HW3 SPR Fall 2022

Deadline 12 Azar

In this homework, you have to implement Bayesian, Quadratic Multiclass Classification, and Density Estimation.

Bayesian Classification

Datasets:BC-Train1, BC-Test1, BC-Train2, BC-Test2

- In this part you have two datasets (Dataset1= BC-Train1, BC-Test1, Dataset2 = BC-Train2, BC-Test2). Each dataset contains two classes and each class generated from one Gaussian distribution. In this part you have to construct two Bayesian classifiers for each dataset so as to classify both train and test data.

- Classify both the train and test datasets with Bayesian classifier and calculate both accuracies, precisions, recalls and f1s (for each dataset separately).
- Plot the decision boundary and classification results while representing the misclassified samples with a different color or shape (for each dataset separately).
- Plot estimated PDFs. (3D for each dataset separately)
- Contour estimated PDFs along with the decision boundary. (2D for each dataset separately)
- What is the main difference between two datasets? Explain your answer using your results and plots.

Quadratic Multiclass Classification

- For this part of the homework, you have to generate two datasets for three classes each with 500 samples from three Gaussian distributions for each dataset described below:

Dataset#1:

$$\text{Class1:}\quad \mu = \begin{pmatrix}3\\6\end{pmatrix} \quad \sum = \begin{pmatrix}1.5 & 0\\0 & 1.5\end{pmatrix}$$
$$\text{Class2:}\quad \mu = \begin{pmatrix}5\\4\end{pmatrix} \quad \sum = \begin{pmatrix}2 & 0\\0 & 2\end{pmatrix}$$
$$\text{Class3:}\quad \mu = \begin{pmatrix}6\\6\end{pmatrix} \quad \sum = \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}$$

Dataset#2:

$$\text{Class1:}\quad \mu = \begin{pmatrix}3\\6\end{pmatrix} \quad \sum = \begin{pmatrix}1.5 & 0.1\\0.1 & 0.5\end{pmatrix}$$
$$\text{Class2:}\quad \mu = \begin{pmatrix}5\\4\end{pmatrix} \quad \sum = \begin{pmatrix}1 & -0.20\\-0.20 & 2\end{pmatrix}$$
$$\text{Class3:}\quad \mu = \begin{pmatrix}6\\6\end{pmatrix} \quad \sum = \begin{pmatrix}2 & -0.25\\-0.25 & 1.5\end{pmatrix}$$

- Consider 80% of the data in each class for the training and 20% for the test.
- Use a Bayesian classifier to classify both the train and test datasets and calculate both accuracies (for each dataset separately).
- Plot the decision boundary and classification results while representing the misclassified samples with a different color or shape (for each dataset separately).
- Plot estimated PDFs. (3D for each dataset separately)
- Contour estimated PDFs along with the decision boundary. (2D for each dataset separately)
- What is the main difference between the two datasets? Explain your answer using your results and plots. Compare this part with the previous part.

Non-Parametric Density Estimation

- For this part of the homework, you have to generate a dataset for three classes, each with 500 samples from three Gaussian distributions described below:

$$\text{Class1:} \quad \mu = \begin{pmatrix} 2 \\ 5 \end{pmatrix} \quad \sum = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

$$\text{Class2:} \quad \mu = \begin{pmatrix} 8 \\ 1 \end{pmatrix} \quad \sum = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$$

$$\text{Class3:} \quad \mu = \begin{pmatrix} 5 \\ 3 \end{pmatrix} \quad \sum = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

  Use your generated data and estimate the density without pre-assuming a model for the distribution, which is done by a non-parametric estimation.
- Implement the below PDF estimation methods using h=0.09,0.3,0.6 and answer the following questions for all of them:
    1. Histogram
    2. Parzen Window
    3. Gaussian kernel (Standard Deviations of 0.2,0.6,0.9)
    4. KNN Estimator (Fork=1,9,99)
- Estimate P(X) and plot the true and estimated PDF, and compare them for each model.
- Find the best value for h in the Gaussian kernel model with a standard deviation of 0.6 using 5-Fold cross-validation.

✓ Pay extra attention to the due date. It will not extend.

✓ Be advised that submissions after the deadline would not grade.

✓ Prepare your full report in PDF format and include the figures and results.

✓ Do not use sklearn or any similar library and write your own code.

✓ Use only the python programming languages.

✓ Submit your assignment using a zipped file with the name of:

"FirstName_LastName_ TeammateFirstName _Teammate LastName.zip"

✓ Email your zip file to 'arezoo7697@gmail.com'.

✓ Using other students' codes or the codes available on the internet will lead to zero grades.