DASC-521
Homework 01: Multivariate Parametric Classification

Ender Erkaya

October 2021

# 1 Introduction

In this homework, our duty is to expand parametric classification to multivariate data for multiple classes, $K > 2$. We assume that classes are gaussian distributed with each $p(x|y) -> N(\mu_c, \Sigma_c)$ with prior densities $p(y = c) = p_c$. To estimate score functions related to posterior class probabilities, $p(y = c|x)$, we need to estimate class conditional probability densities, $p(x|y)$, and prior probabilites, $p(y)$, for each class, using Bayes Rule given below. Noting that, we derive the decision rule using only the numerator of posterior pdf, because $p(x)$ is same for each class, not dependent of class variable $y$, for each data point $x$.

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}$$

Since, conditional probabilities are modeled with Gaussian distribution, only mean and covariance can be used to characterize the distribution. Hence, we need to estimate $\mu_1, \mu_2, \mu_3, \Sigma_1, \Sigma_2, \Sigma_3$ to characterize conditional probabilites, $p(x|y = c)$. Then, we estimate prior class probabilities from the data using $\frac{N_C}{N}$, where $N_C$ is total number of data in class $c$ and $N$ is the total number of data. After estimating parameters for score functions, $g_c(x)$ for each class, we calculate the score functions for a given data $x_i$ and we predict the class $y_i$ using the following decision rule:

$$g_c(x_i) = log(p(x_i|y = c) + log(p(y = c)$$

$$\hat{y_i} = \arg\max_c g_c(x_i)$$

# 2 Data Generation

To simulate multivariate multiclass parameter classification, we generate three class data with Gaussian distribution, $N(\mu_c, \Sigma_c)$ where $x \in R^D$ with $D = 2$, two dimensional space. Data is generated with given parameters and below Figure 1 is obtained.

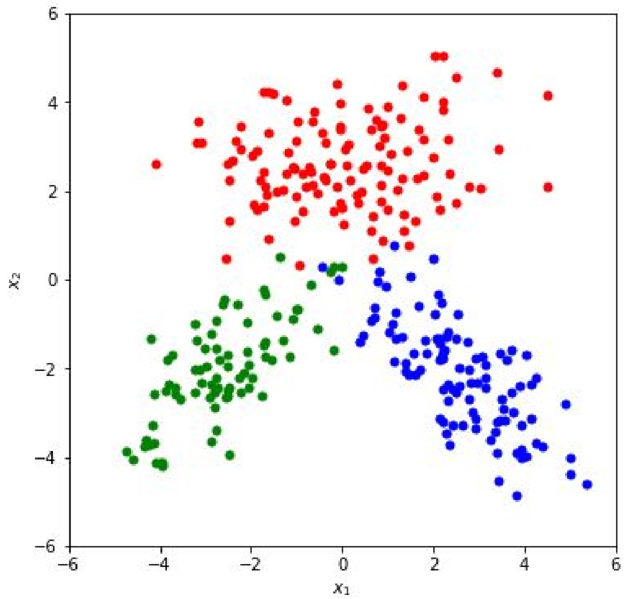


Figure 1: Generated Data

# 3 Parameter Estimation

To estimate conditional proability density for each class, we estimate corresponding Gaussian density parameters, $\{\mu_c, \Sigma_c\}$, using the sample mean and sample covariance statistics, for each class. Sample mean estimates are given in 2 and sample covariance estimates are given in Figure 3:

```
[[ 0.04453807  2.61225128]
 [-2.65871583 -2.04611631]
 [ 2.56054453 -2.12492713]]
```

Figure 2: Sample Mean Estimates

Additionally, to calculate score functions we estimate class prior probabilities, $p(y = c)$, for each class $c$ (Figure 4).

# 4 Prediction Using Parametric Classification Decision

Estimating class parameters, we obtain the score functions, defined as logarithm of the numerator of the posterior density, as quadratic function form given

```
[[[ 2.83985866  0.22625045]
  [ 0.22625045  1.01248436]]

 [[ 1.43826462  1.02345485]
  [ 1.02345485  1.37825958]]

 [[ 1.42107708 -1.08927412]
  [-1.08927412  1.52276646]]]
```

Figure 3: Sample Covariance Estimates

```
[0.4        0.26666667 0.33333333]
```

Figure 4: Class Prior Estimates

below.

$$g_c(x) = x^T W_c x + {w_c}^T x + w_{c0}$$

First, we calculate each quadratic function parameters, $W_c, w_c, w_{c0}$, for each given class $c$. Then, using the training data, we calculate each score functions for each data point. Consequently, we use the below decision rule as Maximum A Posteriori Estimator of $y$ to obtain the predictions for each observed data $x_i$, ie by picking the maximum one of $g_c$s:

$$\hat{y_i} = \arg\max_c g_c(x_i)$$

As a result, we obtain the following confusion matrix in Figure 5:

```
y_truth    1   2   3
y_pred
1        118   1   1
2          1  78   1
3          1   1  98
```

Figure 5: Confusion Matrx

# 5 Decision Boundaries

After calculating quadratic score function parameters for each class, $W_c, w_c, w_{c0}$, to decide decision boundaries, we evaluate each score function for uniformly picked data points on two dimensional real plane using meshgrid function. Then, the 'decision boundary 1', for red, is picked as $g_1(x) > g_2(x)$ and $g_1(x) > g_3(x)$, (dividing red points) and the 'decision boundary 2' is picked as $g_2(x) = g3(x) > g_1(x)$, (dividing blue and green but not red). The following figure obtained as result in Figure 6:

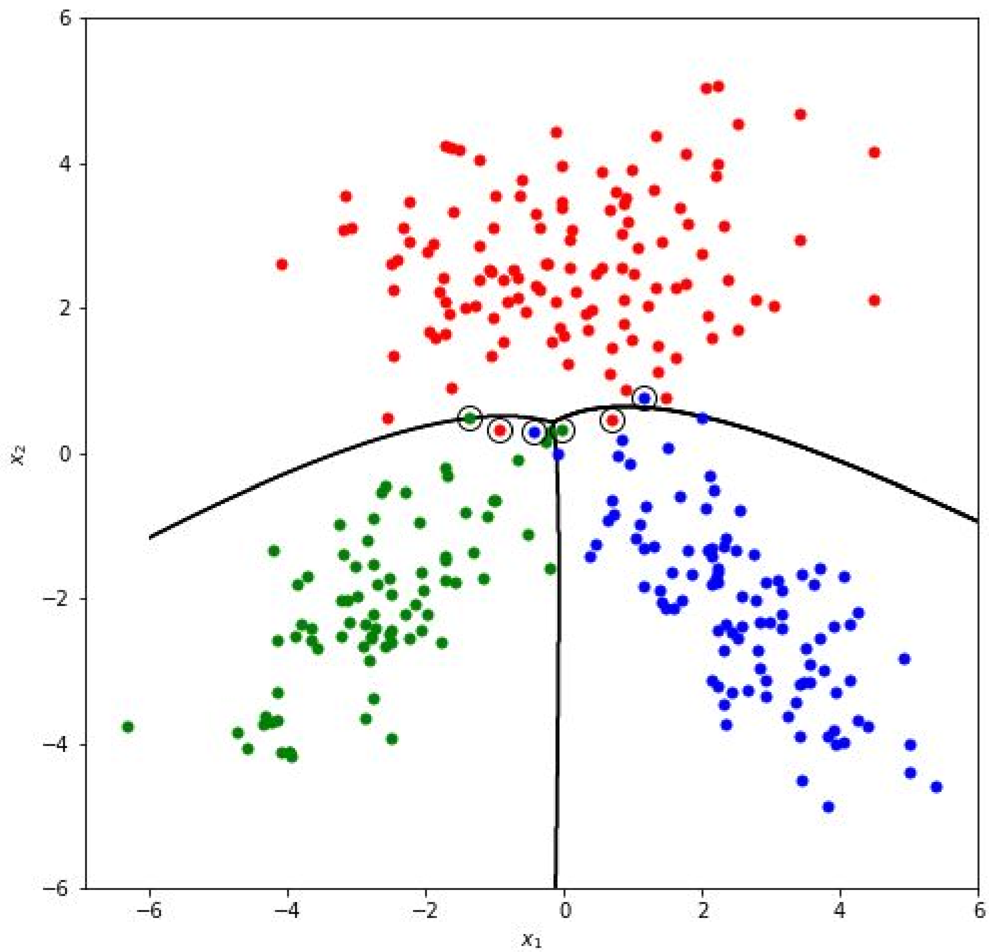


Figure 6: Decision$_Boundaries$

# 6 Conclusion

We calculated quadratic score functions using MLE estimator of class means and class covariance matrices and class prior densities. We made predictions for each training data point using maximum rule and calculated confusion matrix. We calculated score function parameters for each class and quadratic discriminator for each class, by picking maximum, ie we have two quadratic discriminator function $g_1 - g_2 > 0$ and $g_1 - g_3 > 0$. The decision boundaries are plotted using the above discriminator functions. The obtained results are very good, we have a few false predictions and missed points. The homework is very useful for me both to get familiar with python and to understand quadratic score functions and multiclass parametric classification and difficulties with selection decision rule.