2)

Generate random data points from three bivariate Gaussian densities. Firstly, define class\_means and class\_covariances and class\_sizes with given parameters.

Text

Description automatically generated

Then, created random 3 classes based on a multivariate normal distribution. random samples created and its corresponding labels created with given class\_sizes.

Text

Description automatically generated with low confidence

I saved these data point to the csv file to use later. This is not necessary in this Homework, I wanted to save them for later.



Our data points should be checked to see how it looks like like. I used matplotlib.pyplot library top rint data points.

Chart, scatter chart

Description automatically generated

3)

In the next part we will estimate parameters and prior probabilities.

First, I read data points that I saved to the csv file. I got number of classes, by getting max(y\_truth) value which is equal 3. Then I got number of samples from data set (data set shape is (300,3)) which is equal 300.

Graphical user interface, text, application

Description automatically generated

I calculated sample means from given data with corresponding formula below.

Graphical user interface, text, application, email

Description automatically generated

Next, I calculated sample covariances from data. By using following formula. However, in following formula xi vectors is considered as column vectors. However, in the python code this is different. when writing code, each row in matrix X corresponds to a vector xi.

Graphical user interface, text, application, email

Description automatically generated

I calculated prior probabilities:

Graphical user interface, text, application

Description automatically generated

4)

In this step we will need to calculate confusion matrix for the data points in your training set using the parametric classification rule you will develop using the estimated parameters from the previous step.

Before calculating score function, we will need to calculate Wc, wc, wc0 matrices.

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

I used python function to define score function formula.

Graphical user interface, text

Description automatically generated

I used score\_def function to calculate g scores with given X matrix data. Then predict y labels by using g scores.

Text

Description automatically generated

Printed confusion matrix, I used pandas crosstab function.

Graphical user interface, text, application

Description automatically generated

5) Visualization

First plot y\_truth values each point. Then, mark misclassified data points.

Scatter chart

Description automatically generated with medium confidence

A picture containing diagram

Description automatically generated

calculate discriminant values for 3 classes using the score function which is defined above. Then draw decision boundaries of each classes. When I don’t assign nan values to out pf area, it gets unwanted lines. By using discriminant values, plotted contour graph lines

Text

Description automatically generated

Chart, scatter chart

Description automatically generated