Lecture "Digital Signal Processing"

Prof. Dr. Dietrich Klakow, Summer Term 2021

Assignment 10

Submission deadline: 28 June 2021, 23:59

Submission Instructions:

You have one week to solve the assignments.
The code should be well structured and commented. Do not use any Matlab-Toolbox or Python external libraries if it is not mentioned that you can use them.

- You are required to hand in your solutions in a group of two students.
- There are two parts in this assignment: theoretical part and practical part.
- The practical part that is solved with Python should be submitted as an ipynb file (Jupyter Notebook or Google Colab), where every function is written in a separate block.
- The theoretical part can either be written by hand and scanned, or typed with LaTeX in the text / markdown area of ipynb file.
- Submission of both parts should be done via Microsoft Teams by one of the group members.
- Submission should be named as: Ex10_matriculationnumber1_matriculationnumber2.zip

The submission should contain the following files:

- file "README" that contains an information on all team members:
  name, matriculation number, and email address.
- code files
- file "answers.pdf" which contains answers to the questions appearing in the exercise sheet. *Note: If you use ipynb file, you don't have to submit "answers.pdf". You can embed your scanned copy or write your answers in the text / markdown area.*

1 (1.5P) Discriminant Function

1. What is the likelihood expression of a class k given input x, considering $p_k(x)$ is the class density of x in class k, π_k is the prior probability, and $K \geq 2$ is the total number of classes. *Hint: Bayes' theorem.*

2. Find the discriminant function that tells in which class the input x belongs to, by substituting $p_k(x)$ with a Gaussian distribution of mean μ_k and variance σ_k^2, and show that it is linear in x.

2 (3P) Fisher's LDA Derivation

Consider a two-class problem in which there are N_1 vector points of class C_1 and N_2 vector points of class C_2. Each vector point $\mathbf{x}_n$ has D-dimensional features. The mean vector of class C_k is given by:

$$\mathbf{m}_k = \frac{1}{N_k} \sum_{n \in C_k} \mathbf{x}_n \tag{1}$$

and the within-class variance s_k^2 is given by:

$$s_k^2 = \sum_{n \in C_k} (\mathbf{w}^T \mathbf{x}_n - \mathbf{w}^T \mathbf{m}_k)^2 \tag{2}$$

where $\mathbf{w}$ is the weight vector. The Fisher criterion is defined to be as follows:

$$J(\mathbf{w}) = \frac{(m_2 - m_1)^2}{s_1^2 + s_2^2} \tag{3}$$

where

$$m_2 - m_1 = \mathbf{w}^T (\mathbf{m}_2 - \mathbf{m}_1) \tag{4}$$

1. From the Fisher criterion given above, derive the between-class covariance matrix S_B, and the total within-class covariance matrix S_W.

2. Additionally, we want to find a scalar λ and the corresponding vector $\mathbf{v}$ such that

$$S_B \, \mathbf{v}_i = \lambda_i \, S_W \, \mathbf{v}_i \text{ for i = 1, ..., n.} \tag{5}$$

   This problem is called the generalized eigenvalue problem. By rewriting the Fisher criterion to be

$$J(\mathbf{w}) = \frac{\mathbf{w}^T S_B \mathbf{w}}{\mathbf{w}^T S_W \mathbf{w}}, \tag{6}$$

   solve the generalized eigenvalue problem to obtain the optimal $\mathbf{w}$. *Hint: this article may help.*

3 (5.5P) Fisher's LDA Multi-classes Implementation

Allowed Python libraries: numpy, math, and matplotlib.pyplot, as well as requests to directly download the dataset from a provided url. Matlab libraries would be the equivalent ones.

seeds_dataset.txt from this site will be used for this implementation. It contains 210 instances belonging to 3 different classes. Use the last column as the class label, where 1 refers to class Kama, 2 to class Rosa and 3 to class Canadian. Each instance has 7 measured parameters. The detailed description of the dataset can be read here.

1. Write a function that computes the mean vectors of each class in the dataset
2. Write a function to find the within-class scatter matrix S_W
3. Write a function to find the between-class scatter matrix S_B
4. Solve the generalized eigenvalue problem: $S_B \, \mathbf{v}_i = \lambda_i \, S_W \, \mathbf{v}_i$
5. Using the first two eigenvectors that correspond to the largest two eigenvalues, represent the data in two dimensions. *Hint: $Y = XW$ where X is the original feature matrix, W the matrix containing the eigenvectors, and Y is the low dimensional representation of the feature matrix.*
6. Plot Y using different colors for different classes. What do you observe?