

ENGR 421 / DASC 521: Introduction to Machine Learning

Homework 01: Multivariate Parametric Classification

Deadline: March 14, 2022, 11:59 PM

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

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

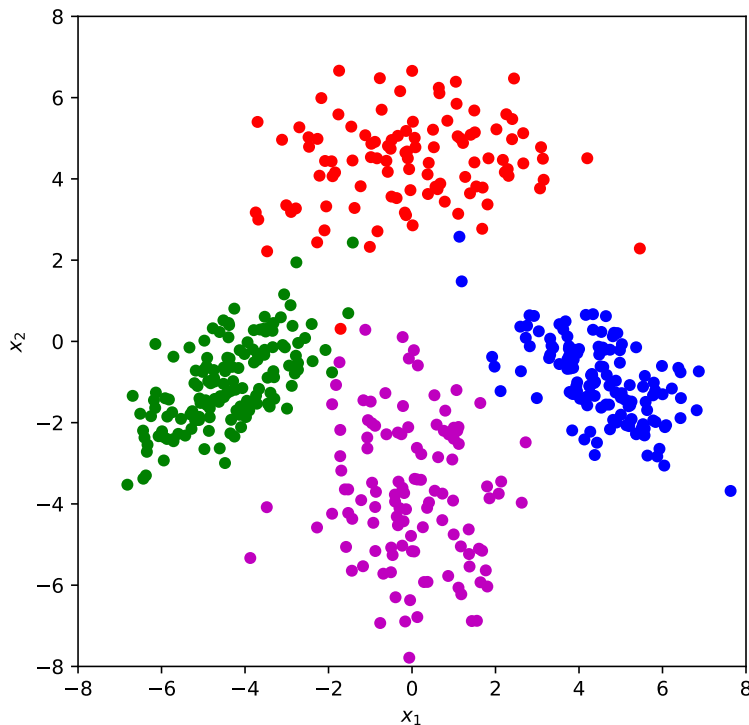
$$\boldsymbol{\mu}_1 = \begin{bmatrix} +0.0 \\ +4.5 \end{bmatrix}, \quad \boldsymbol{\Sigma}_1 = \begin{bmatrix} +3.2 & +0.0 \\ +0.0 & +1.2 \end{bmatrix}, \quad N_1 = 105$$

$$\boldsymbol{\mu}_2 = \begin{bmatrix} -4.5 \\ -1.0 \end{bmatrix}, \quad \boldsymbol{\Sigma}_2 = \begin{bmatrix} +1.2 & +0.8 \\ +0.8 & +1.2 \end{bmatrix}, \quad N_2 = 145$$

$$\boldsymbol{\mu}_3 = \begin{bmatrix} +4.5 \\ -1.0 \end{bmatrix}, \quad \boldsymbol{\Sigma}_3 = \begin{bmatrix} +1.2 & -0.8 \\ -0.8 & +1.2 \end{bmatrix}, \quad N_3 = 135$$

$$\boldsymbol{\mu}_4 = \begin{bmatrix} +0.0 \\ -4.0 \end{bmatrix}, \quad \boldsymbol{\Sigma}_4 = \begin{bmatrix} +1.2 & +0.0 \\ +0.0 & +3.2 \end{bmatrix}, \quad N_4 = 115$$

Your data points should be like the following figure. (10 points)



3. Estimate the parameters $\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3, \hat{\mu}_4, \hat{\Sigma}_1, \hat{\Sigma}_2, \hat{\Sigma}_3, \hat{\Sigma}_4, \hat{P}(y = 1), \hat{P}(y = 2), \hat{P}(y = 3)$, and $\hat{P}(y = 4)$ using the data points you generated in the previous step. Your parameter estimations should be like the following figures. (30 points)

```
print(sample_means)
[[-2.43085714e-04  4.41475305e+00]
 [-4.40159367e+00 -1.00817799e+00]
 [ 4.53185568e+00 -9.79534452e-01]
 [-3.20267739e-02 -3.79497784e+00]]

print(sample_covariances)
[[[ 3.46382957  0.26022464]
   [ 0.26022464  1.19547019]]

  [[ 1.34545849  0.78772458]
   [ 0.78772458  1.11187005]]

  [[ 1.27229804 -0.66903494]
   [-0.66903494  0.96283015]]

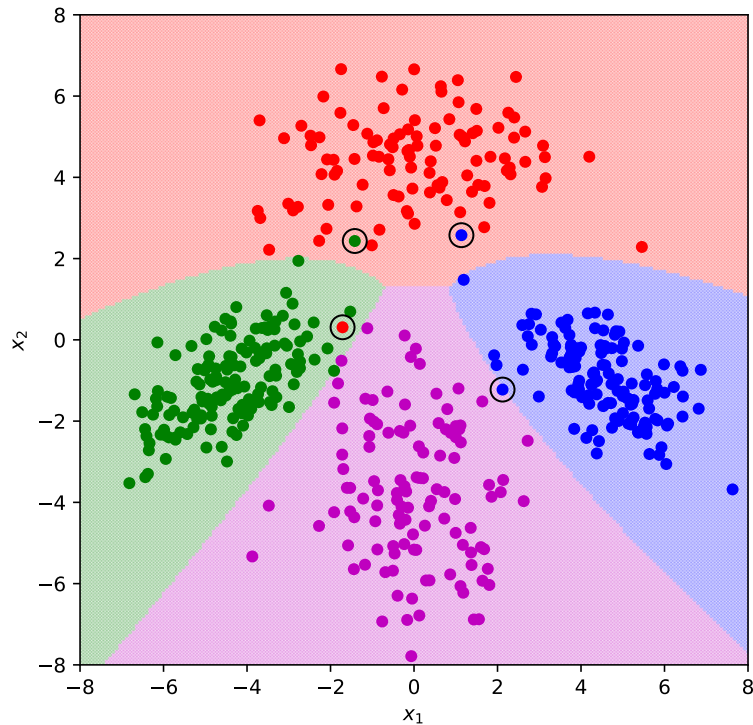
  [[ 1.44282286 -0.20544896]
   [-0.20544896  3.2734625 ]]]

print(class_priors)
[0.21 0.29 0.27 0.23]
```

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 like the following matrix. (30 points)

```
print(confusion_matrix)
y_truth  1    2    3    4
y_pred
1         104    1    1    0
2          1   144    0    0
3          0    0   133    0
4          0    0    1   115
```

5. Draw your decision boundaries that 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 like the following figure. (30 points)



What to submit: You need to submit your source code in a single file (.py file) named as ***STUDENTID.py***, where ***STUDENTID*** should be replaced with your 7-digit student number.

How to submit: Submit the file you created to Blackboard. Please follow the exact style mentioned and do not send a file named as ***STUDENTID.py***. 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.