

ENGR 421 / DASC 521: Introduction to Machine Learning

Homework 01: Multivariate Parametric Classification

Deadline: November 9, 2020, 11:59 PM

In this homework, you will implement a multivariate parametric classification algorithm in Matlab, Python, or R. Here are the steps you need to follow:

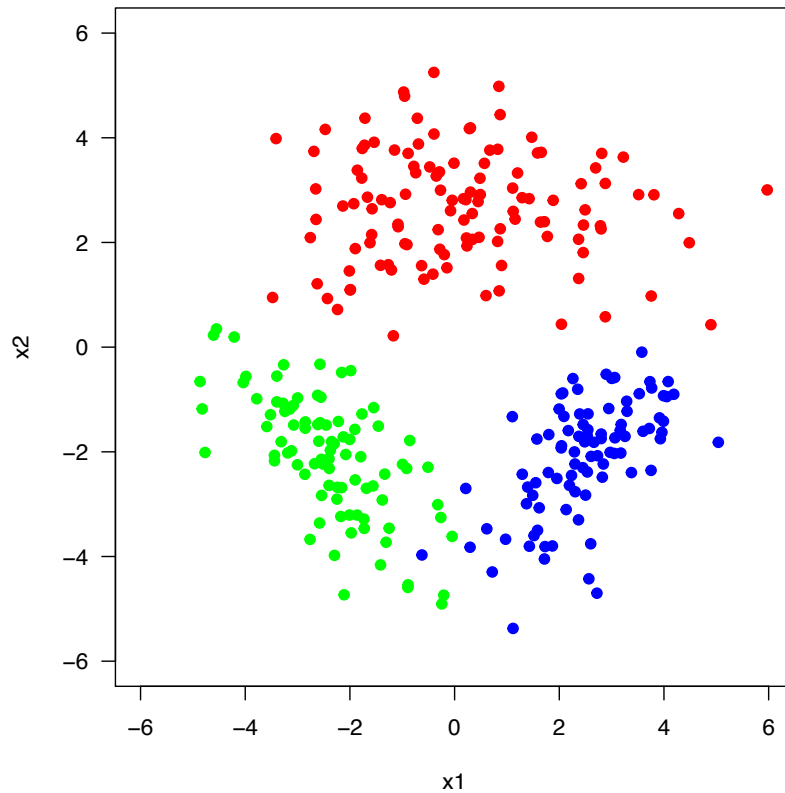
1. Read Chapter 5 from the textbook.
2. Generate random data points from three bivariate Gaussian densities with the following parameters:

$$\mu_1 = \begin{bmatrix} +0.0 \\ +2.5 \end{bmatrix}, \quad \Sigma_1 = \begin{bmatrix} +3.2 & +0.0 \\ +0.0 & +1.2 \end{bmatrix}, \quad N_1 = 120$$

$$\mu_2 = \begin{bmatrix} -2.5 \\ -2.0 \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} +1.2 & -0.8 \\ -0.8 & +1.2 \end{bmatrix}, \quad N_2 = 90$$

$$\mu_3 = \begin{bmatrix} +2.5 \\ -2.0 \end{bmatrix}, \quad \Sigma_3 = \begin{bmatrix} +1.2 & +0.8 \\ +0.8 & +1.2 \end{bmatrix}, \quad N_3 = 90$$

Your data points should be similar to the following figure.



3. Estimate the parameters $\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3, \hat{\Sigma}_1, \hat{\Sigma}_2, \hat{\Sigma}_3, \hat{P}(y = 1), \hat{P}(y = 2),$ and $\hat{P}(y = 3)$ using the data points you generated in the previous step. Your parameter estimations should be similar to the following figures.

```

print(sample_means)
##           [,1]      [,2]      [,3]
## [1,] 0.1555793 -2.394584  2.483595
## [2,] 2.6628991 -2.098373 -2.116054

print(sample_covariances)
## , , 1
##
##           [,1]      [,2]
## [1,] 3.54503797 -0.09391921
## [2,] -0.09391921  1.13613311
##
## , , 2
##
##           [,1]      [,2]
## [1,] 1.1513251 -0.8314826
## [2,] -0.8314826  1.3774658
##
## , , 3
##
##           [,1]      [,2]
## [1,] 0.9969161 0.6507806
## [2,] 0.6507806 1.1967356

print(class_priors)
## [1] 0.4 0.3 0.3

```

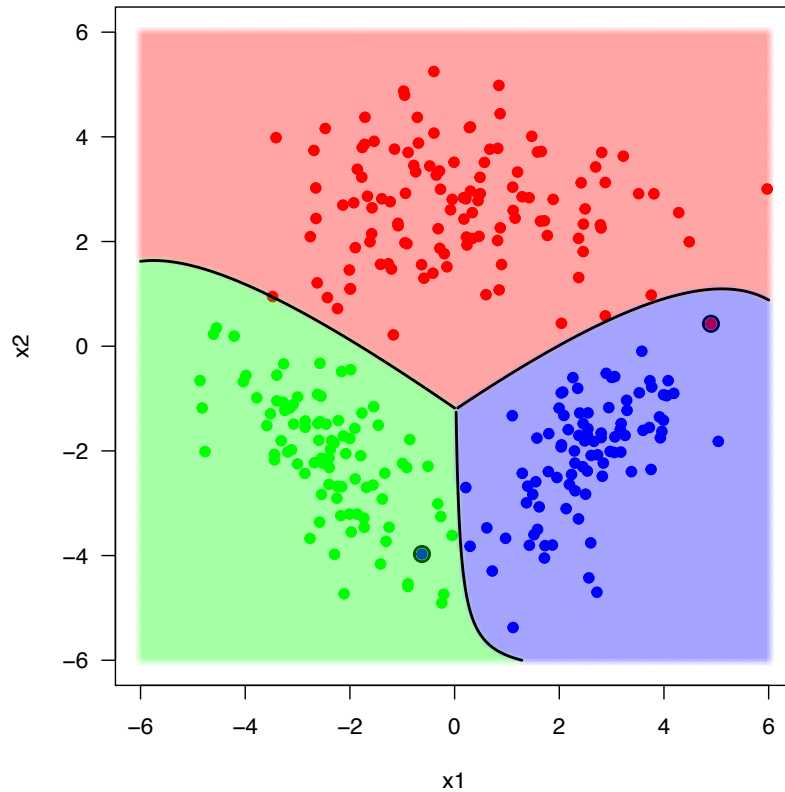
4. Calculate the confusion matrix for the data points in your training set using the parametric classification rule you will develop using the estimated parameters from the previous step. Your confusion matrix should be similar to the following matrix.

```

##           y_truth
## y_predicted  1  2  3
##           1 119  0  2
##           2   0 90  1
##           3   1  0 89

```

5. Draw your decision boundaries you will calculate using the parametric classification rule from the previous step together with data points and clearly mark misclassified data points. Your figure should be similar to the following figure.



What to submit: You need to submit your source code in a single file (.m file if you are using Matlab, .py file if you are using Python, or .R file if you are using R) and a short report explaining your approach (.doc, .docx, or .pdf file). You will put these two files in a single zip file named as ***STUDENTID.zip***, where ***STUDENTID*** should be replaced with your 7-digit student number.

How to submit: Submit the zip file you created to Blackboard. Please follow the exact style mentioned and do not send a zip file named as ***STUDENTID.zip***. Submissions that do not follow these guidelines will not be graded.

Late submission policy: Late submissions will not be graded.

Cheating policy: Very similar submissions will not be graded.