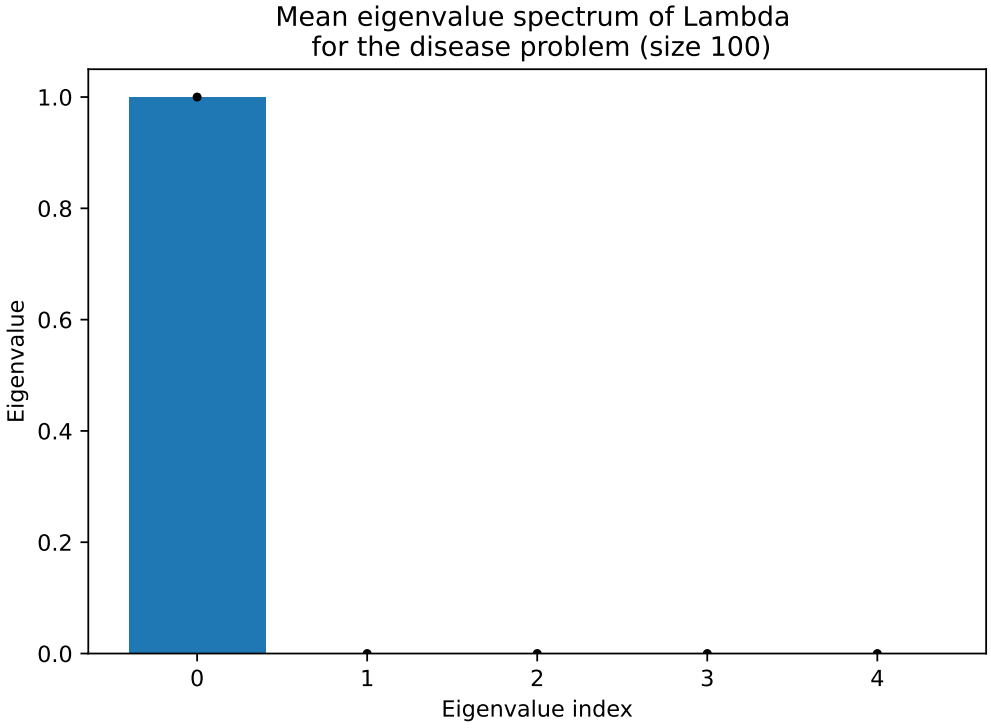

Figure 48: Mean eigenvalue spectrum of Lambda of case A with $n = 100$.

B.5 Data set stability

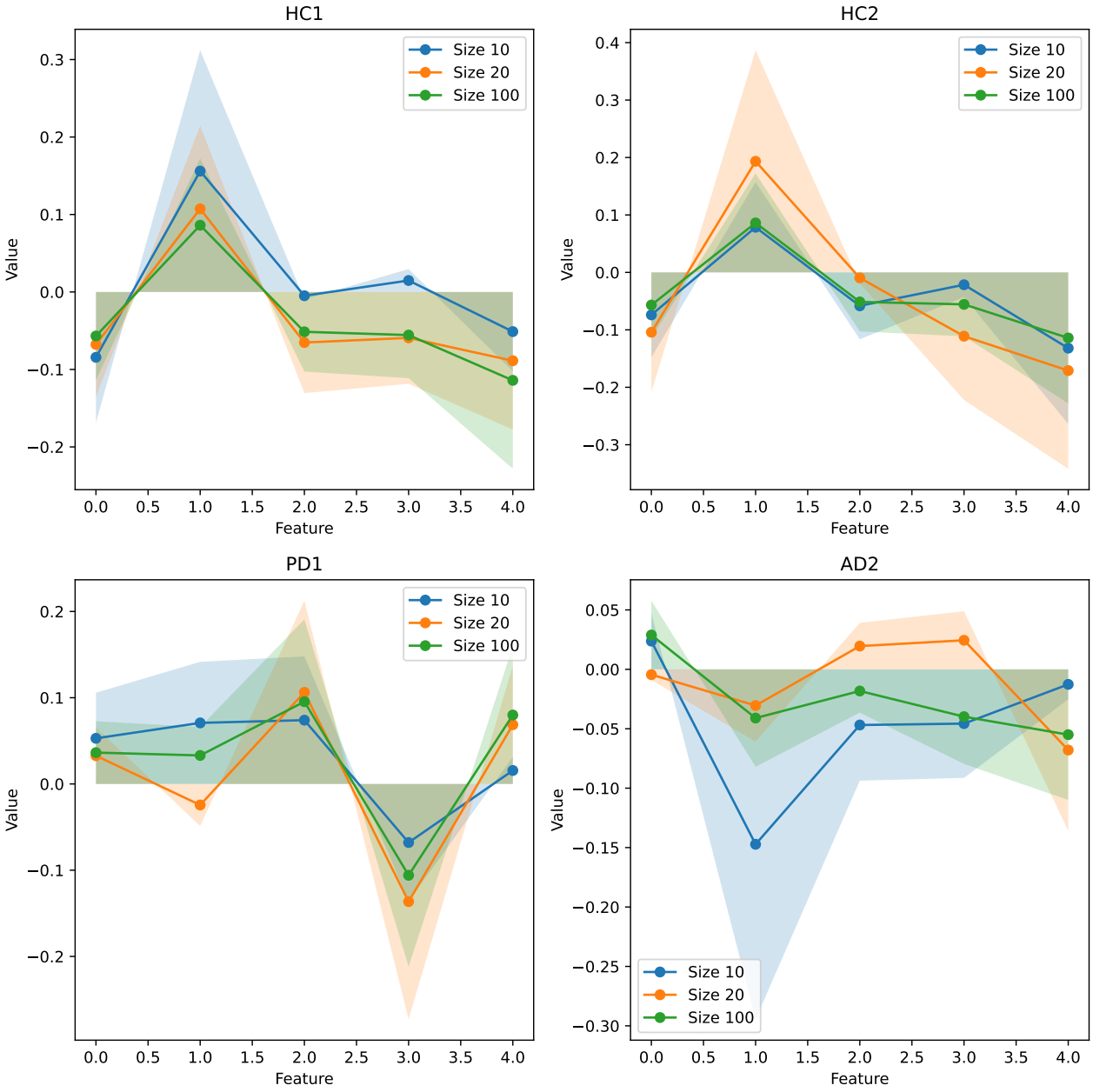

Figure 49: Mean (estimated mean - ground truth mean) as a function of the number of features.

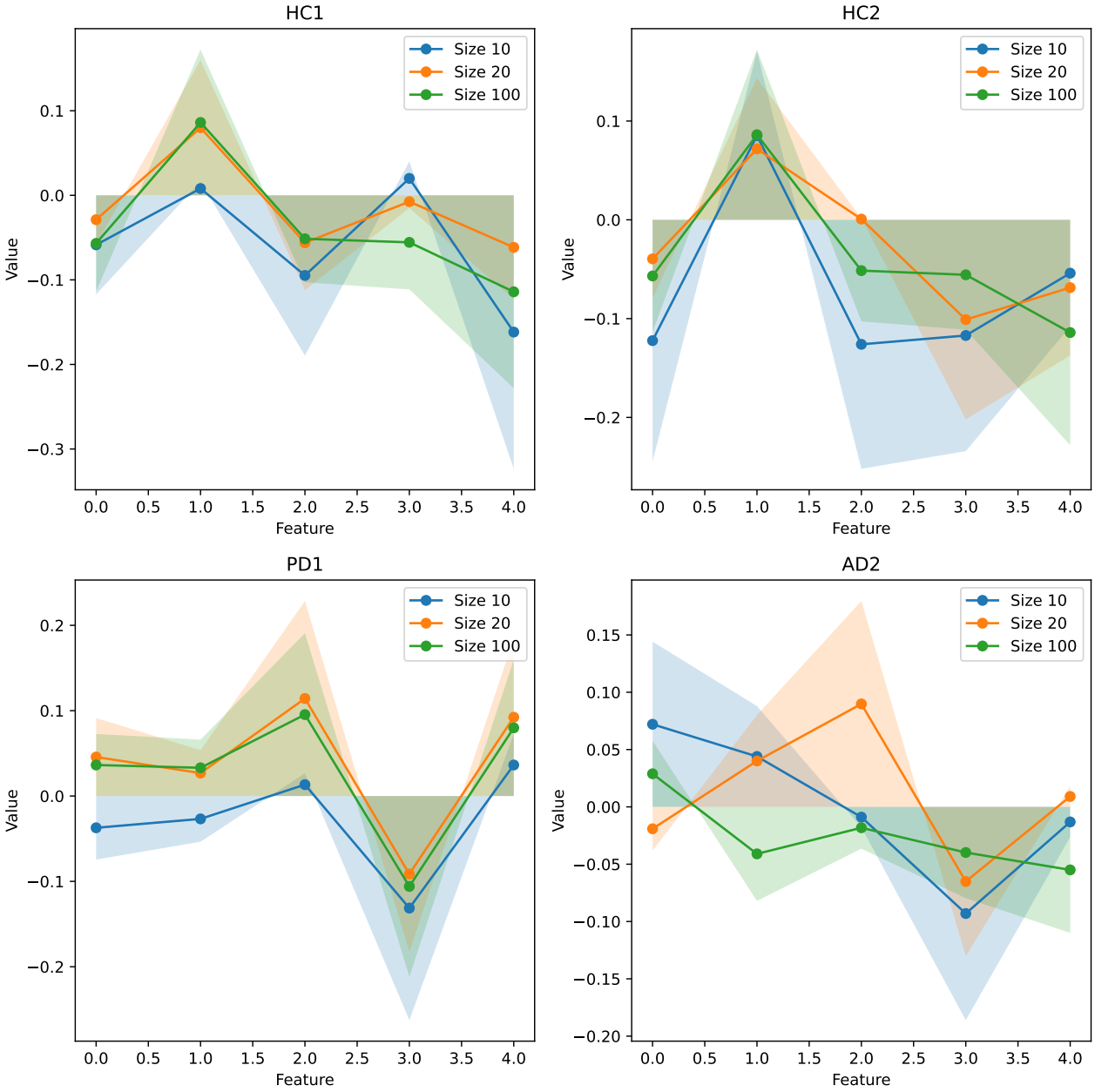

Figure 50: Mean (estimated mean - ground truth mean) as a function of the number of features of the centre-wise z-scored data.

C Case B

In this appendix several plots supplementing the results in Section 5.2 are presented, including the mean relevance profiles, eigenvalue spectrum of Lambda over the iterations and the difference between the data set mean and ground truth mean to show the stability of the data sets used for the experiments.

C.1 Results for the centre classification

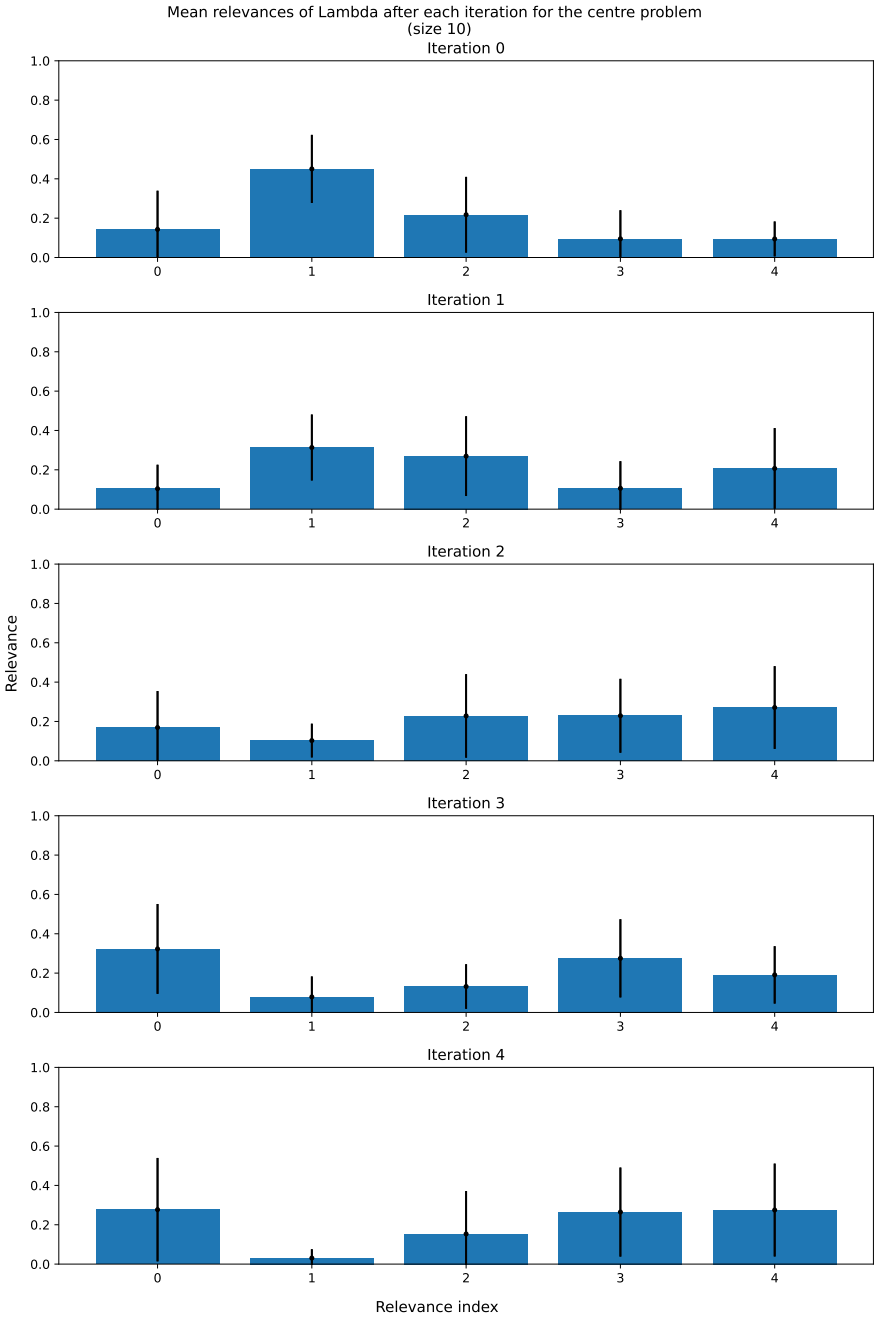

Figure 51: Mean relevance profile of case B with $n = 10$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

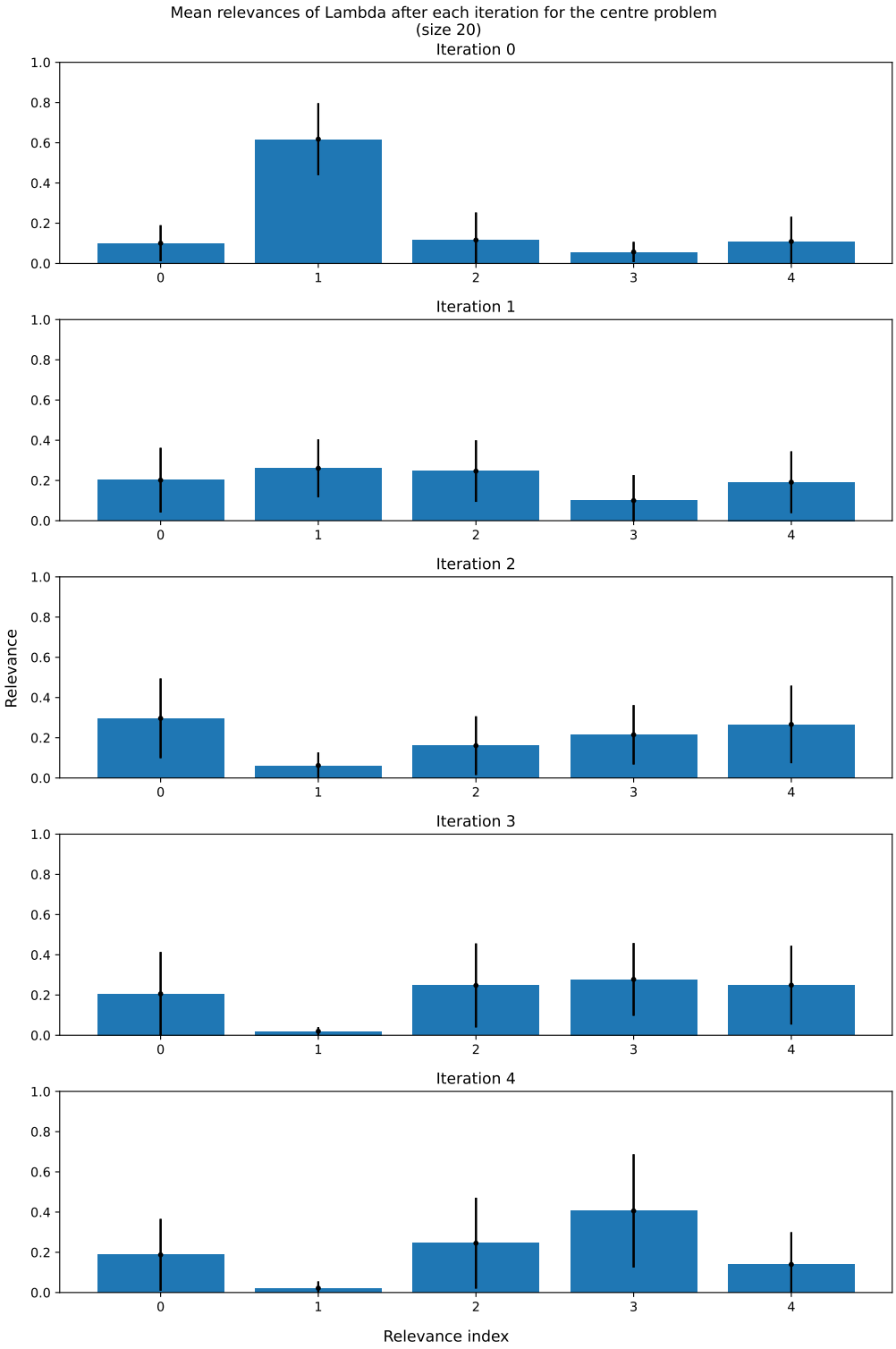

Figure 52: Mean relevance profile of case B with $n = 20$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

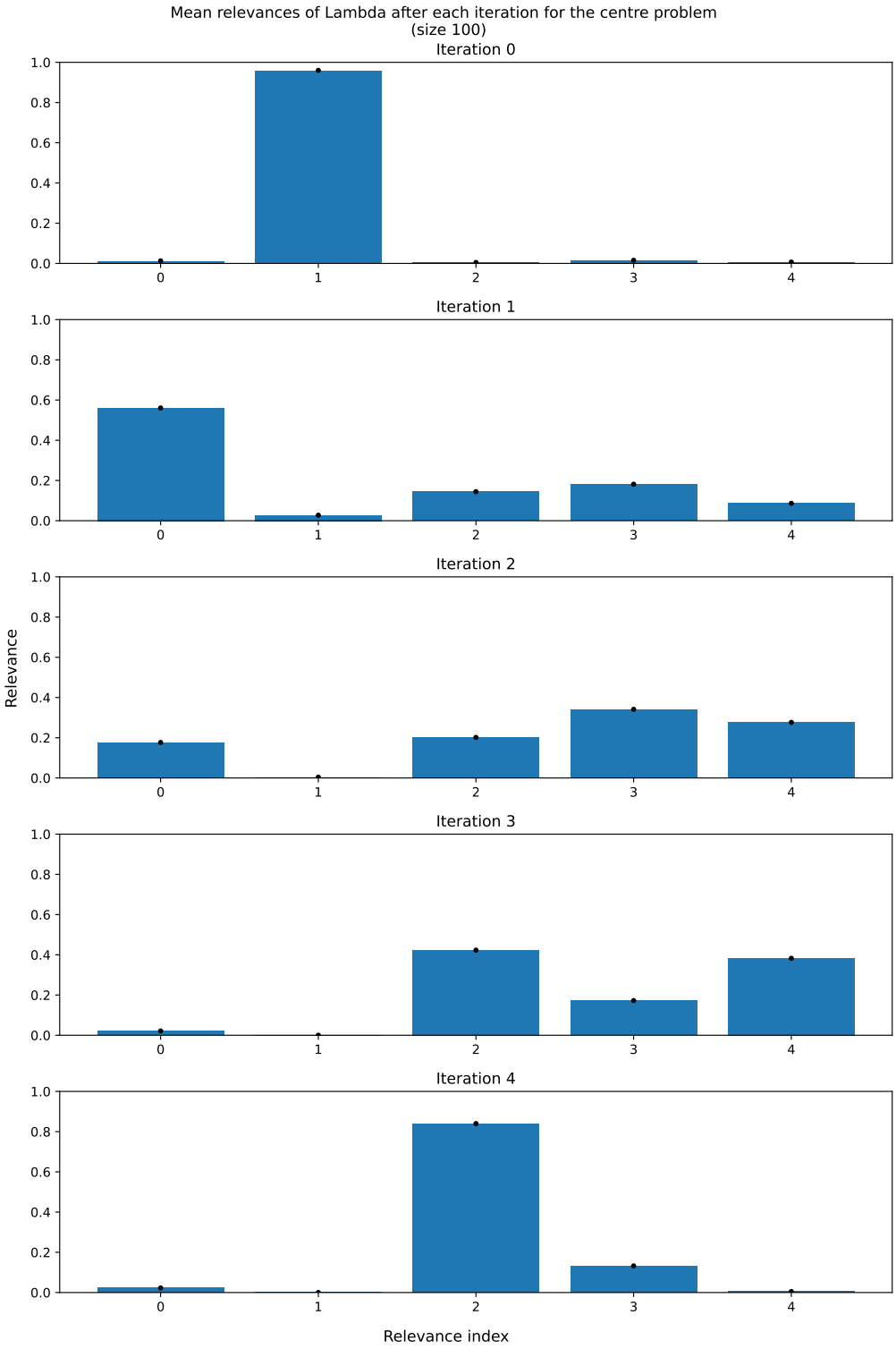

Figure 53: Mean relevance profile of case B with $n = 100$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

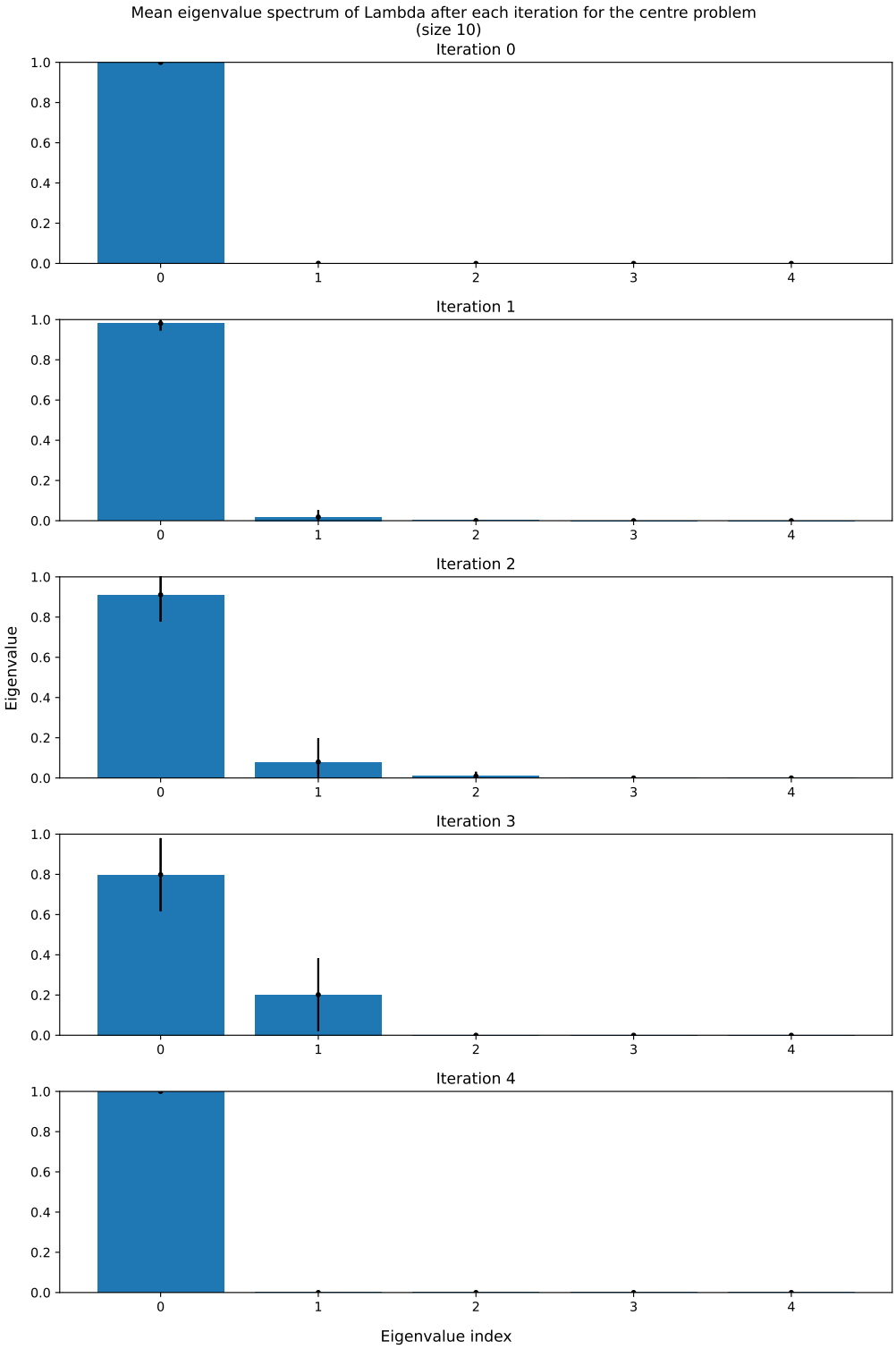

Figure 54: Mean eigenvalue spectrum of Lambda of case B with $n = 10$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

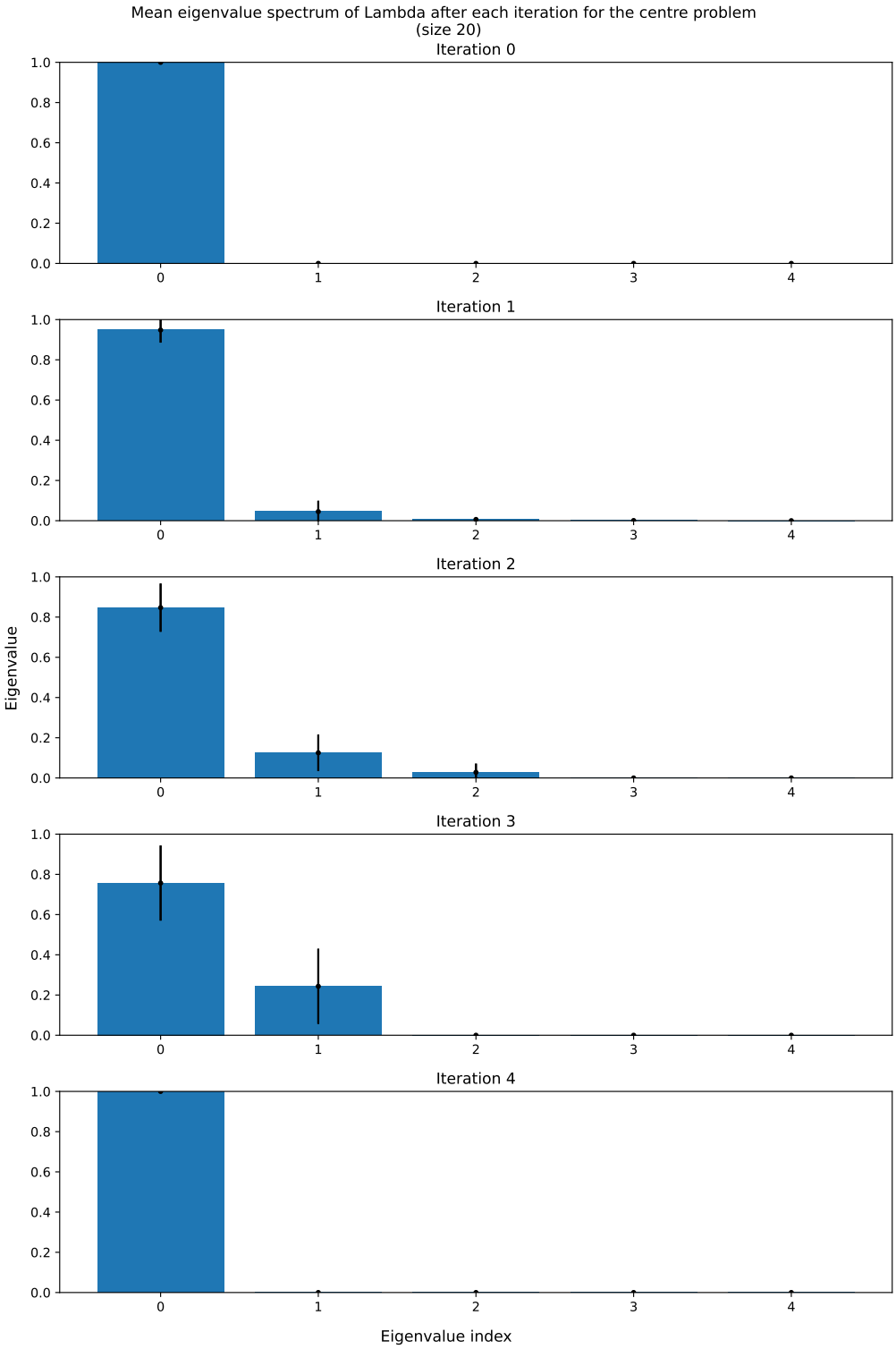

Figure 55: Mean eigenvalue spectrum of Lambda of case B with $n = 20$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

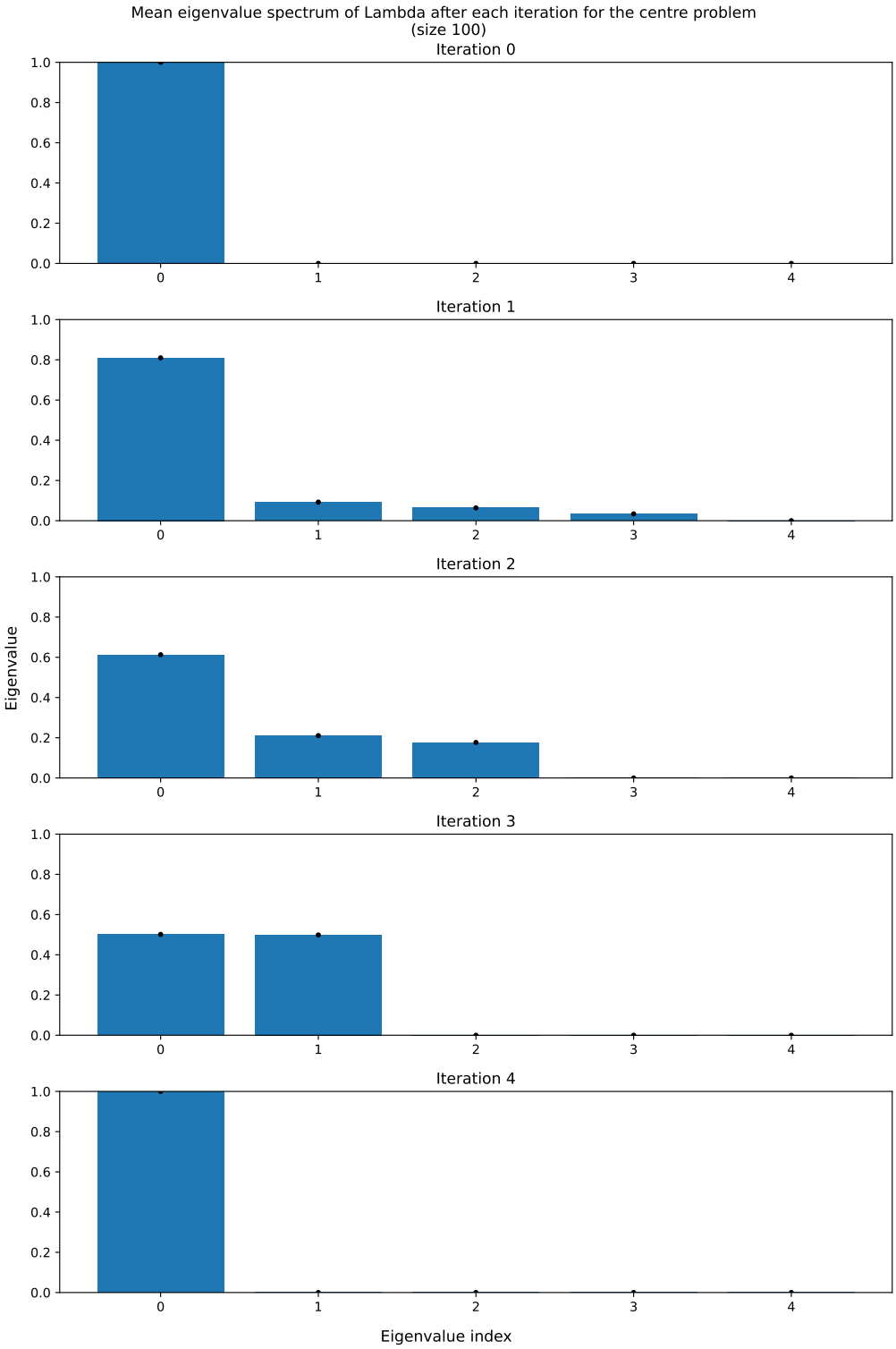

Figure 56: Mean eigenvalue spectrum of Lambda of case B with $n = 100$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

C.2 Results for the disease classification

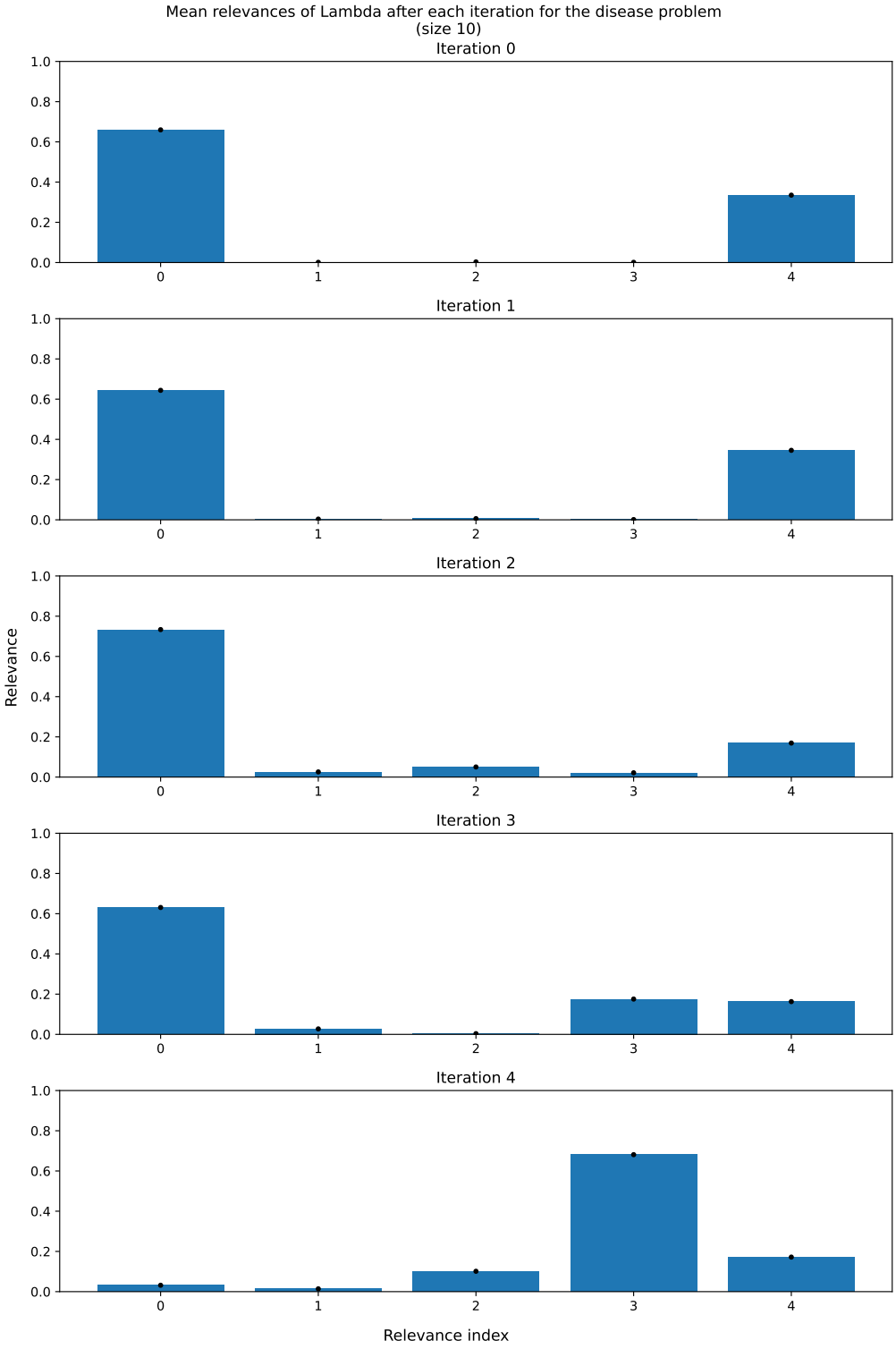

Figure 57: Mean relevance profile of case B with $n = 10$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system..

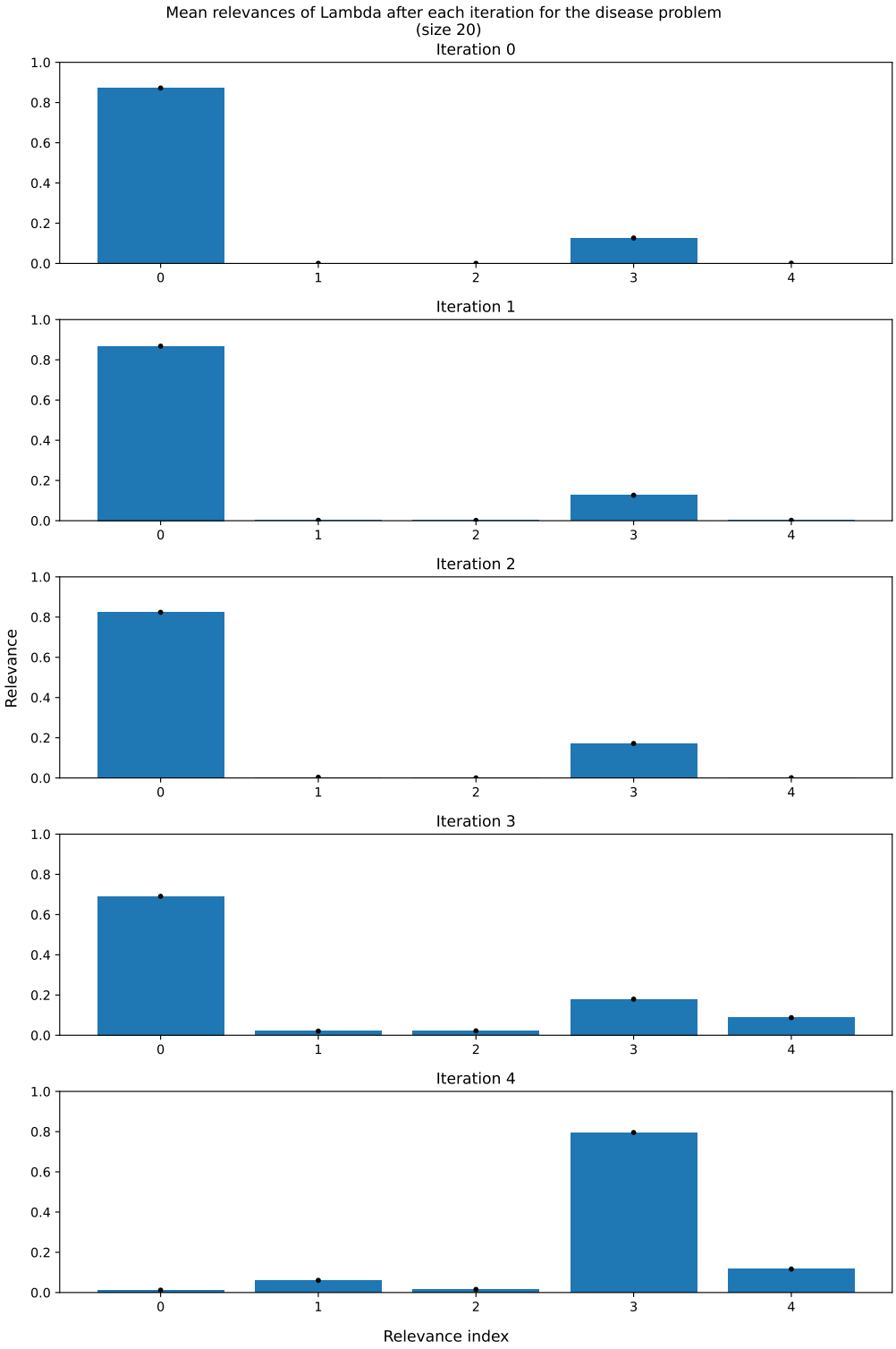

Figure 58: Mean relevance profile of case B with $n = 20$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system..

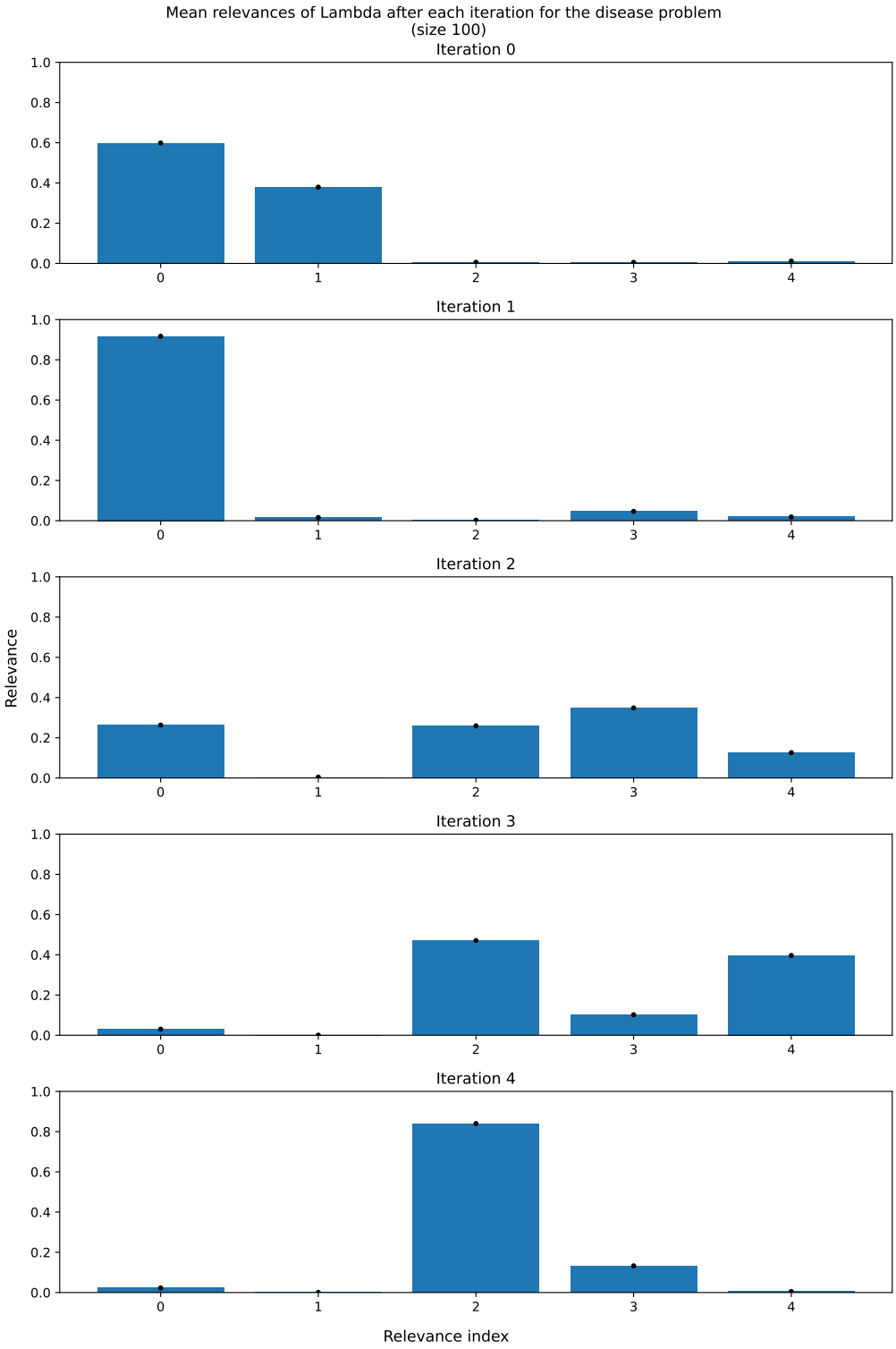

Figure 59: Mean relevance profile of case B with $n = 100$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system..

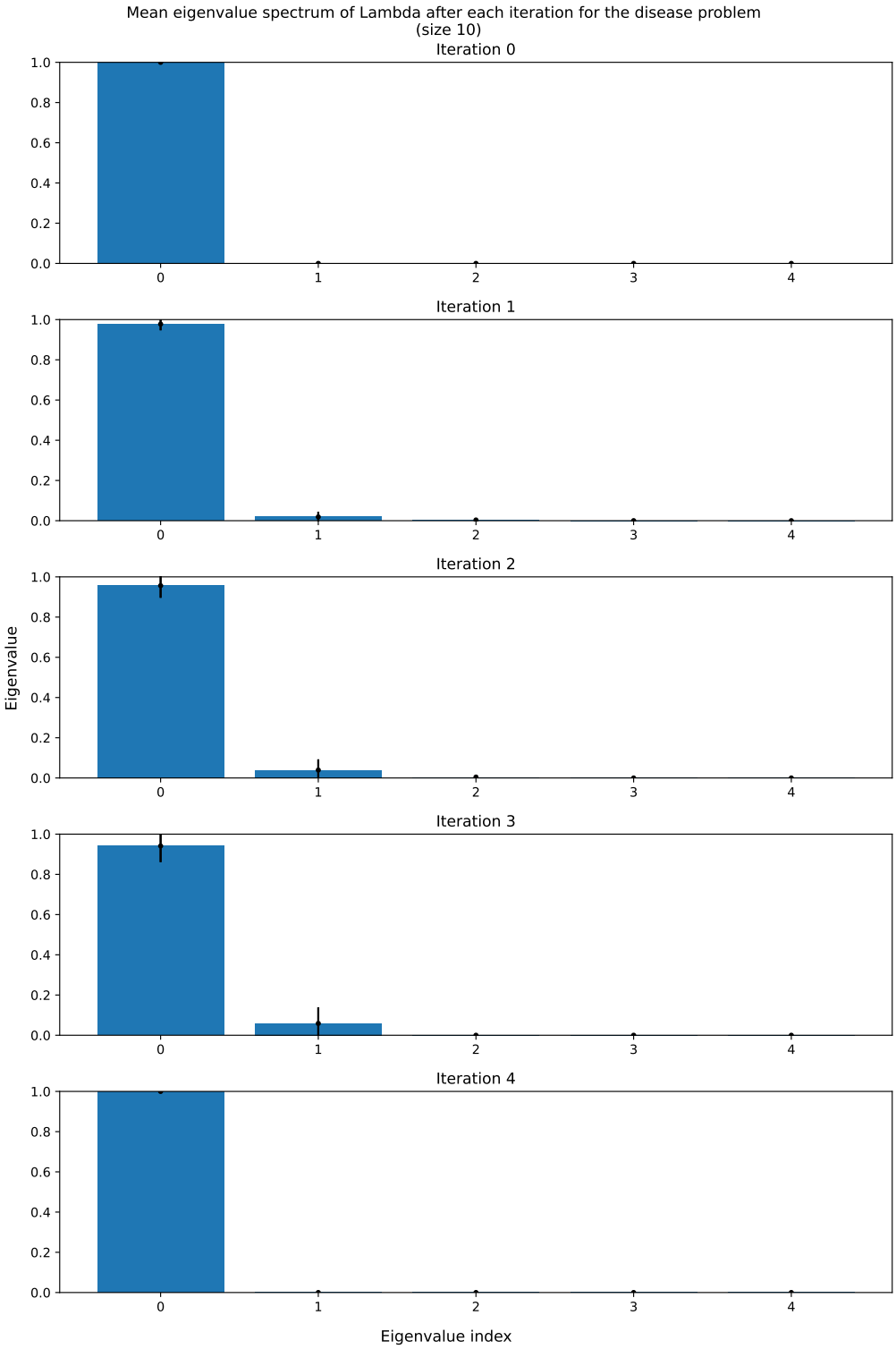

Figure 60: Mean eigenvalue spectrum of Lambda of case B with $n = 10$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system..

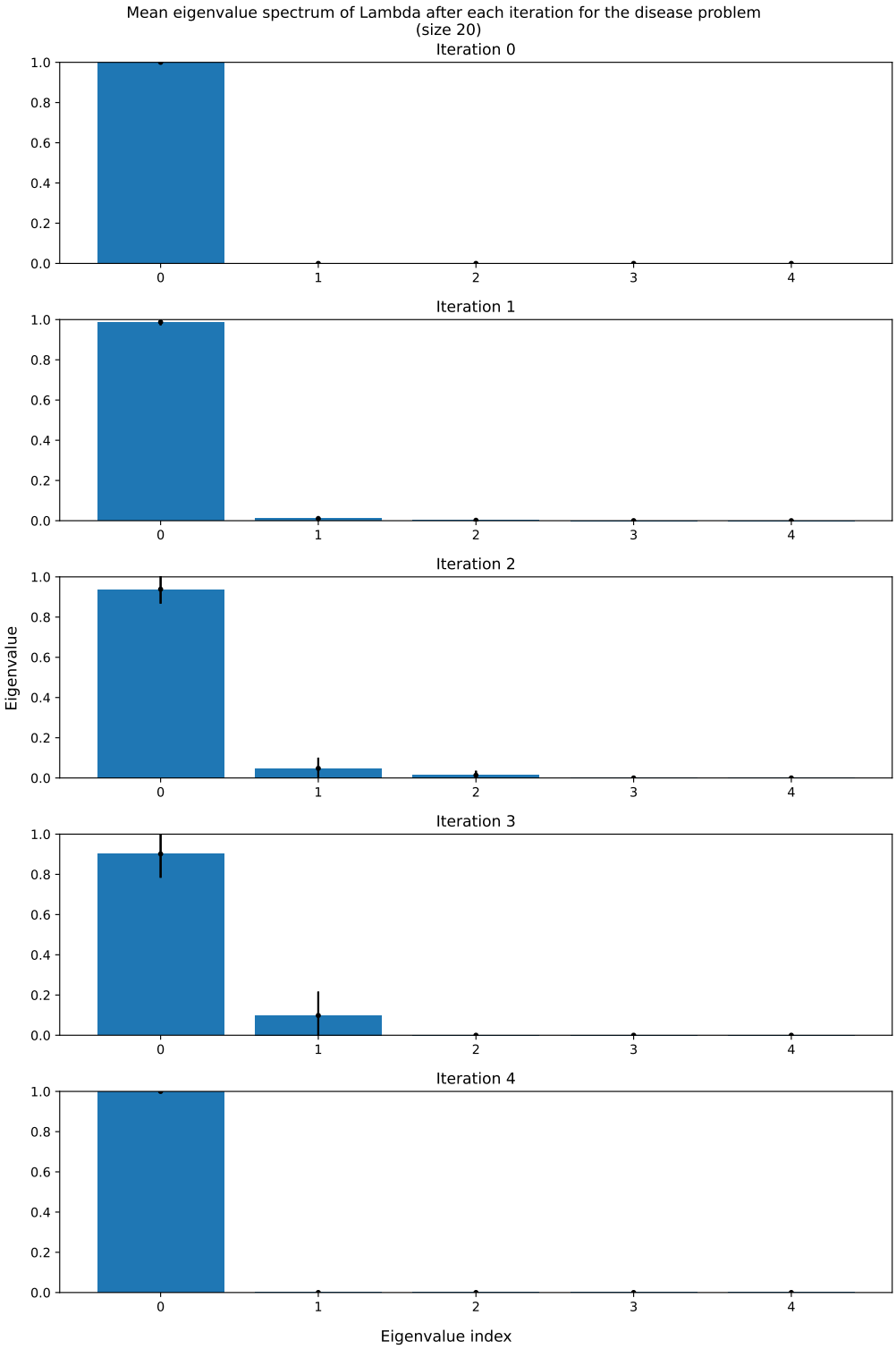

Figure 61: Mean eigenvalue spectrum of Lambda of case B with $n = 20$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

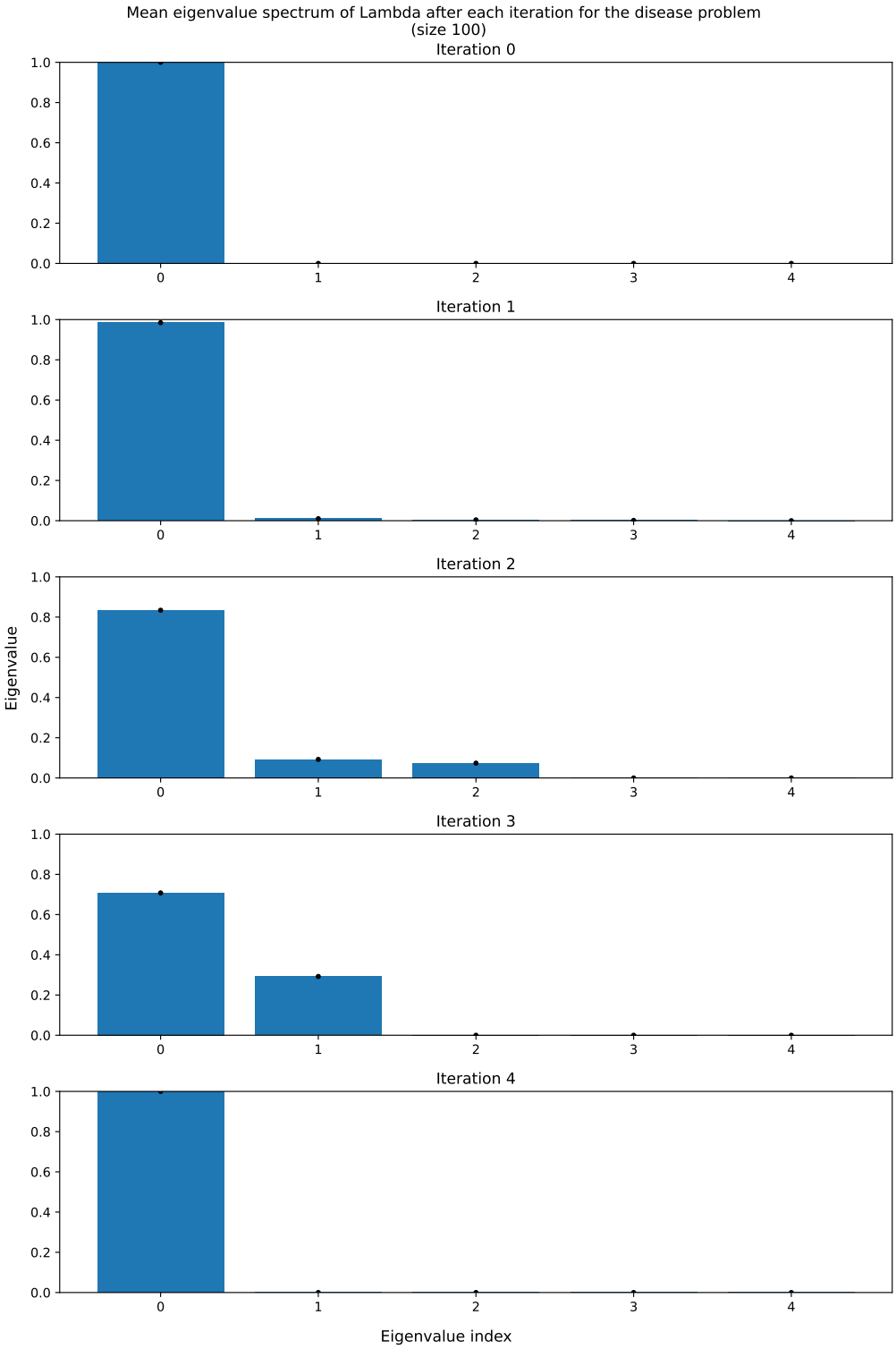

Figure 62: Mean eigenvalue spectrum of Lambda of case B with $n = 100$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

C.3 Centre-wise z-score transformed data: results for the centre classification

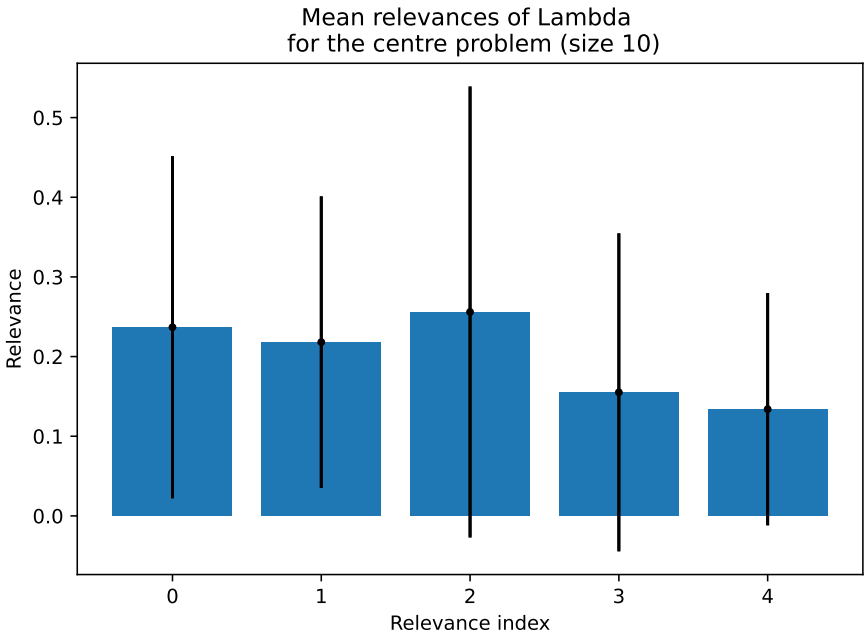

Figure 63: Mean relevance profile of case B with $n = 10$.

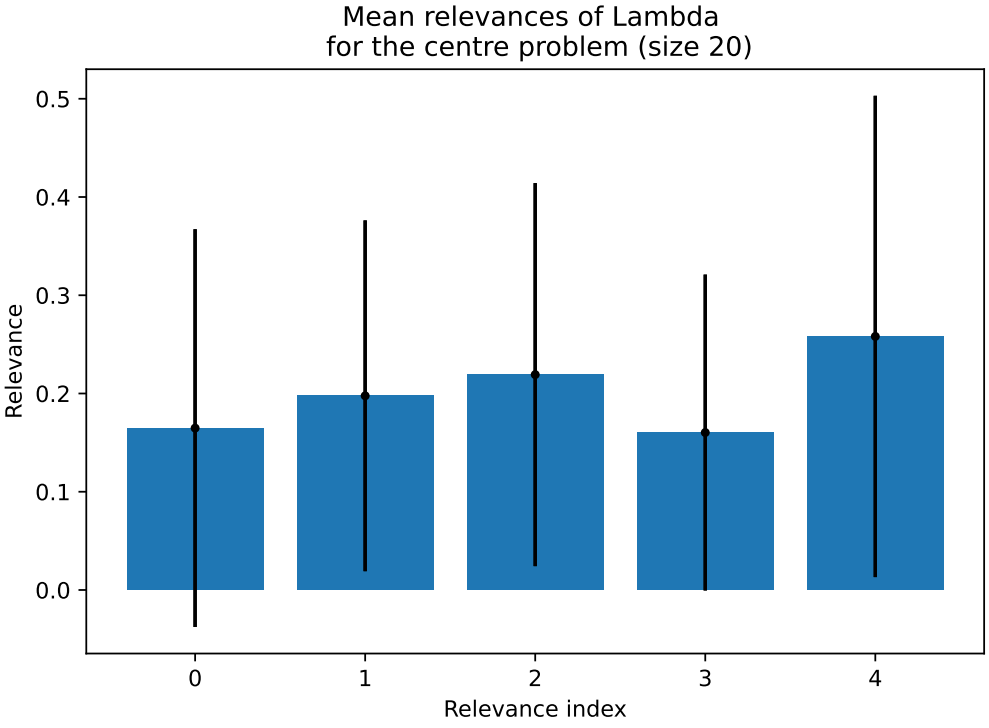

Figure 64: Mean relevance profile of case B with $n = 20$.

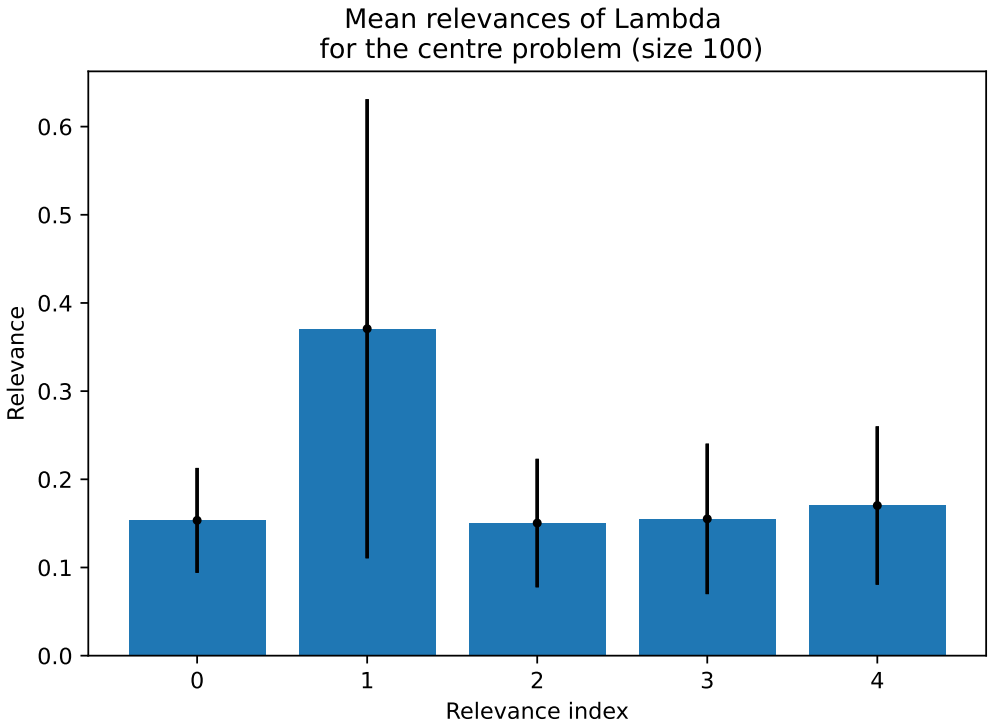

Figure 65: Mean relevance profile of case B with $n = 100$.

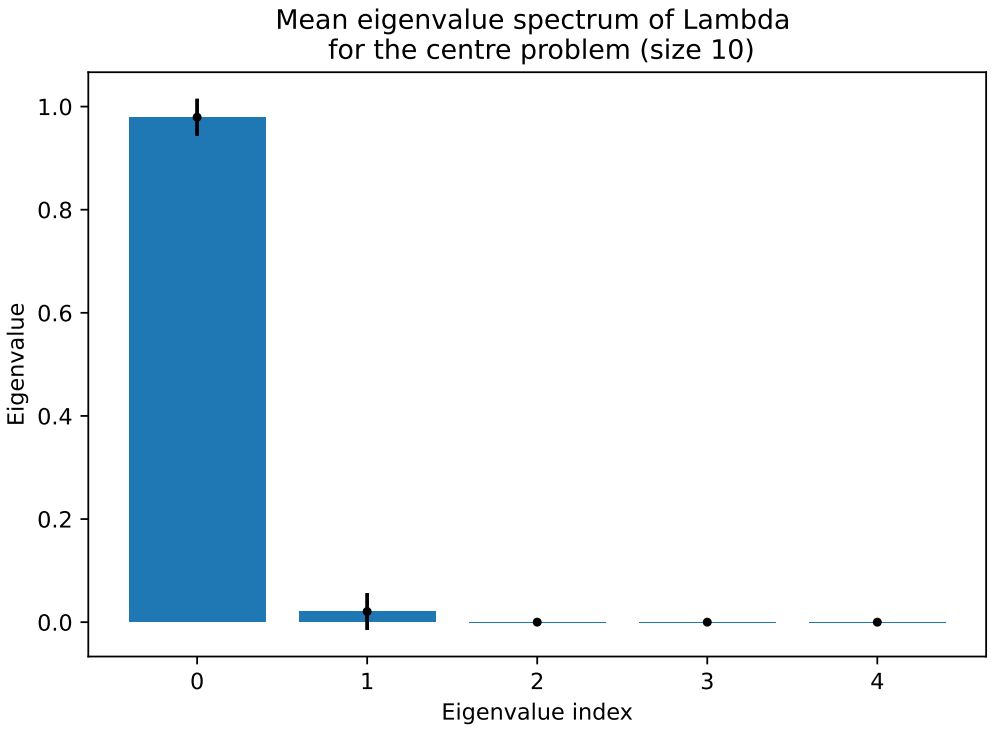

Figure 66: Mean eigenvalue spectrum of Lambda of case B with $n = 10$.

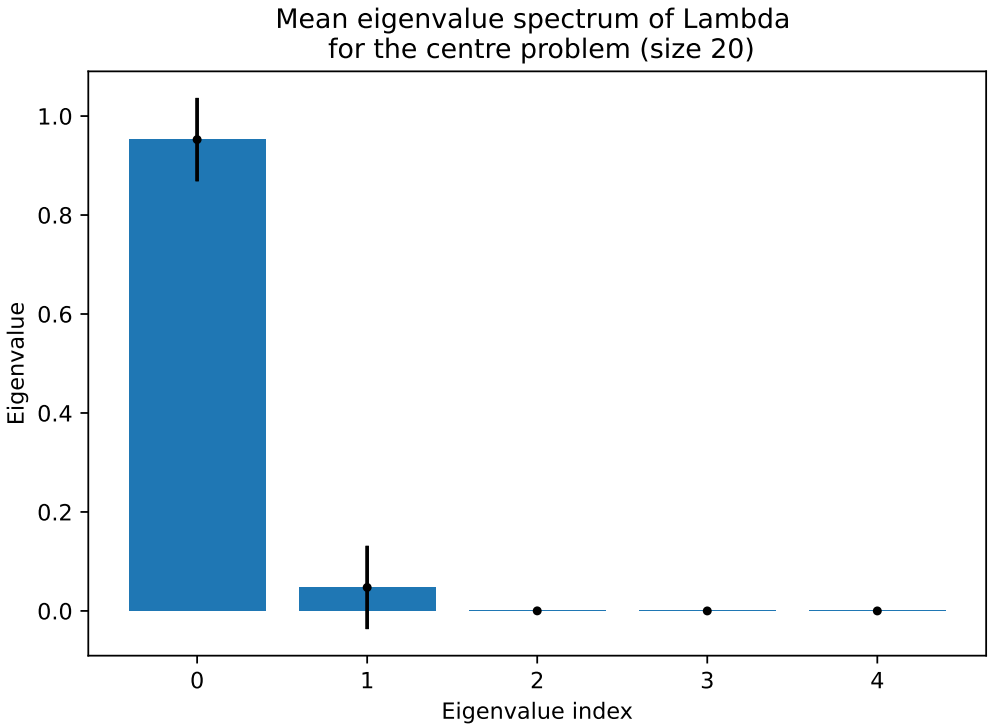

Figure 67: Mean eigenvalue spectrum of Lambda of case B with $n = 20$.

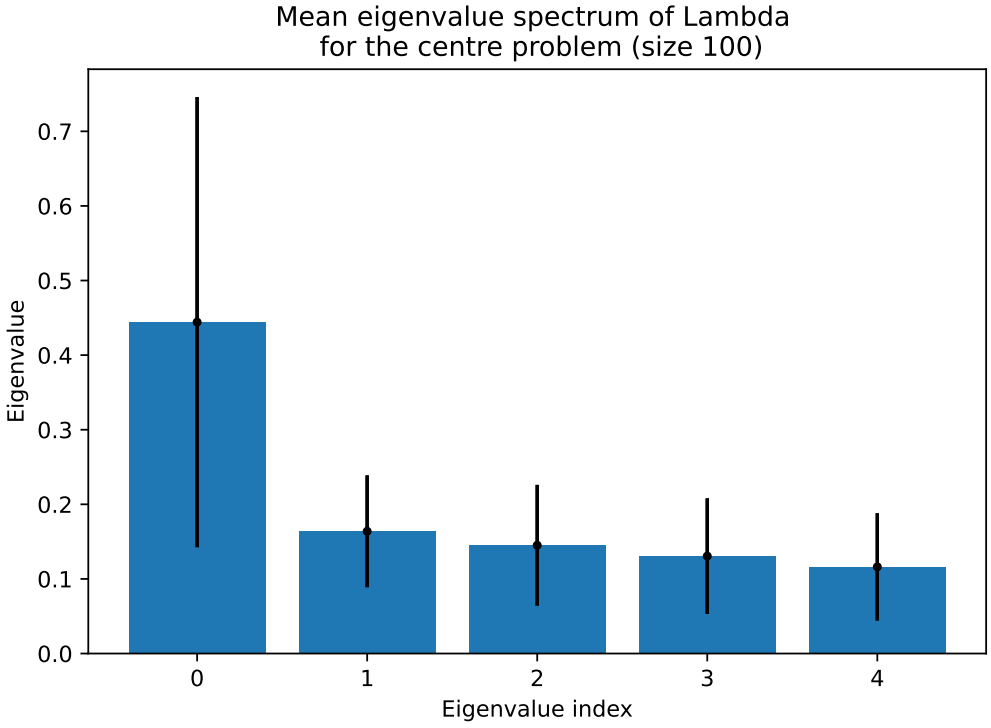

Figure 68: Mean eigenvalue spectrum of Lambda of case B with $n = 100$.

C.4 Centre-wise z-score transformed data: results for the disease classification

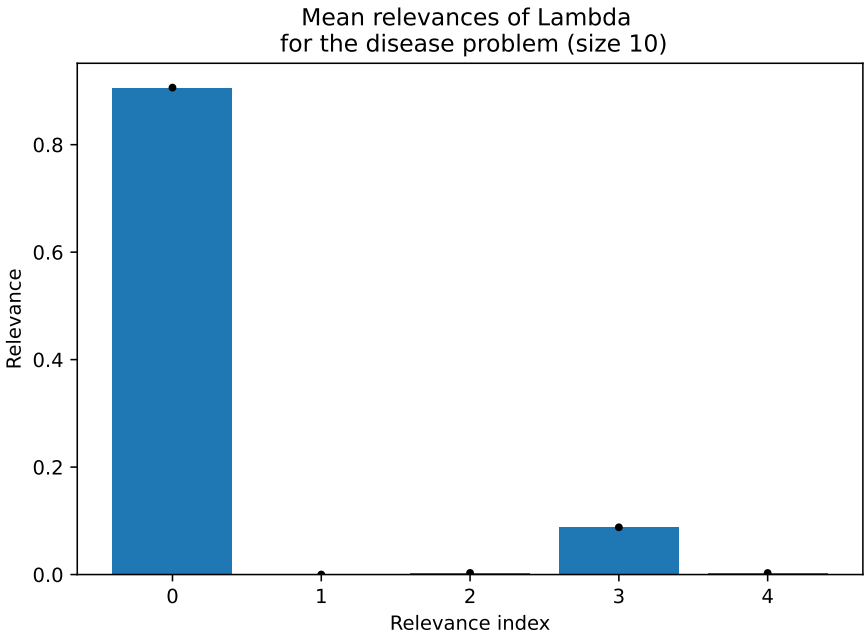

Figure 69: Mean relevance profile of case B with $n = 10$.

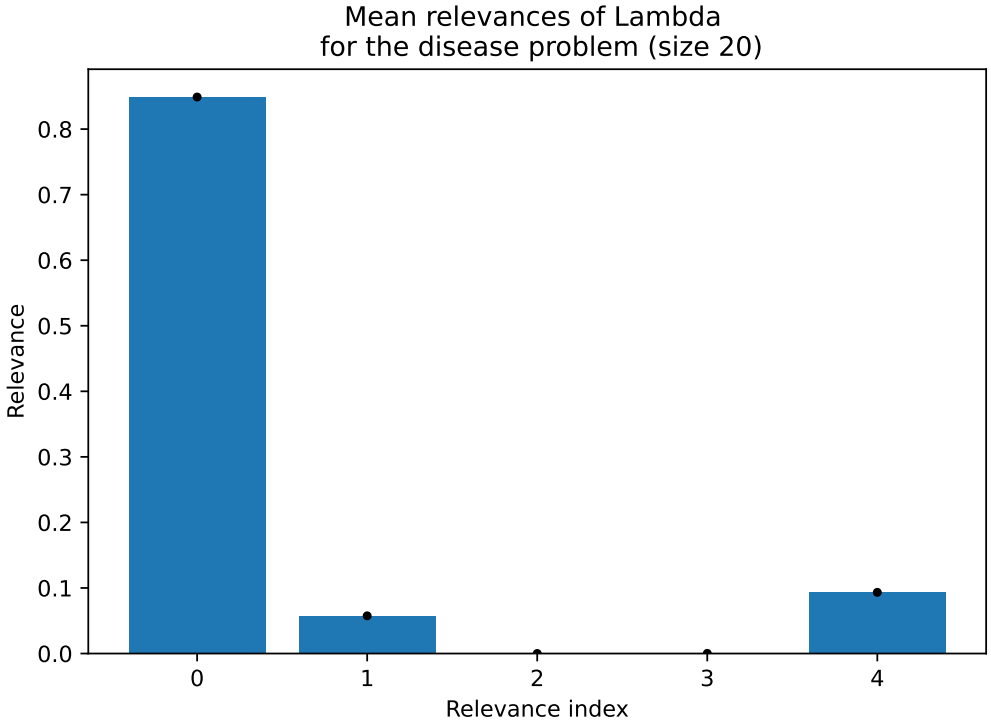

Figure 70: Mean relevance profile of case B with $n = 20$.

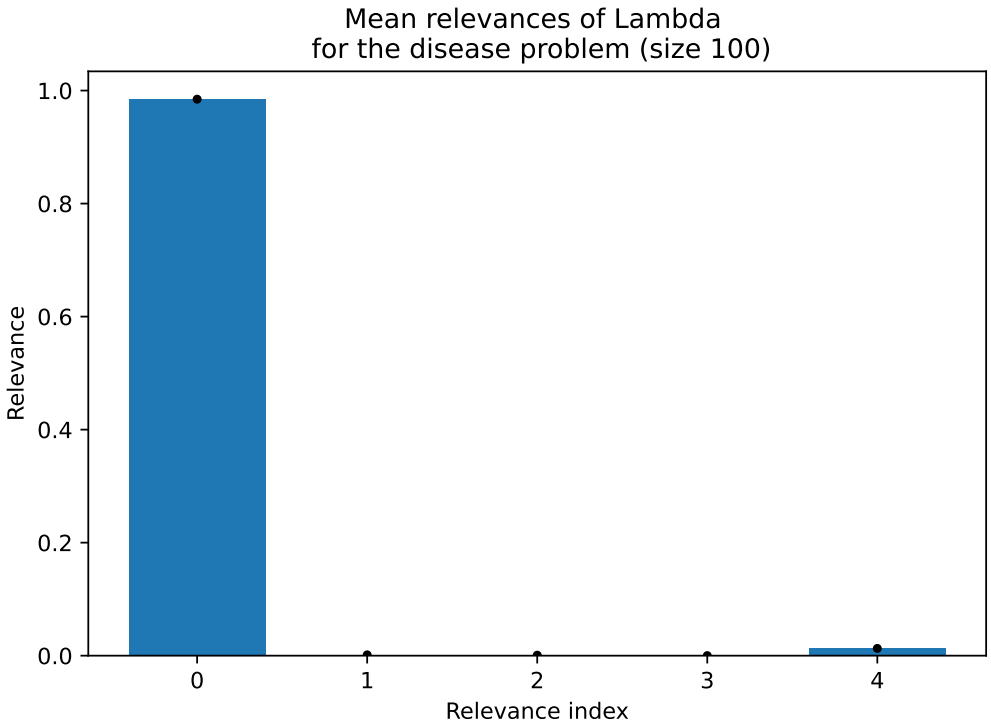

Figure 71: Mean relevance profile of case B with $n = 100$.

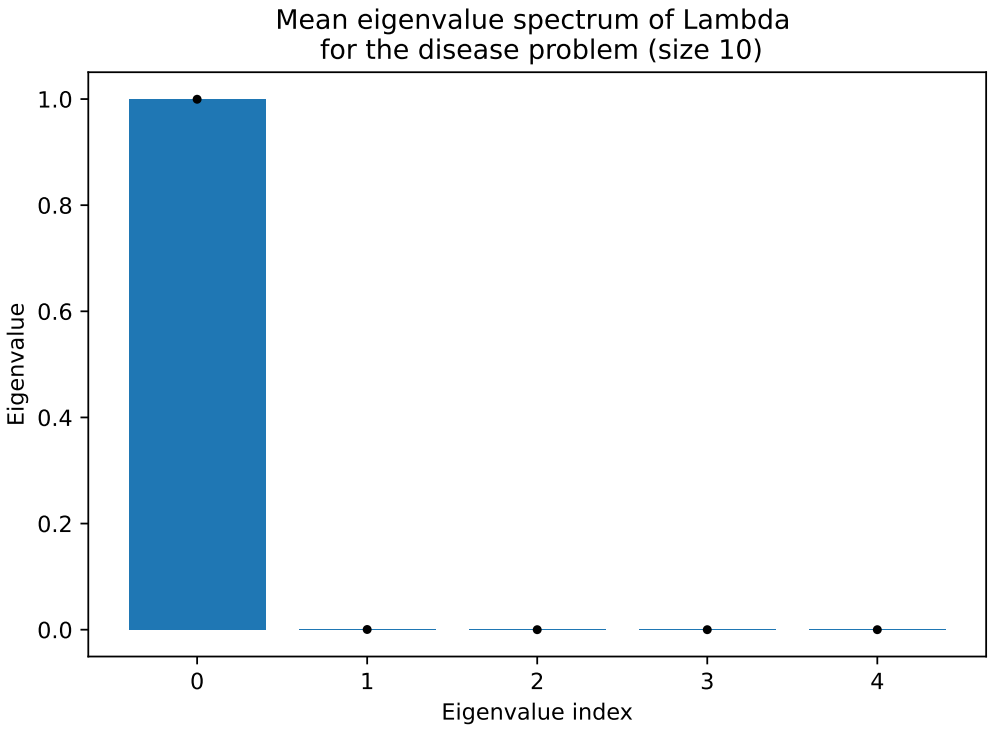

Figure 72: Mean eigenvalue spectrum of Lambda of case B with $n = 10$.

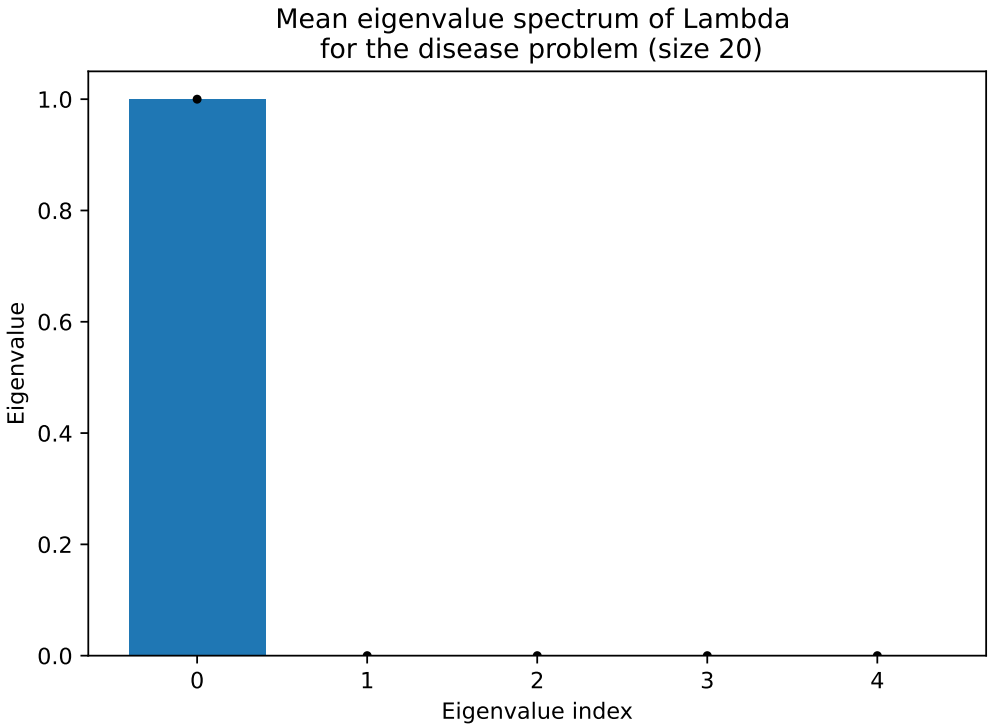

Figure 73: Mean eigenvalue spectrum of Lambda of case B with $n = 20$.

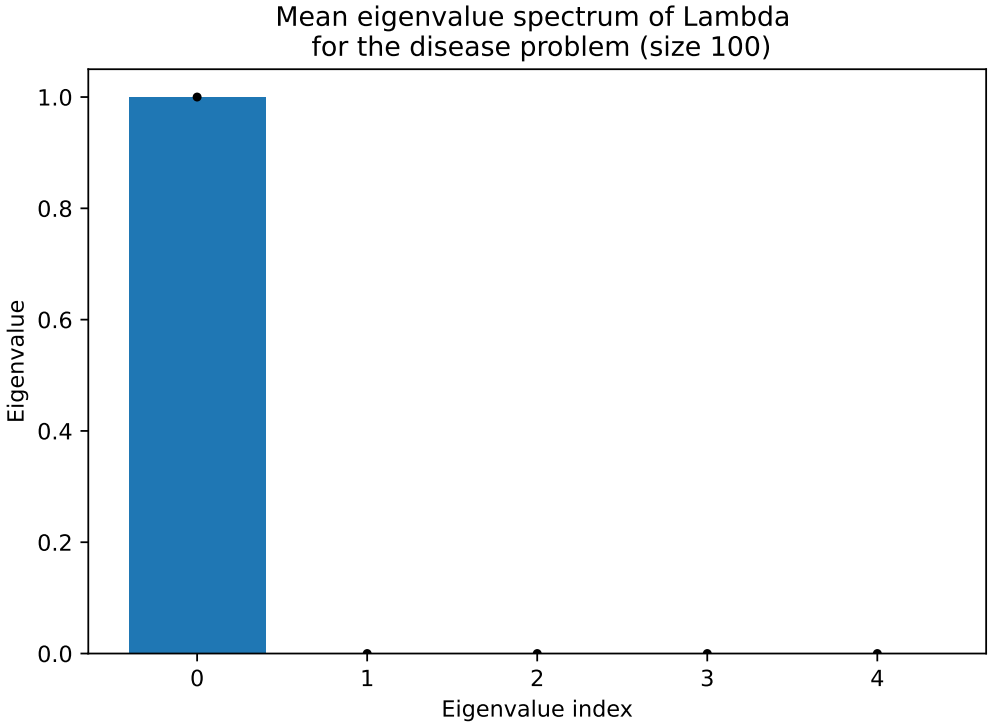

Figure 74: Mean eigenvalue spectrum of Lambda of case B with $n = 100$.

C.5 Data set stability

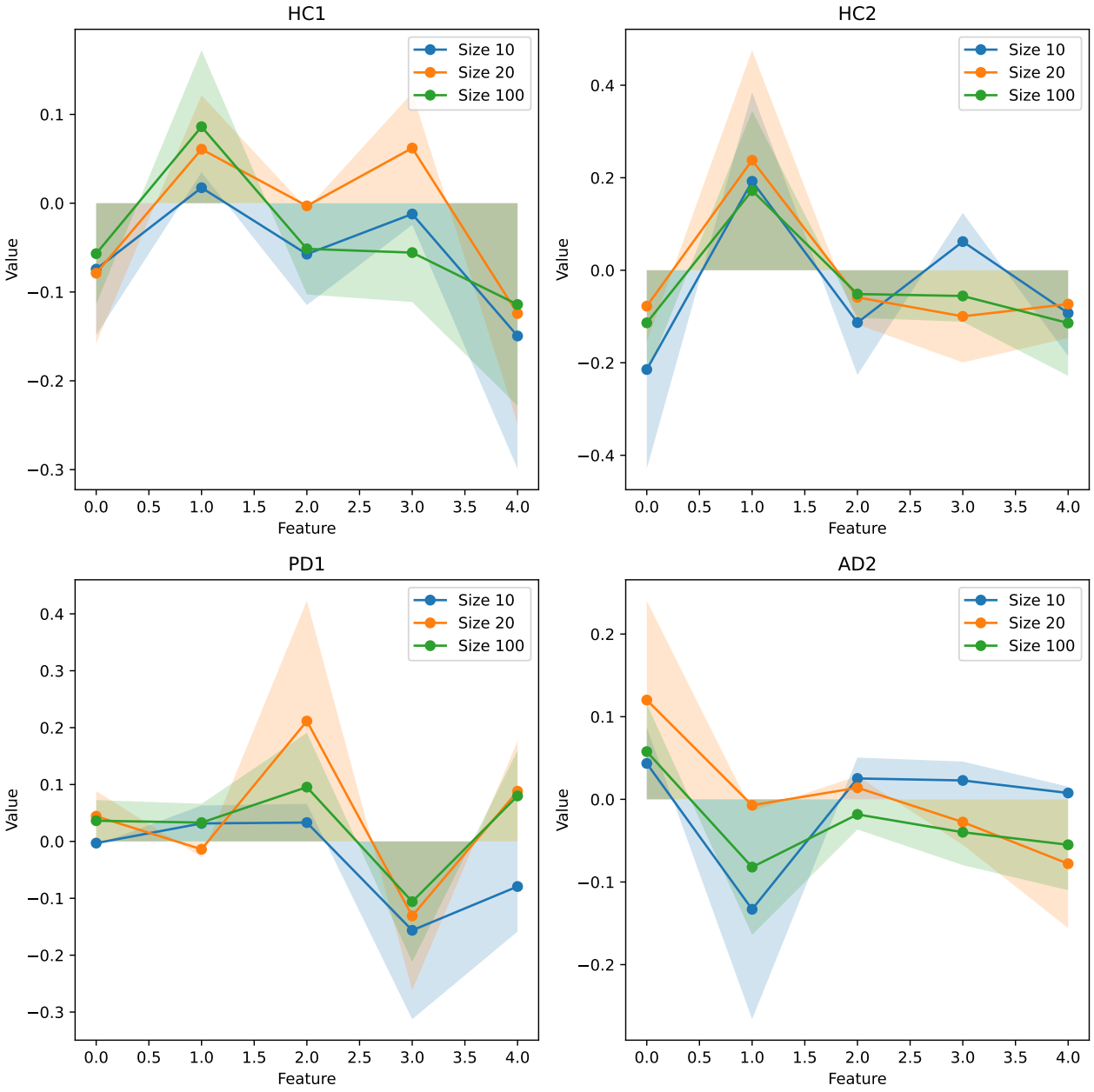

Figure 75: Mean (estimated mean - ground truth mean) as a function of the number of features.

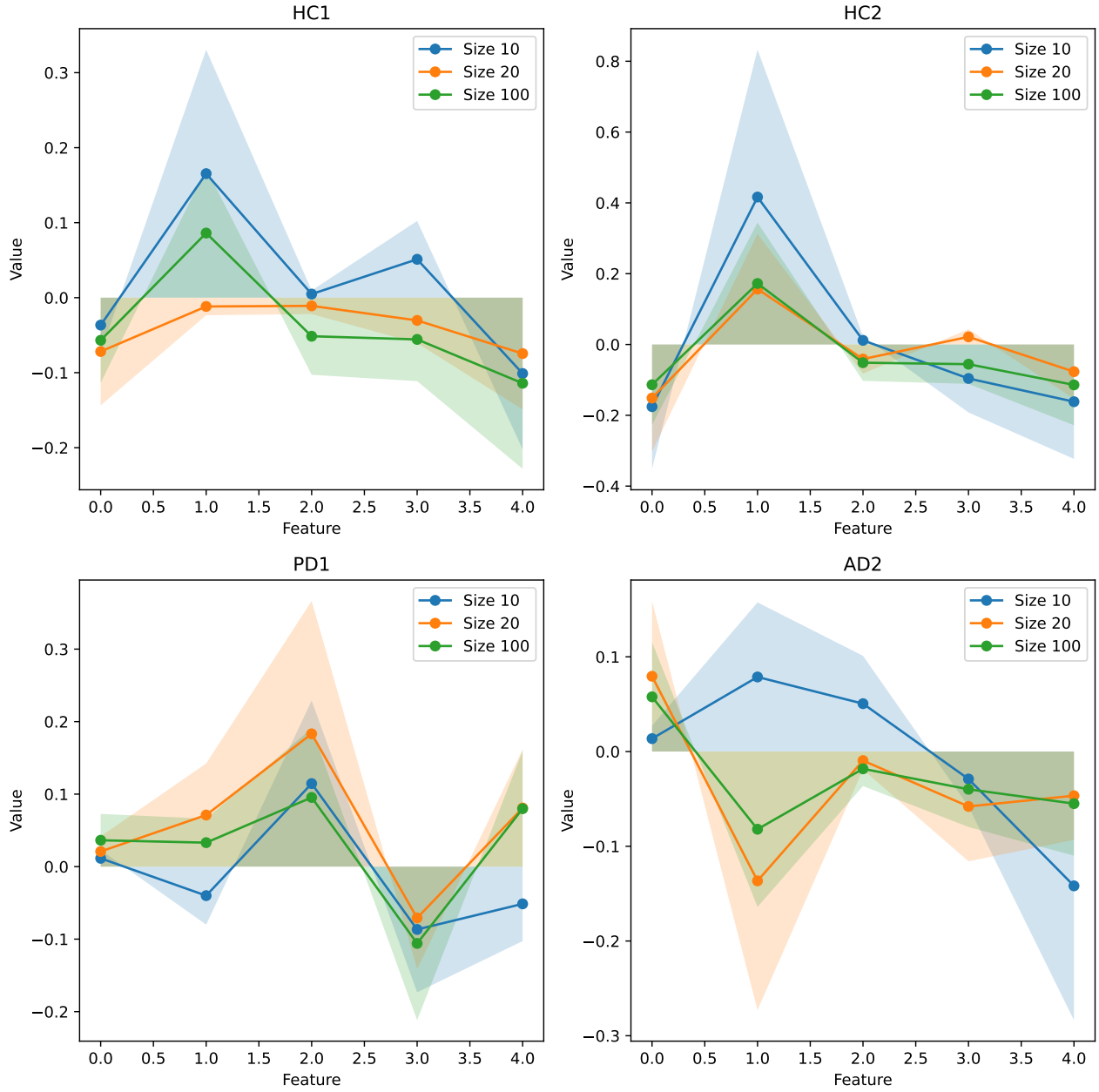

Figure 76: Mean (estimated mean - ground truth mean) as a function of the number of features of the centre-wise z-scored data.

D Case C

In this appendix several plots supplementing the results in Section 5.3 are presented, including the mean relevance profiles, eigenvalue spectrum of Lambda over the iterations and the difference between the data set mean and ground truth mean to show the stability of the data sets used for the experiments.

D.1 Results for the centre classification

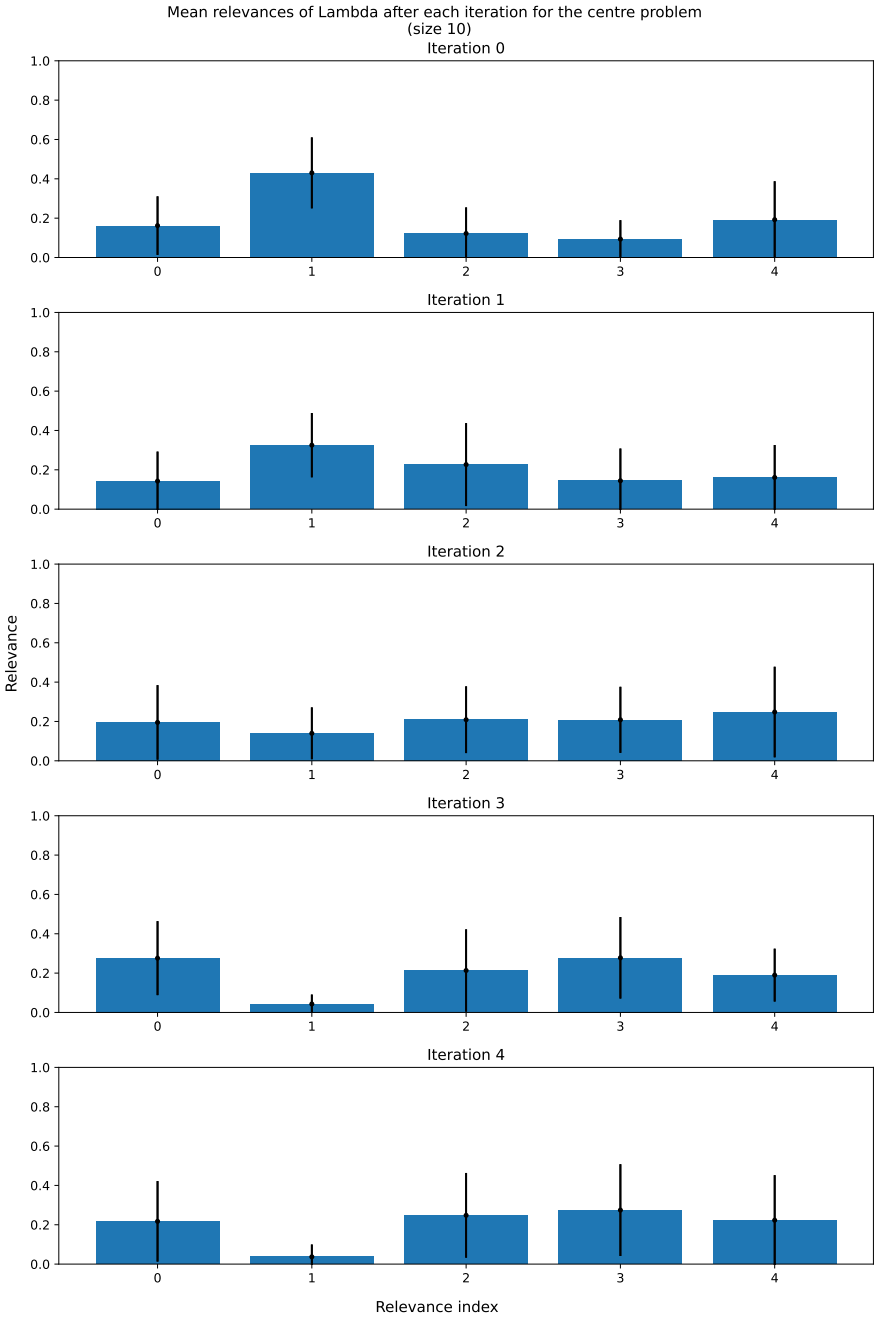

Figure 77: Mean relevance profile of case C with $n = 10$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

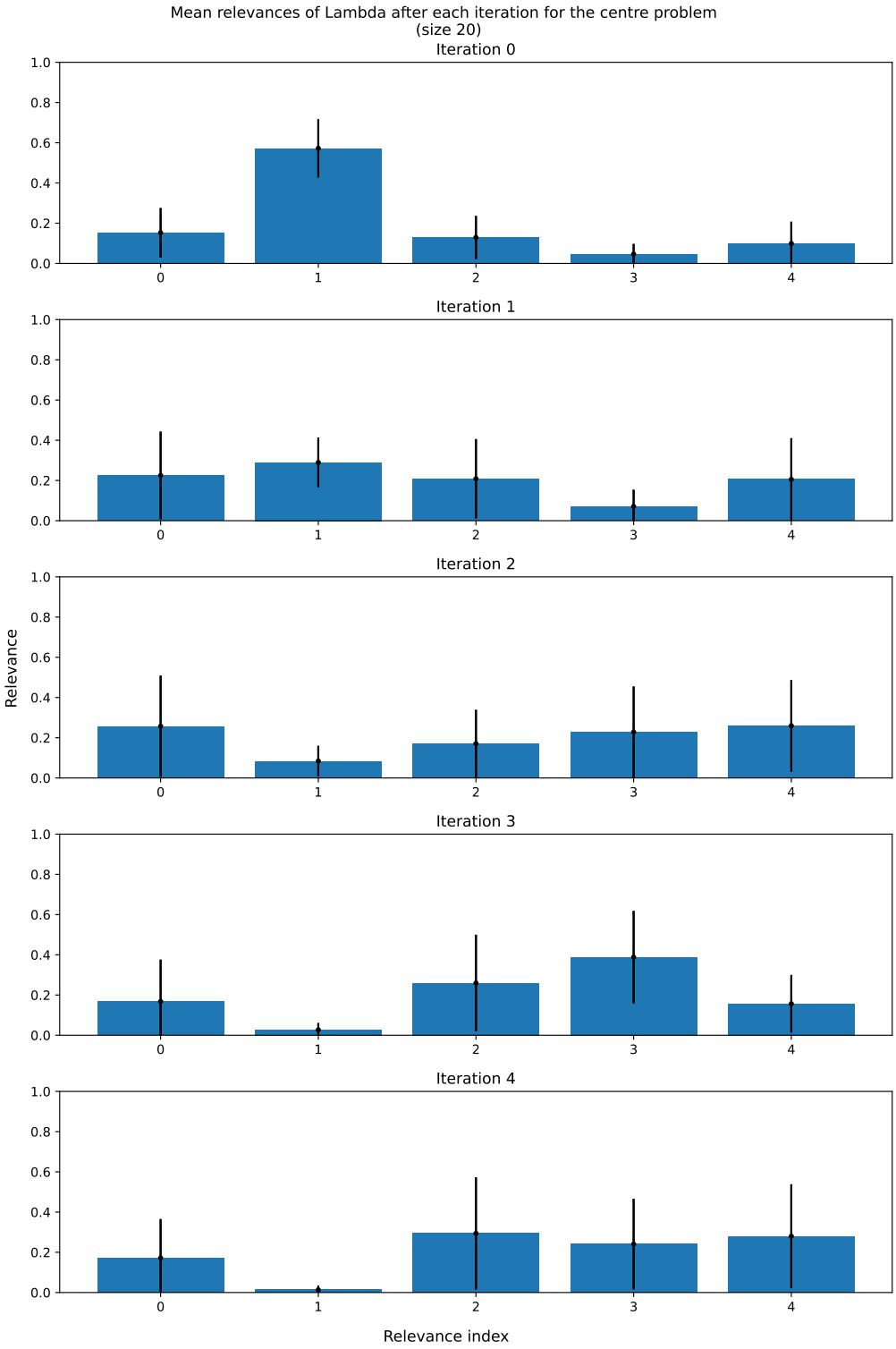

Figure 78: Mean relevance profile of case C with $n = 20$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

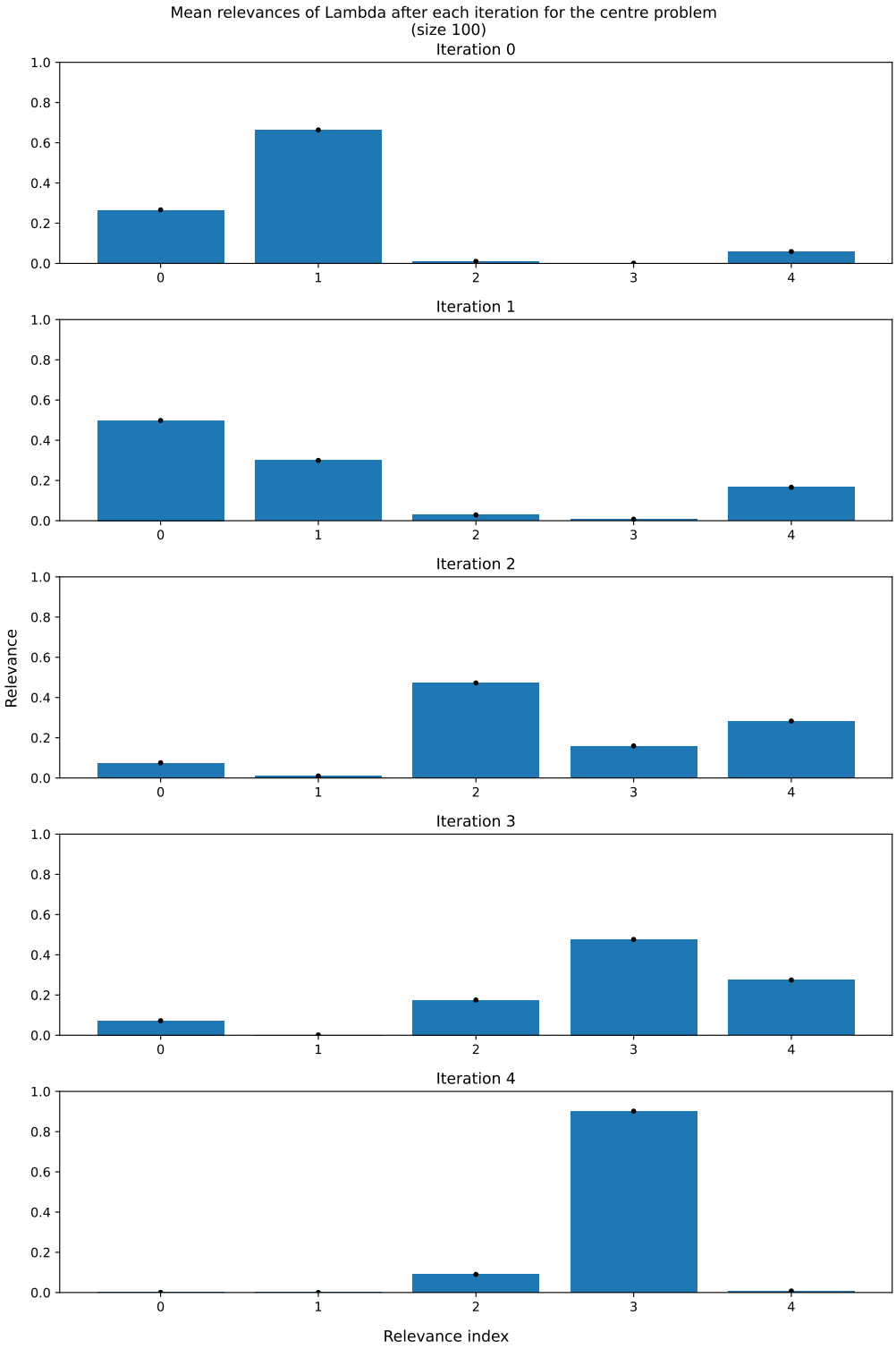

Figure 79: Mean relevance profile of case C with $n = 100$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

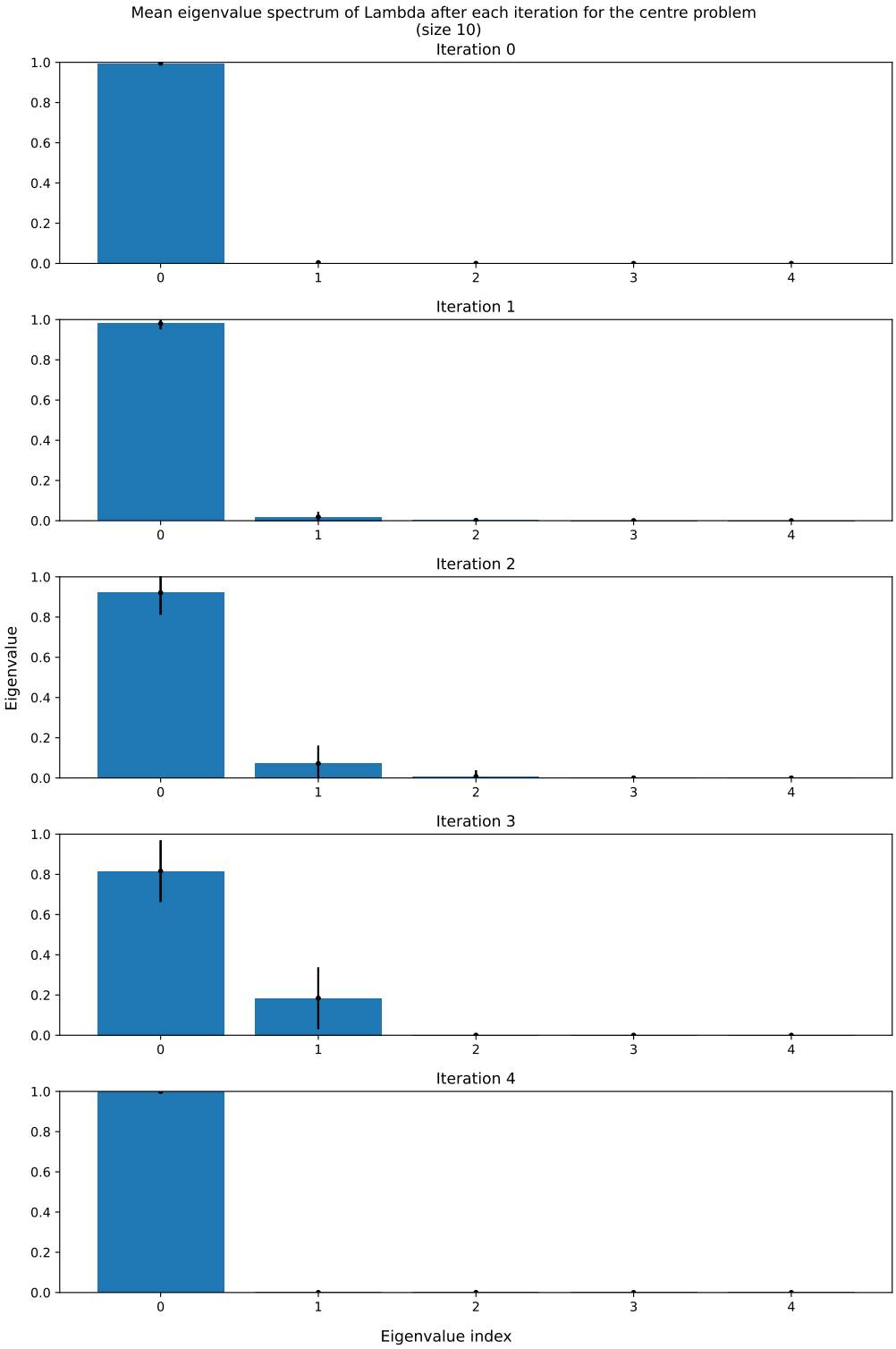

Figure 80: Mean eigenvalue spectrum of Lambda of case C with $n = 10$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

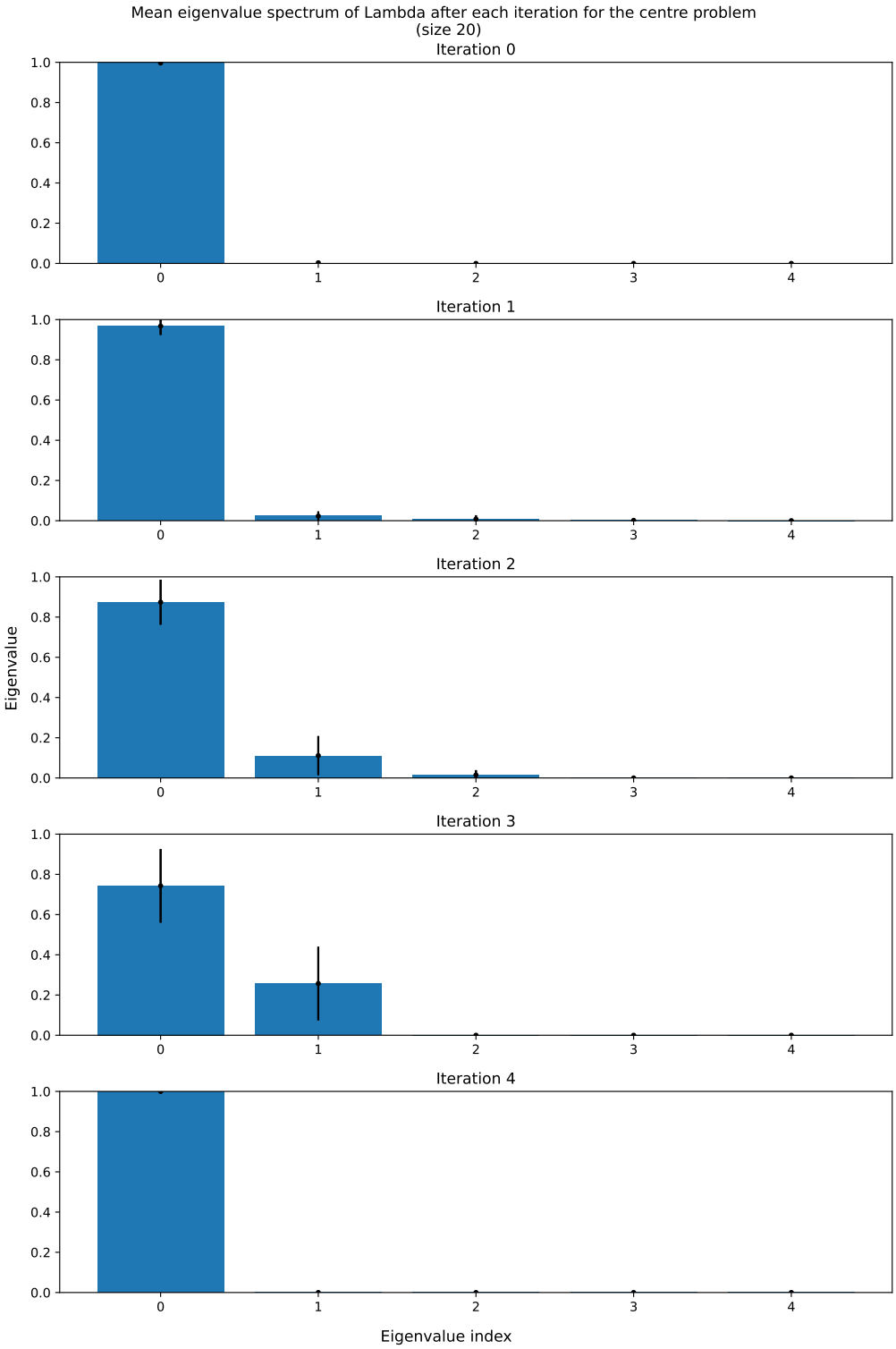

Figure 81: Mean eigenvalue spectrum of Lambda of case C with $n = 20$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

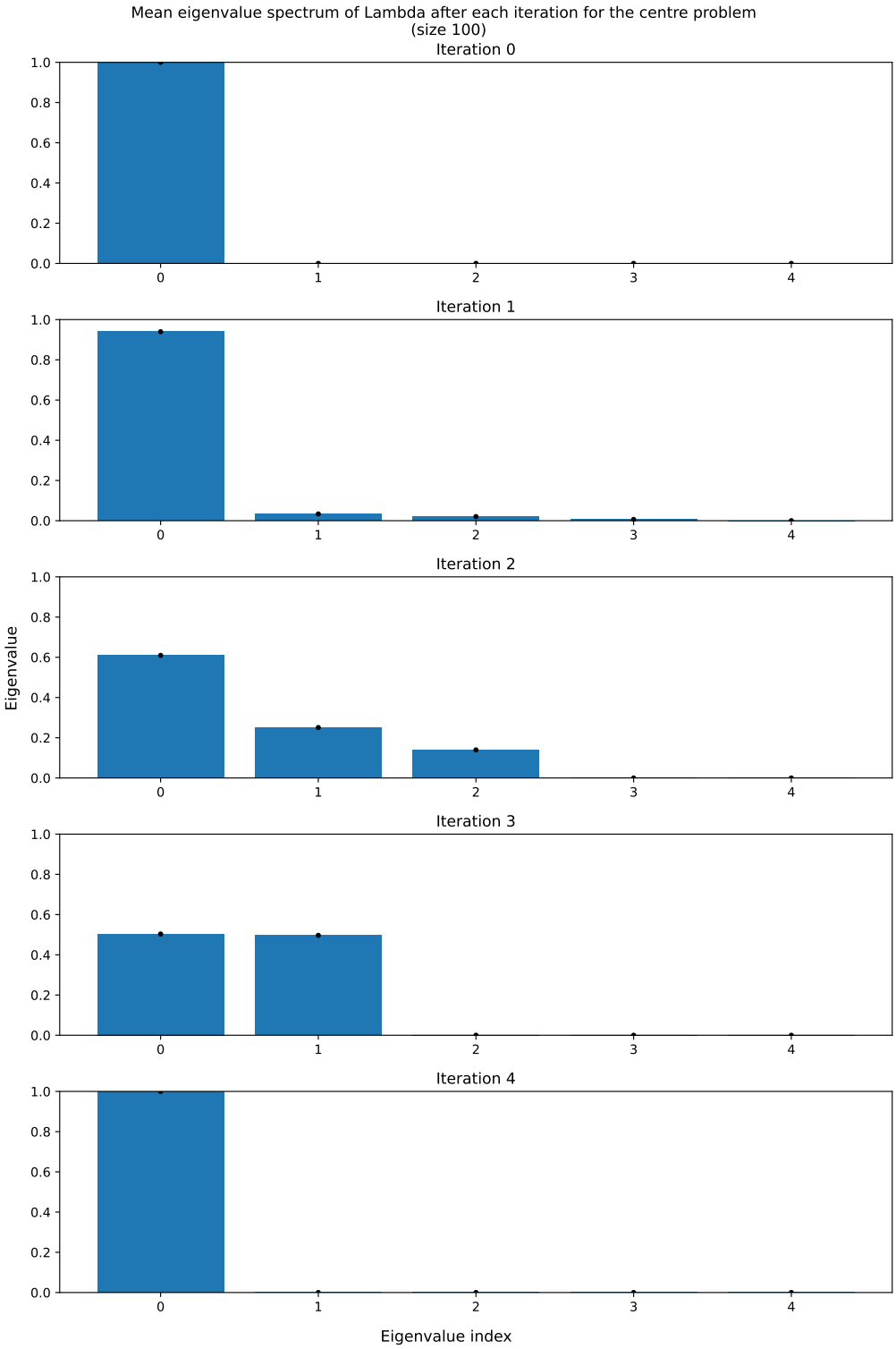

Figure 82: Mean eigenvalue spectrum of Lambda of case C with $n = 100$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

D.2 Results for the disease classification

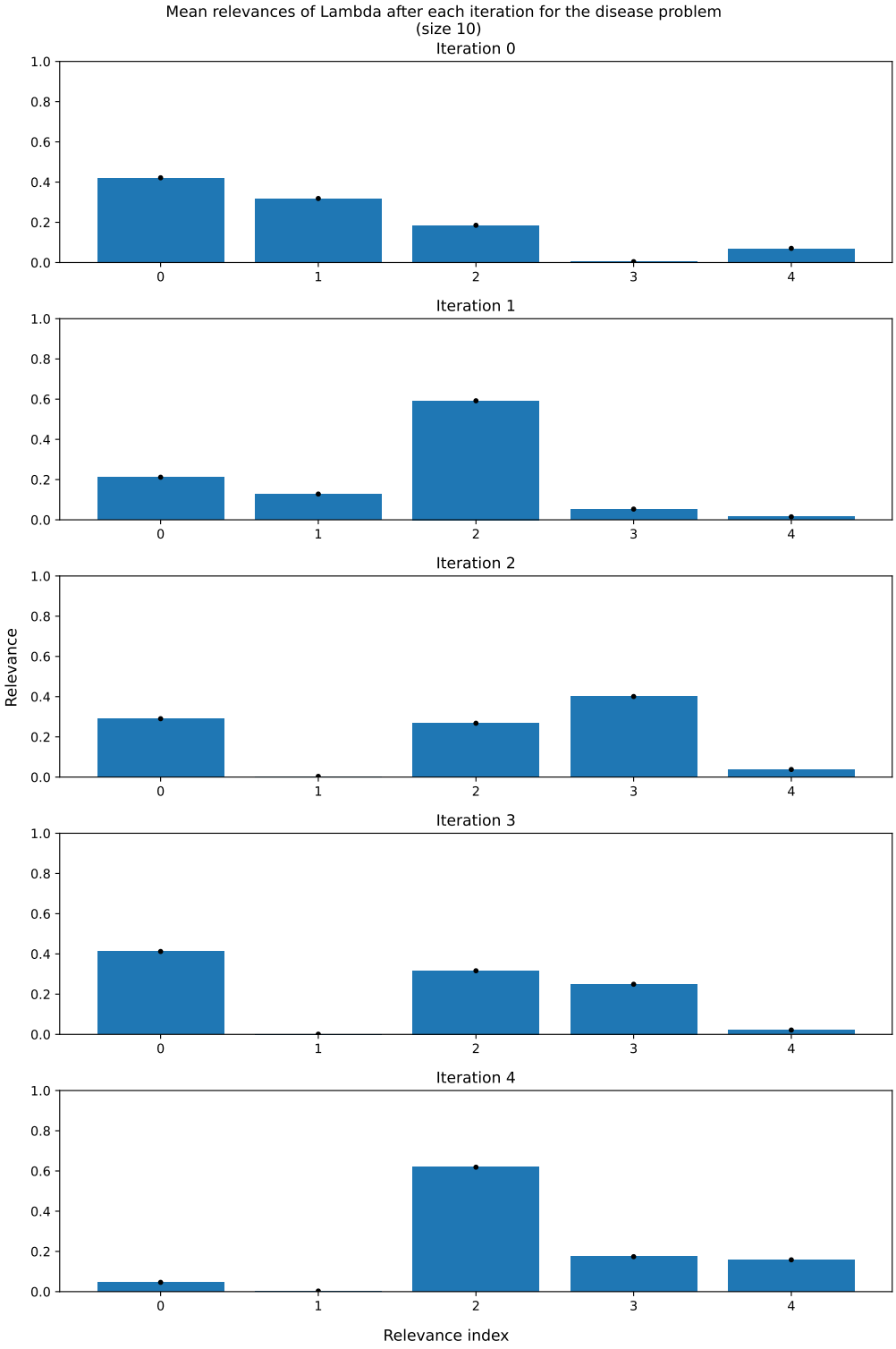

Figure 83: Mean relevance profile of case C with $n = 10$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

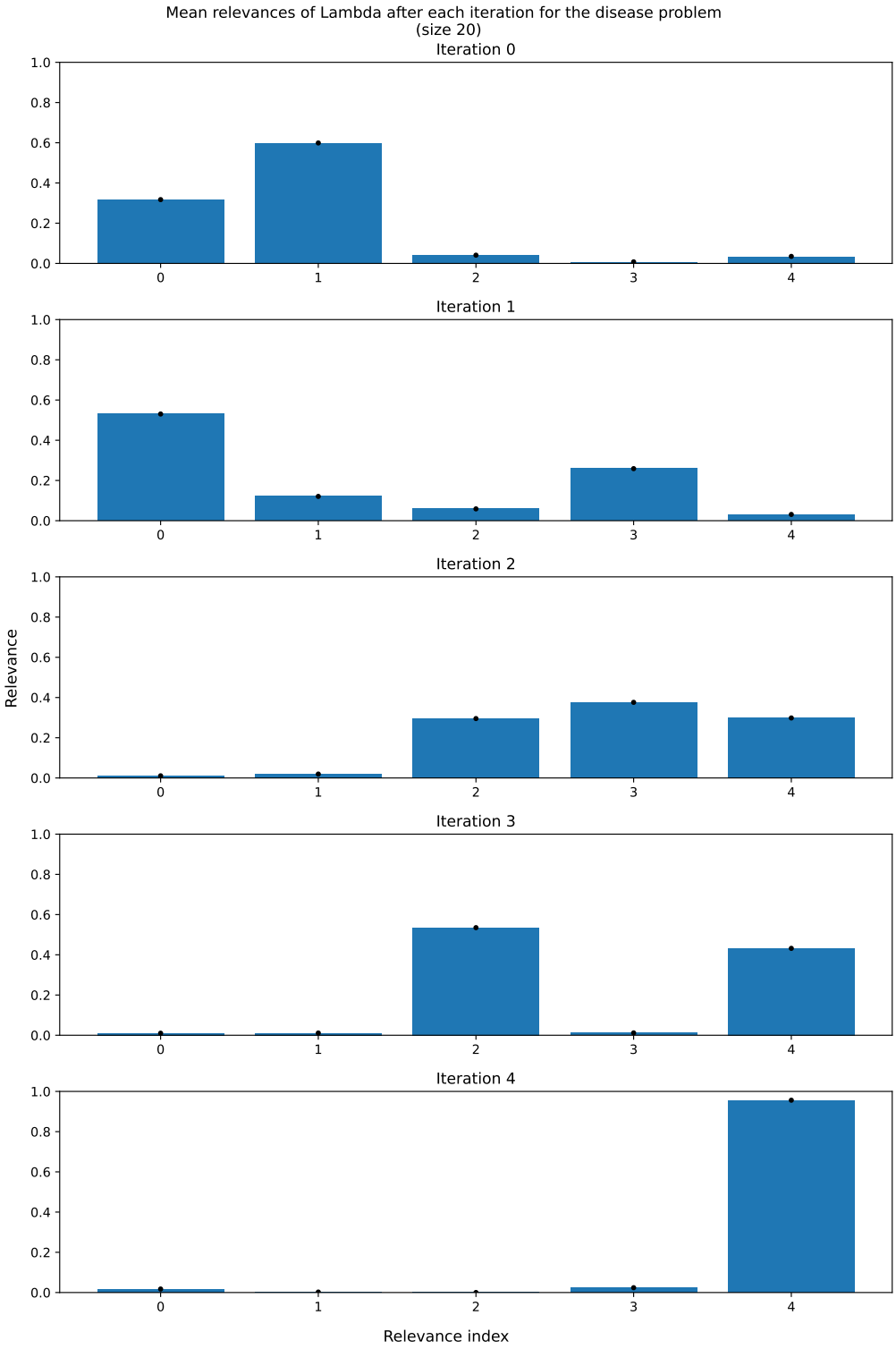

Figure 84: Mean relevance profile of case C with $n = 20$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

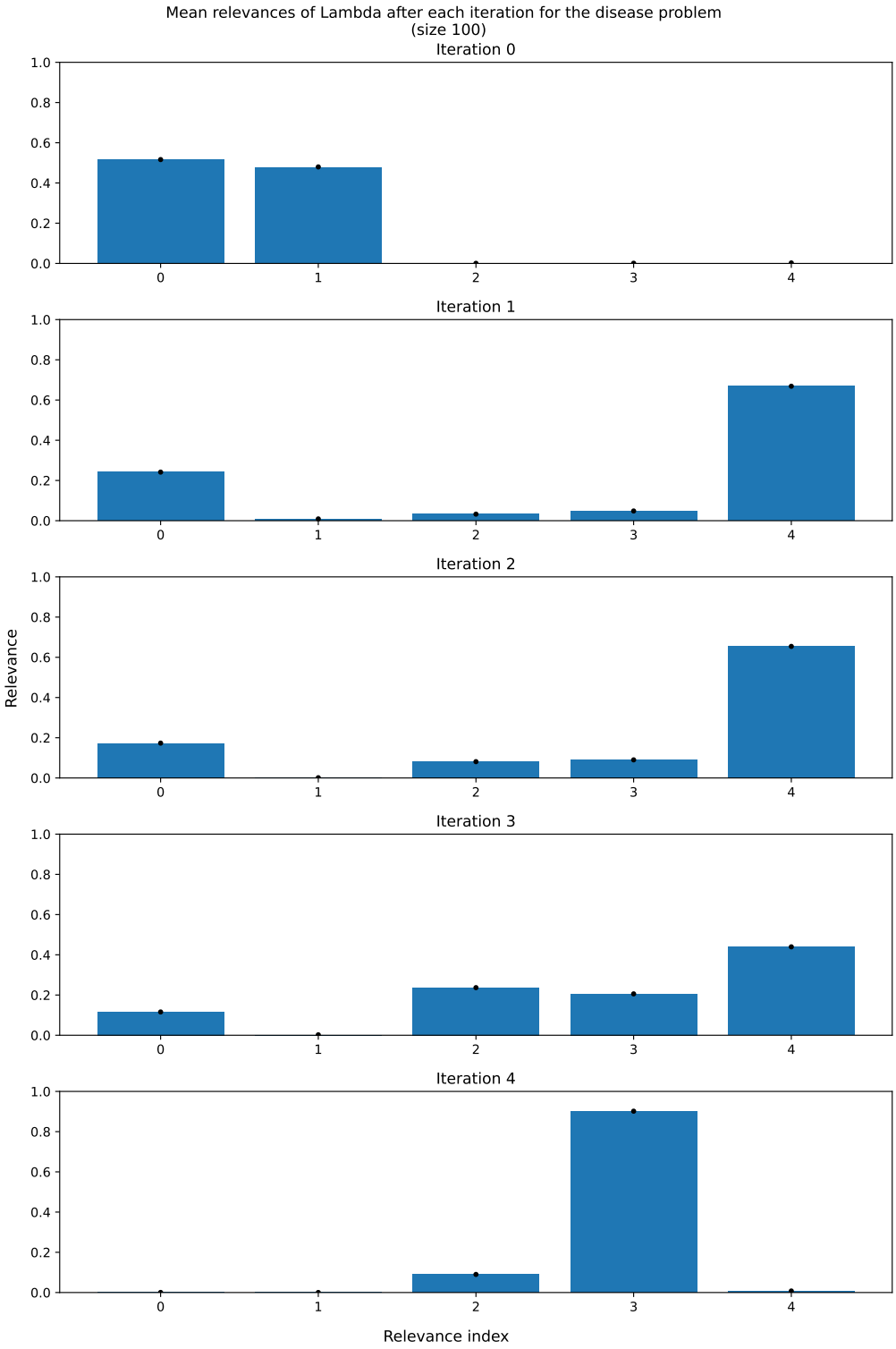

Figure 85: Mean relevance profile of case C with $n = 100$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

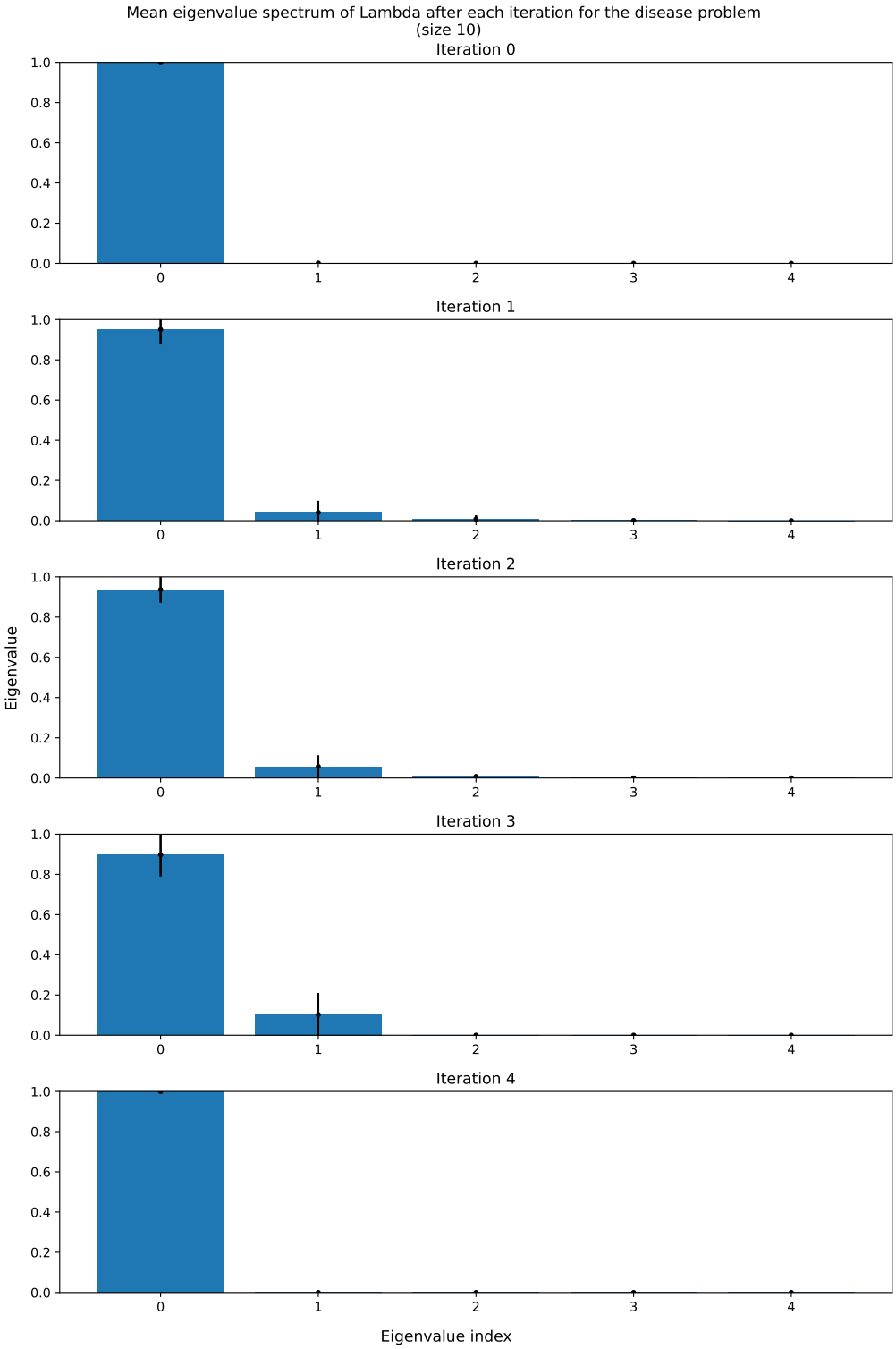

Figure 86: Mean eigenvalue spectrum of Lambda of case C with $n = 10$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

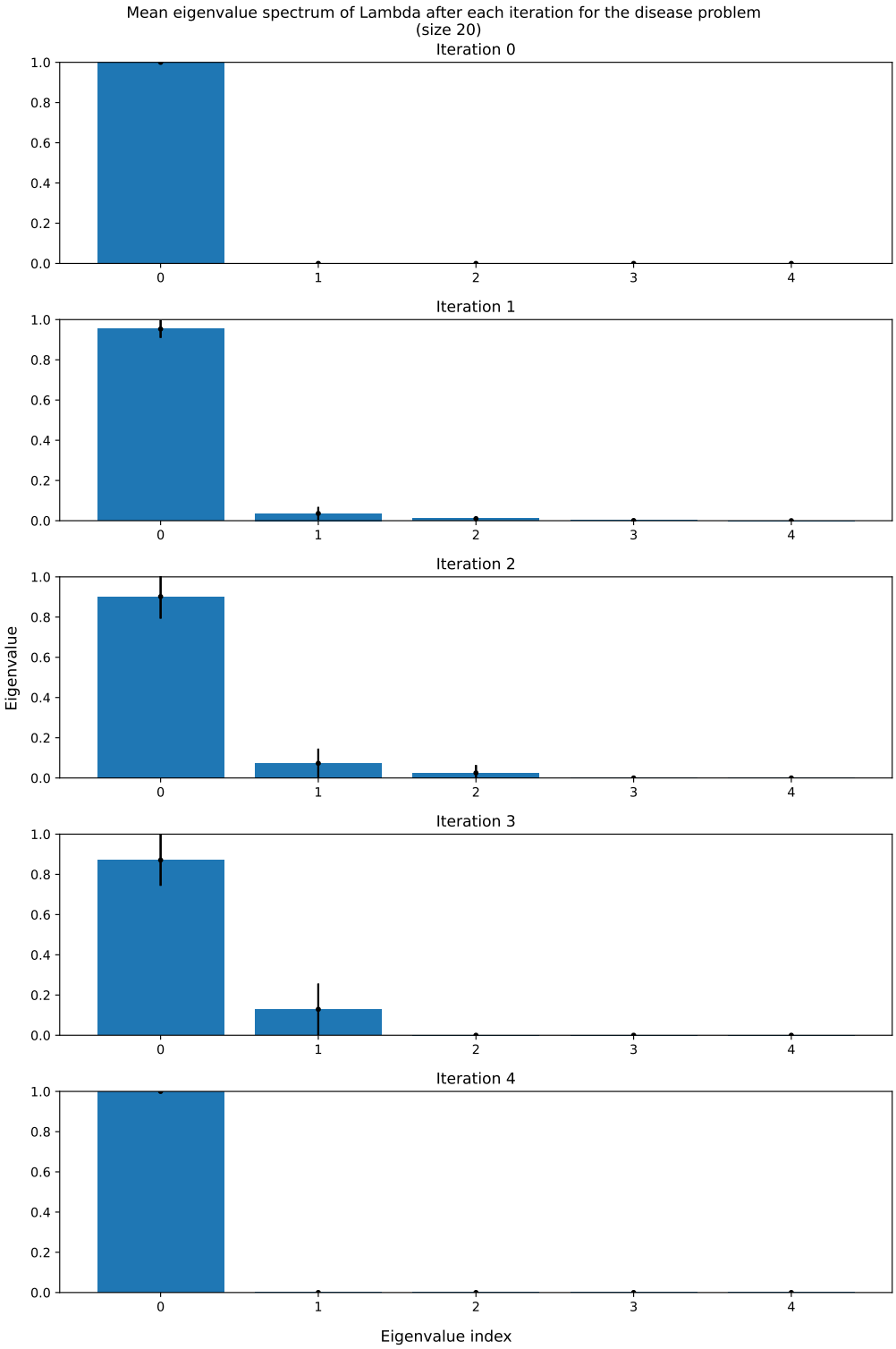

Figure 87: Mean eigenvalue spectrum of Lambda of case C with $n = 20$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

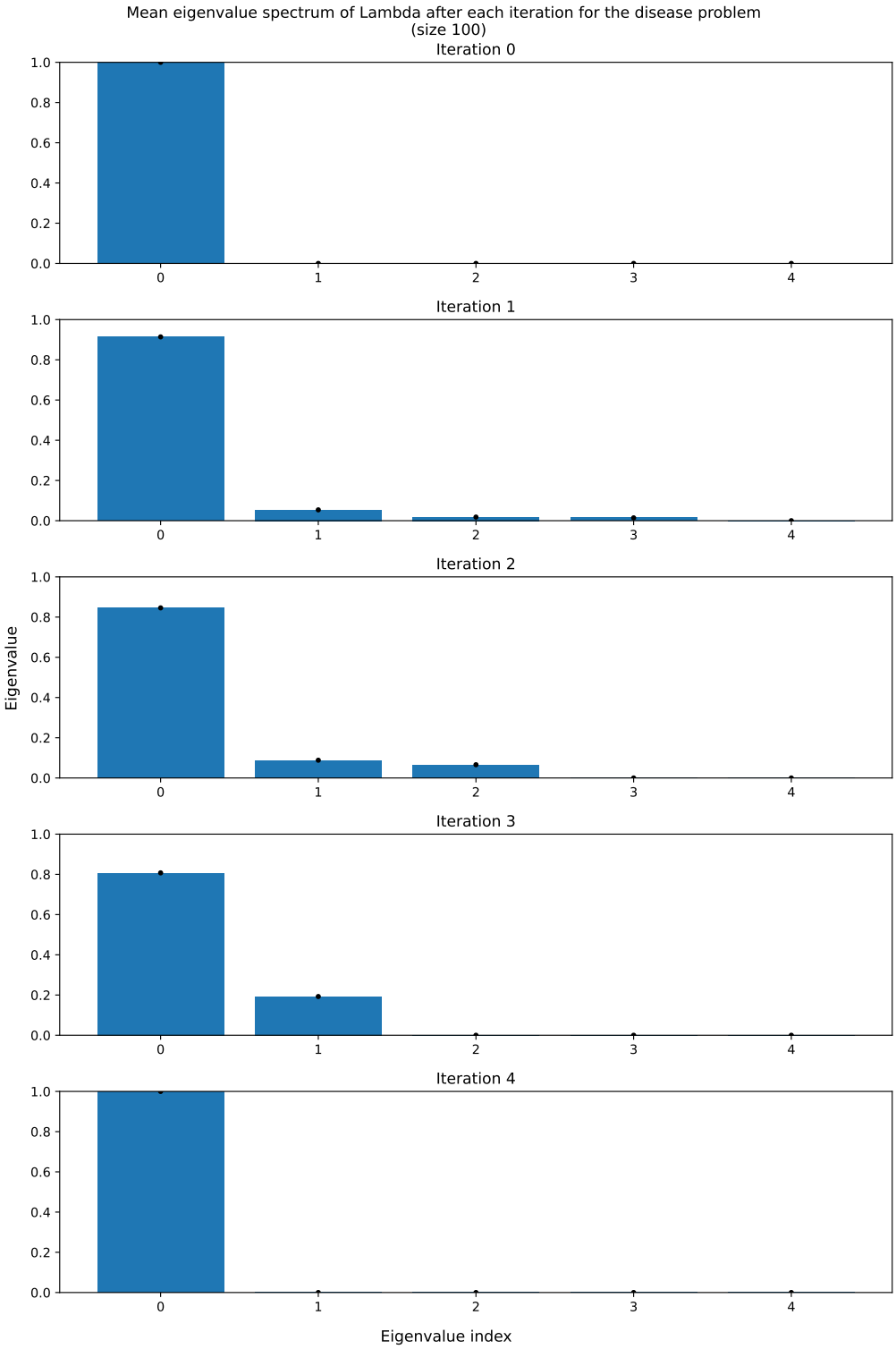

Figure 88: Mean eigenvalue spectrum of Lambda of case C with $n = 100$ over the iterations, where the 0-*th iteration* is the uncorrected GMLVQ system.

D.3 Centre-wise z-score transformed data: results for the centre classification

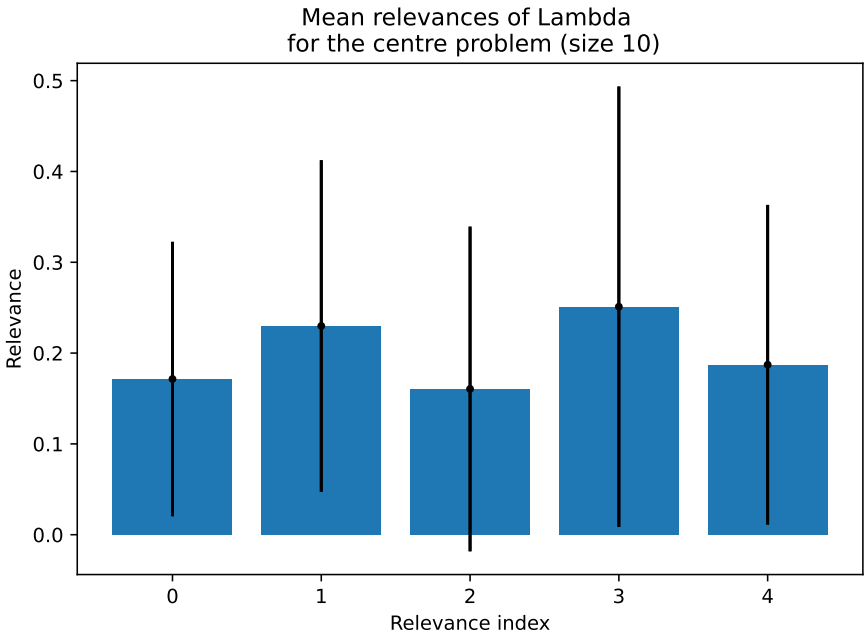

Figure 89: Mean relevance profile of case C with $n = 10$.

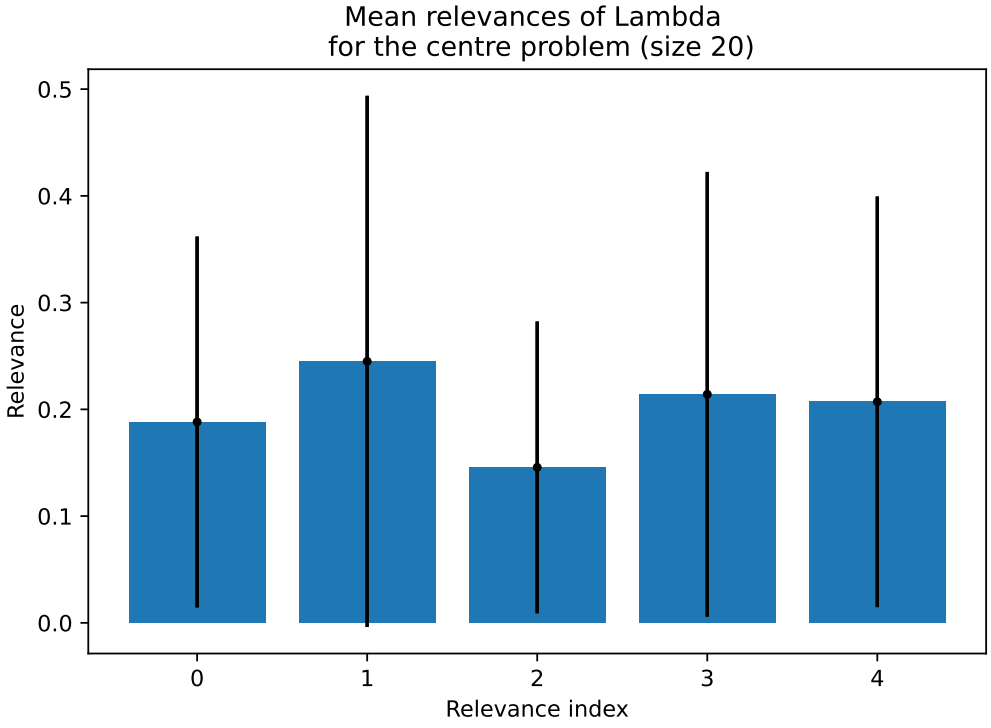

Figure 90: Mean relevance profile of case C with $n = 20$.

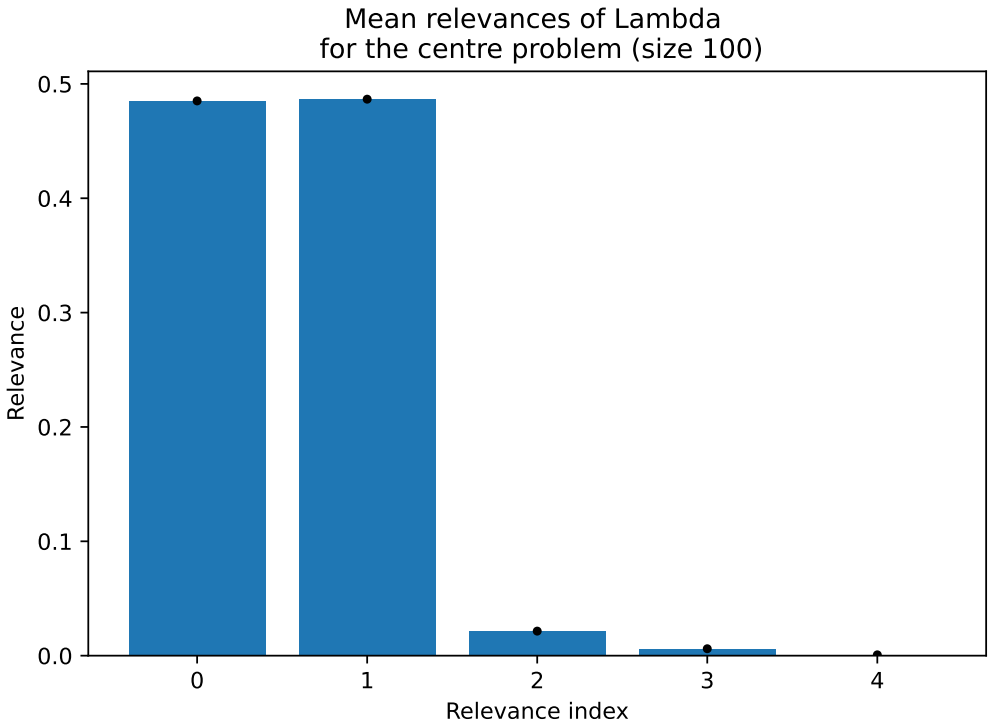

Figure 91: Mean relevance profile of case C with $n = 100$.

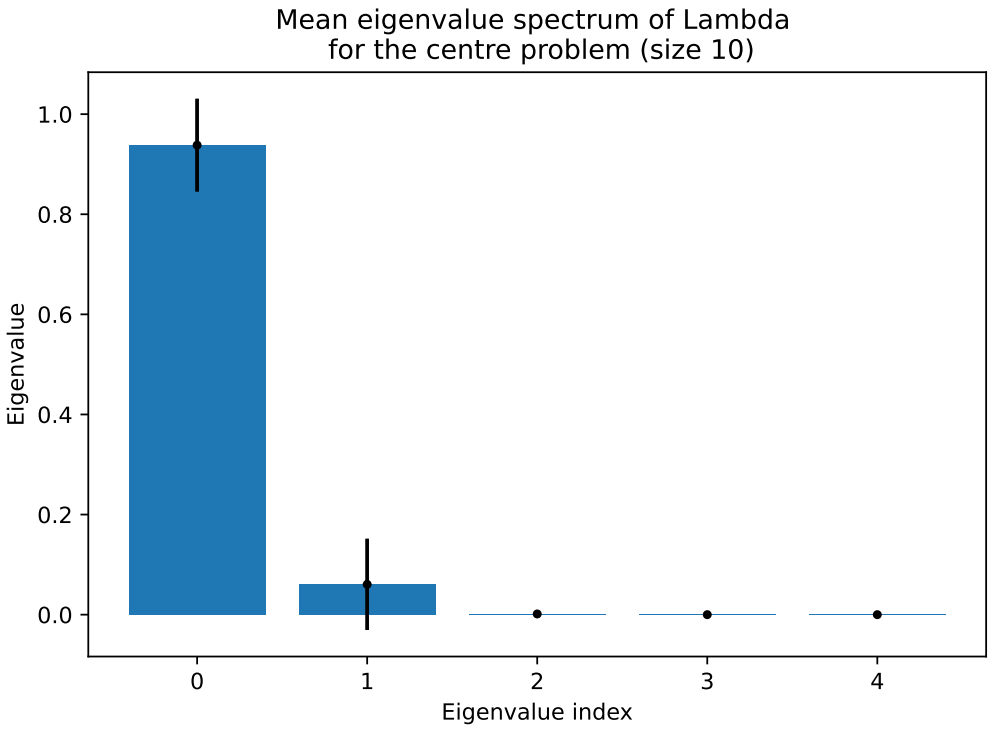

Figure 92: Mean eigenvalue spectrum of Lambda of case C with $n = 10$.

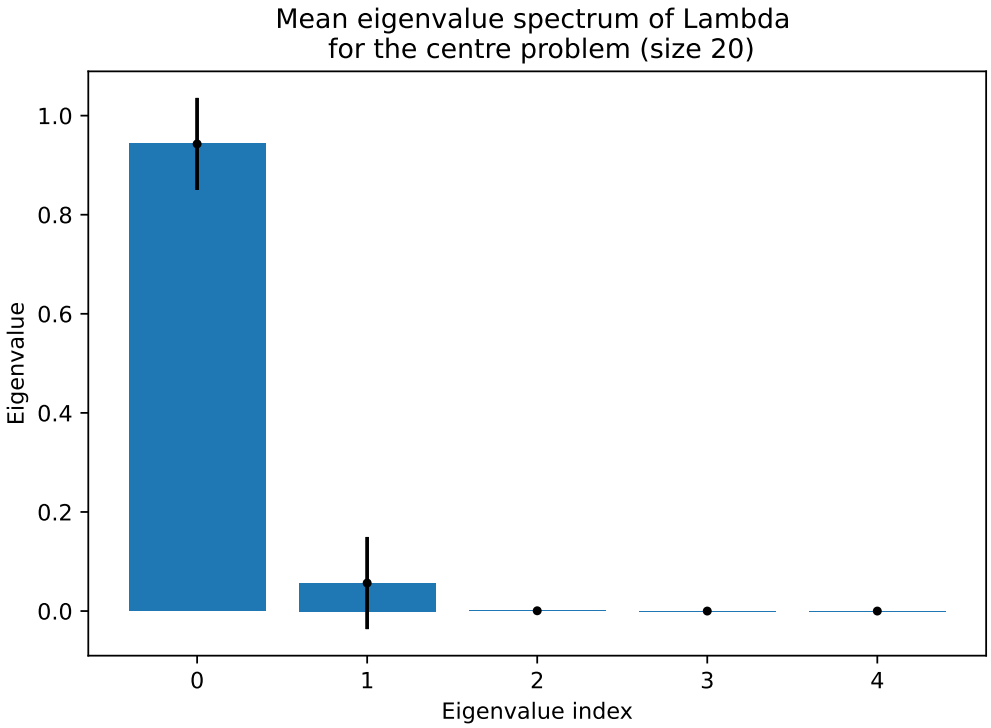

Figure 93: Mean eigenvalue spectrum of Lambda of case C with $n = 20$.

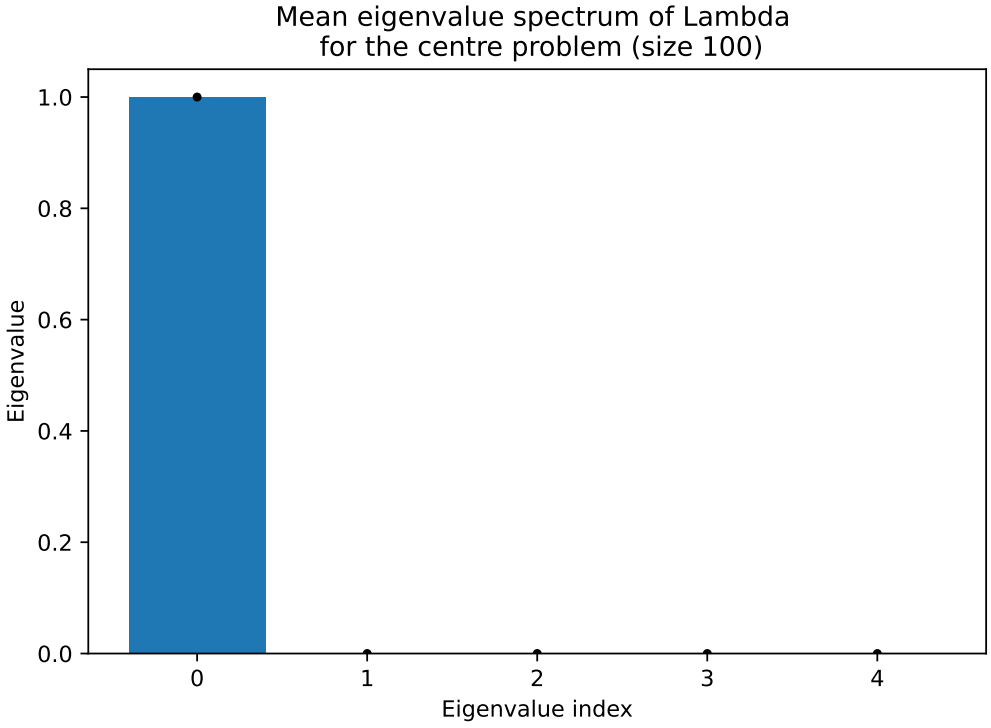

Figure 94: Mean eigenvalue spectrum of Lambda of case C with $n = 100$.

D.4 Centre-wise z-score transformed data: results for the disease classification

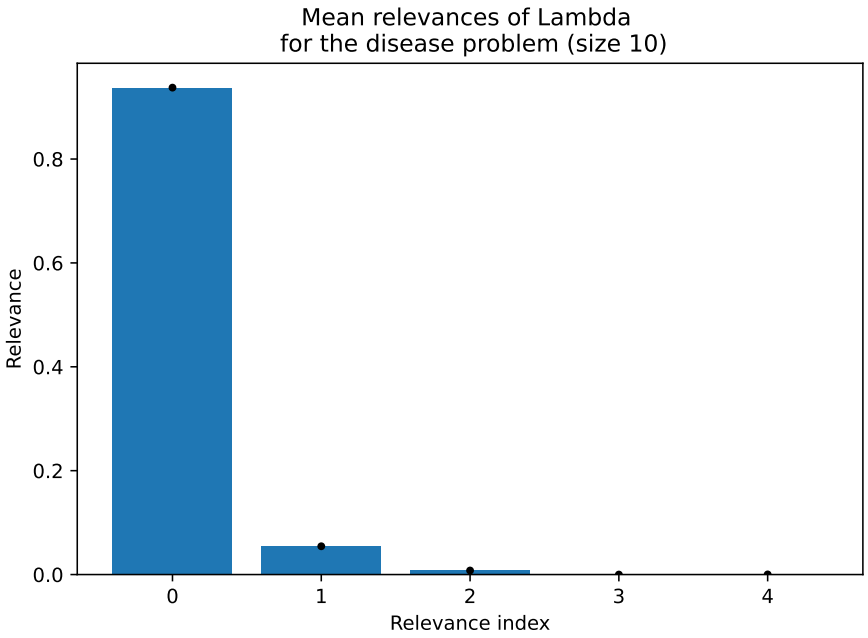

Figure 95: Mean relevance profile of case C with $n = 10$.

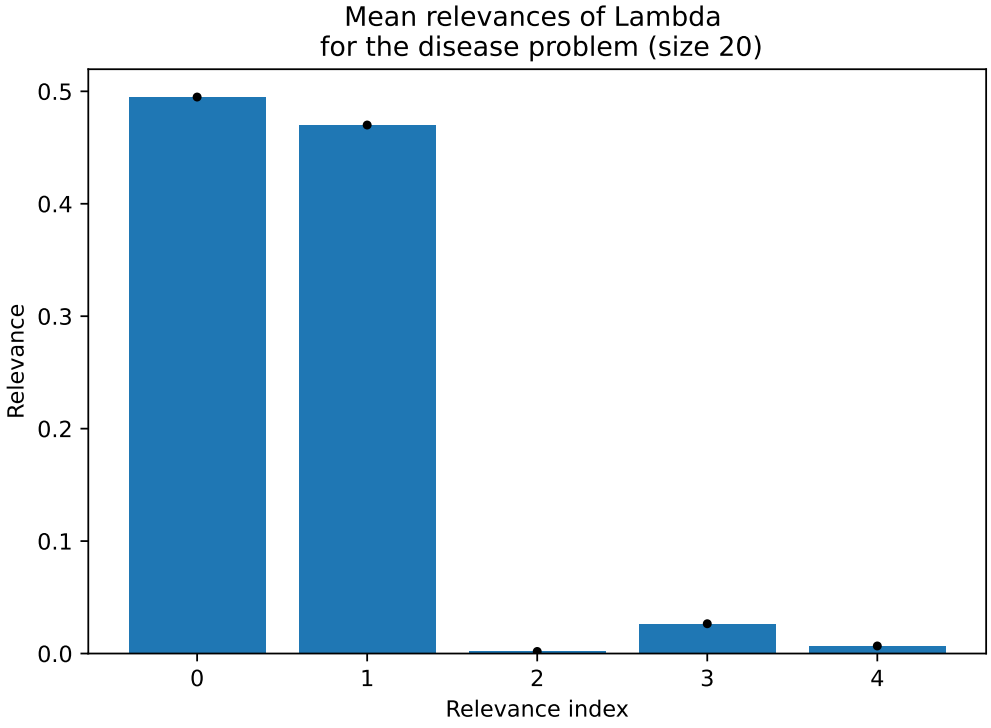

Figure 96: Mean relevance profile of case C with $n = 20$.

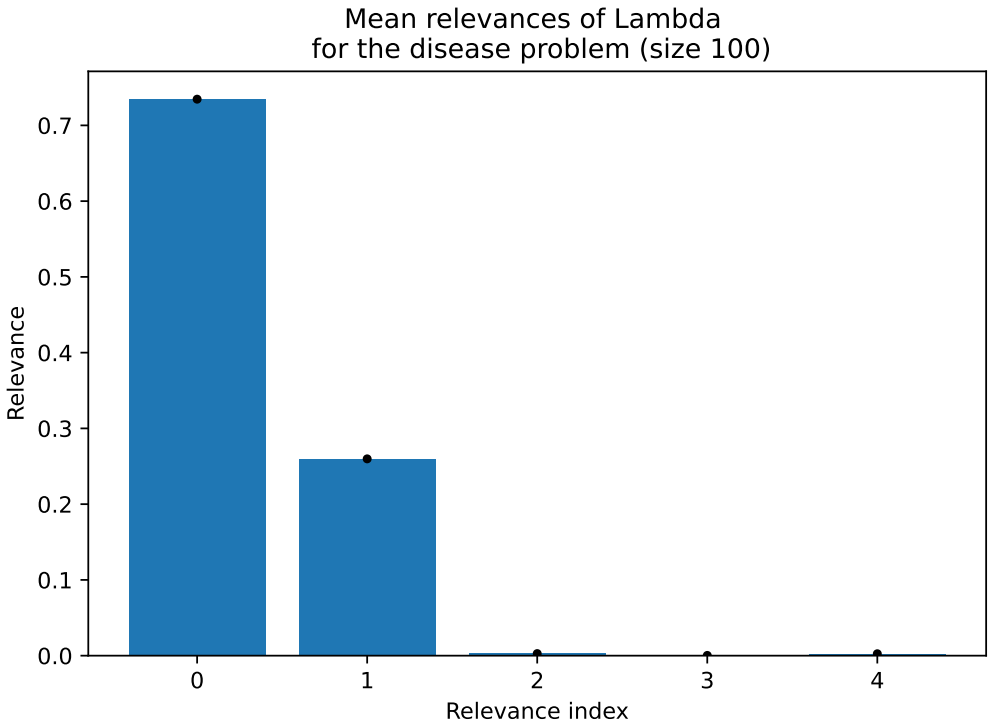

Figure 97: Mean relevance profile of case C with $n = 100$.

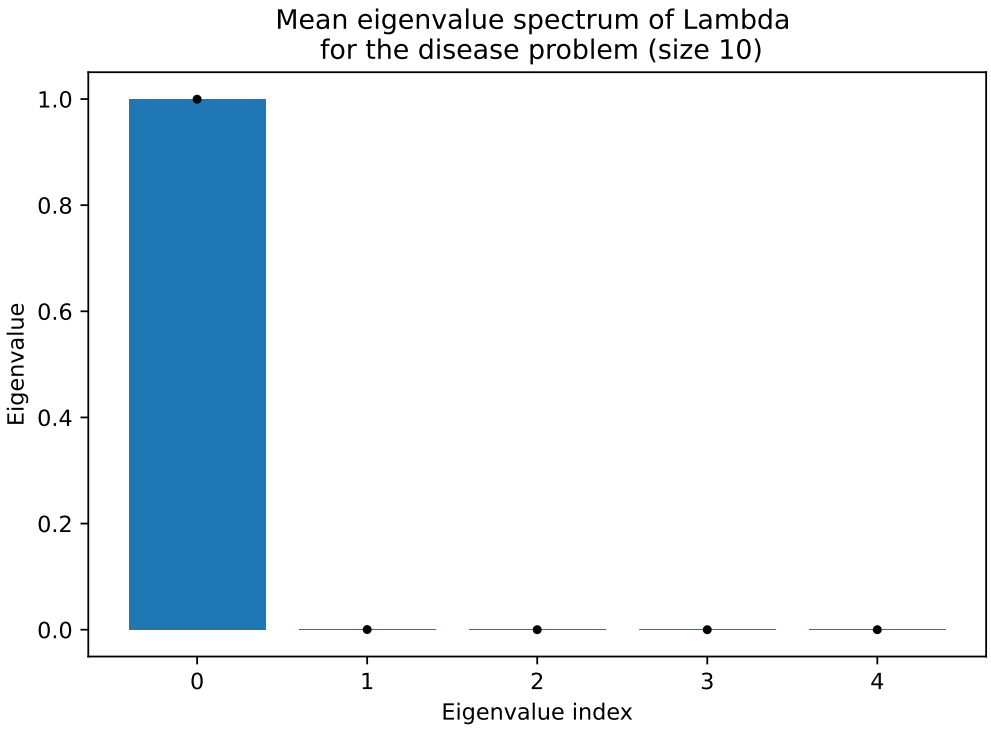

Figure 98: Mean eigenvalue spectrum of Lambda of case C with $n = 10$.

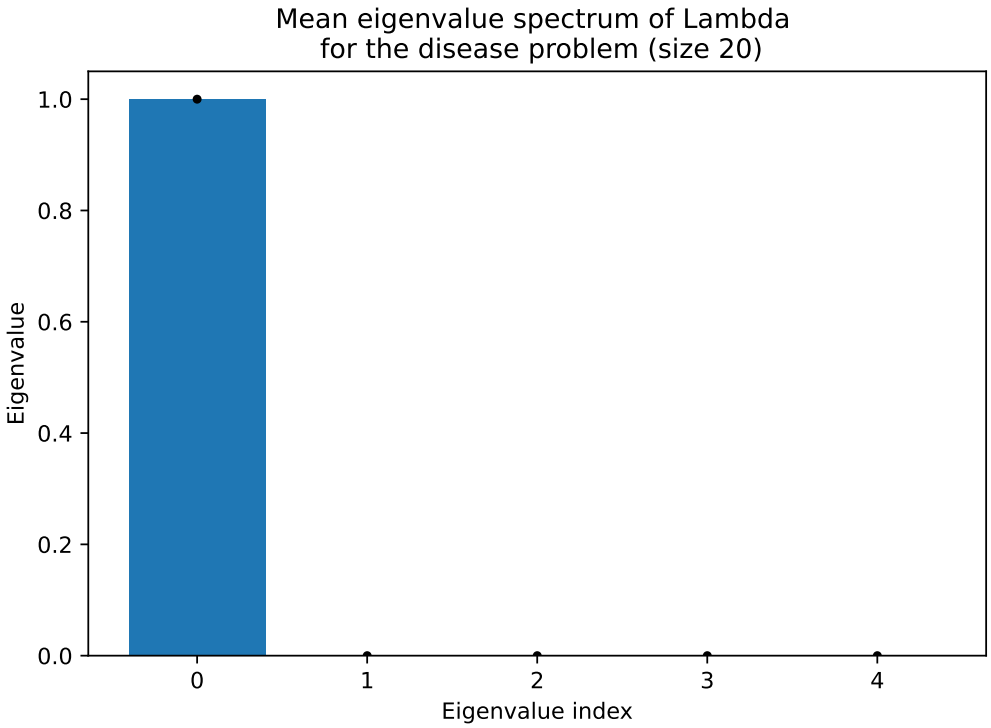

Figure 99: Mean eigenvalue spectrum of Lambda of case C with $n = 20$.

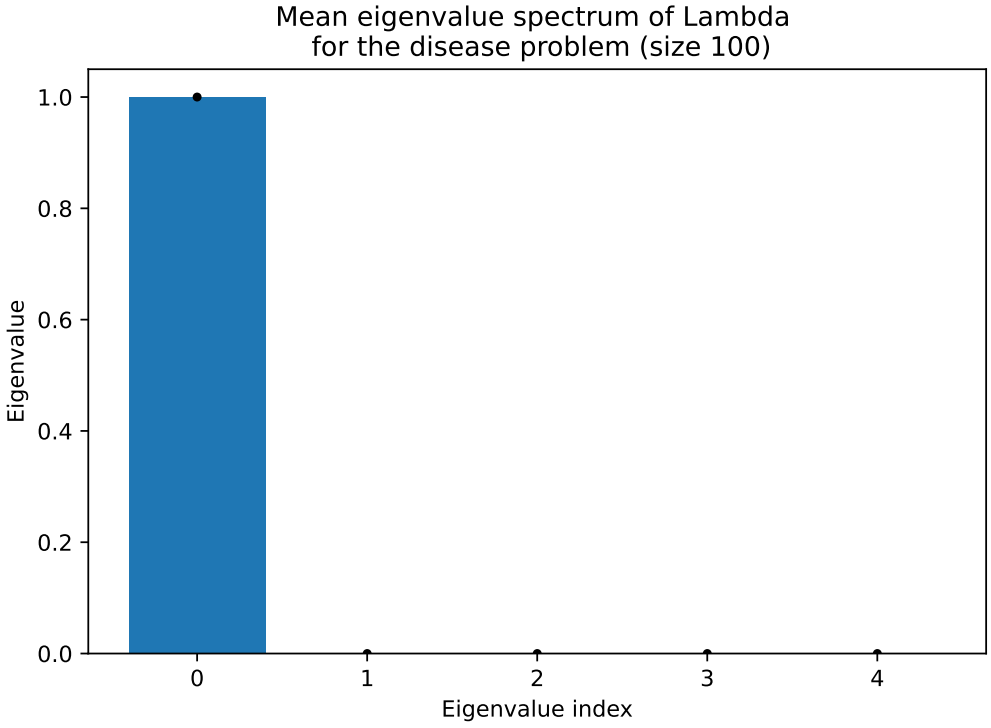

Figure 100: Mean eigenvalue spectrum of Lambda of case C with $n = 100$.

D.5 Data set stability

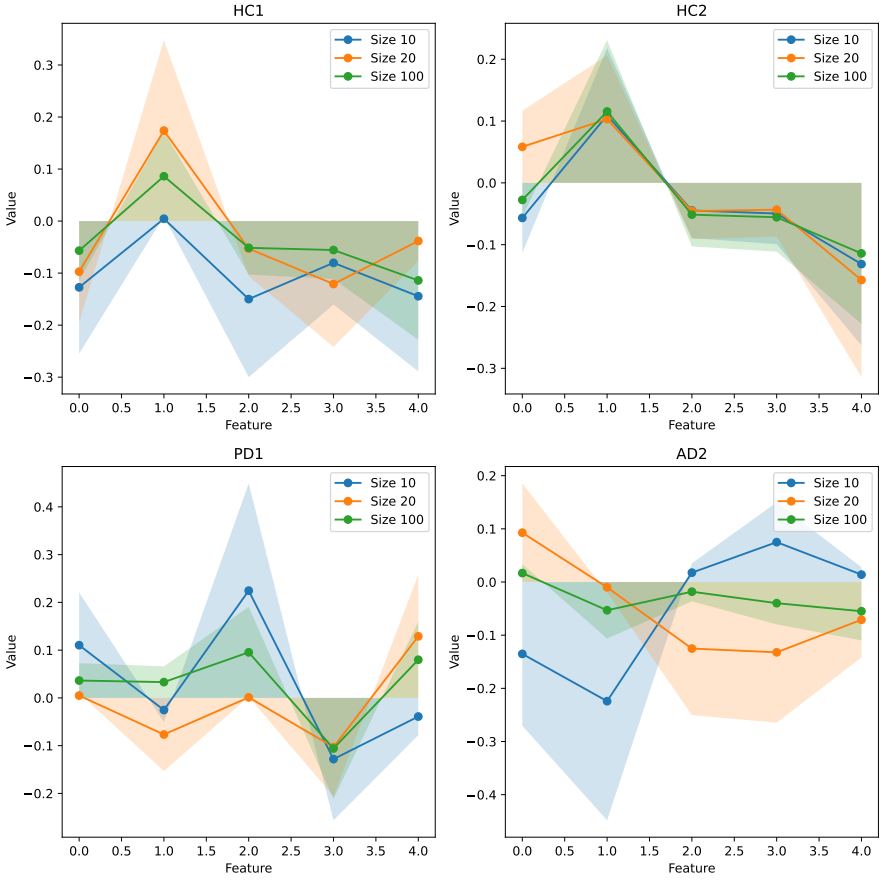

Figure 101: Mean (estimated mean - ground truth mean) as a function of the number of features.

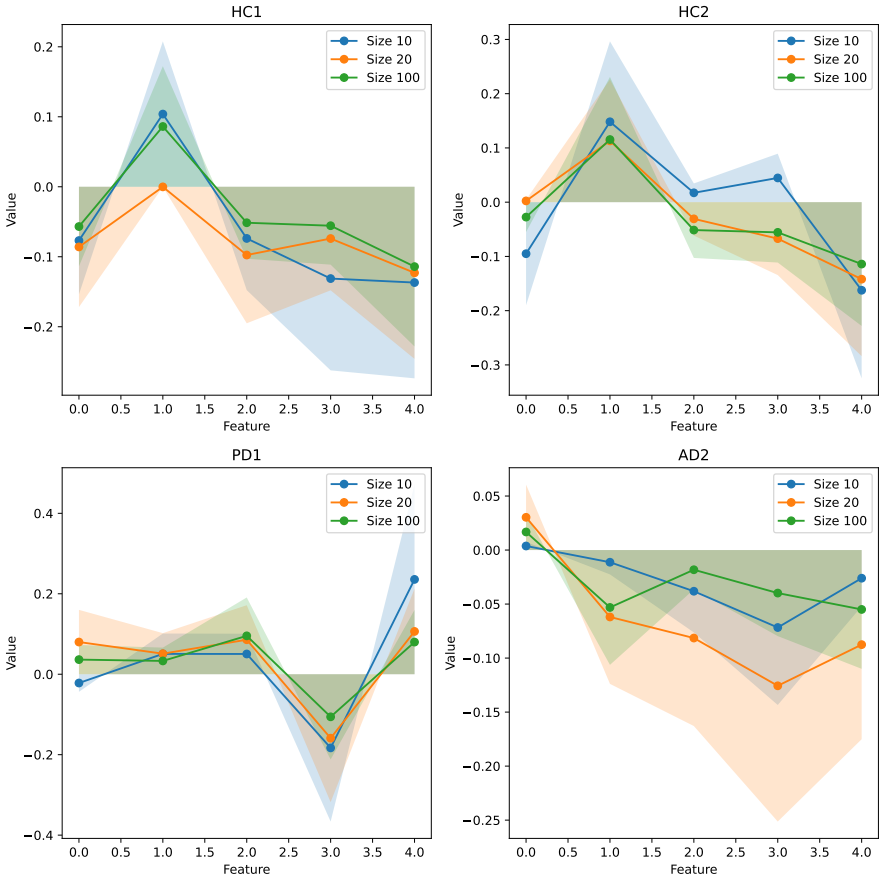

Figure 102: Mean (estimated mean - ground truth mean) as a function of the number of features of the centre-wise z-scored data.