Homework 01: Multivariate Parametric Classification

Doruk Özer

70192

At first part, I am asked to create data points from bivariate Gaussian densities with specific parameters. I put given means, covariances and sample sizes for each class to arrays in the same order(parallelly).Later I created the Normally distributed points and stacked them, then I created and stacked the y values parallel to the X matrix which is 300x2 matrix. Later I created the csv file of my datapoints with X and Y values.

Text

Description automatically generated with medium confidenceAfter creating the csv files I started my approximations of mean values, covariances, and class prior possibilities from the data points that I created. In order to, calculate the covariance matrix for each class from the given data points ,I used the formula at the Figure1. In order to approximate the sample means I took the means of the vectors of data points for each classes. I also calculated the frequencies of each classes which is same as prior probabilities. I stored covariance, means, and prior probabilities in arrays. Those values will be used in each class score function which is g(x).

Figure1(Taken from lecture notes at blackboard)

Polygon

Description automatically generated with low confidenceAfter the calculation of mean values, covariances, and class prior possibilities, I wrote the score function method which is called as score. That method of score function uses the formula at Figure2, in a way that picks the maximum score value, and identifies the which class’s score function gives the maximum value and returns the class label of that class. What I mean is, it return class label 2 if the score function of class 2’s score function is the maximum among other class. Later I defined a function called confusionMatrix2 that returns an array with the predicted class labels(like y\_truth which is actual array that contains the real values not predictions). By using the predicted label array and y\_truth I created the Confusion Matrix . I observed the 5 misclassified points.

Figure2(Taken from lecture notes at blackboard)

Finally, I was able to find and marked the misclassified points by the help of Visualization part of lab3. I used the plot function that was used in lab3 at visualization part.

Calendar

Description automatically generated with medium confidenceChart, scatter chart

Description automatically generated

Figure3 (Outputs of my Program)