KOÇ UNIVERSITY

DASC521: Introduction to Machine Learning

Fall 2021

Instructor: Mehmet Gönen

Homework 03- Discrimination by Regression

Gamze Keçibaş- 60211

09.11.2021

As in the first assignment, three different classes are generated from bivariate Gaussian densities (1) by provided size, mean array and covariance matrix.

$$p(x) = \frac{1}{(2\pi)^{d/2}|\sum|^{1/2}} exp\left[-\frac{1}{2}(x-\mu)^T\sum^{-1}(x-\mu)\right] \quad (1)$$

These datasets are presented below:

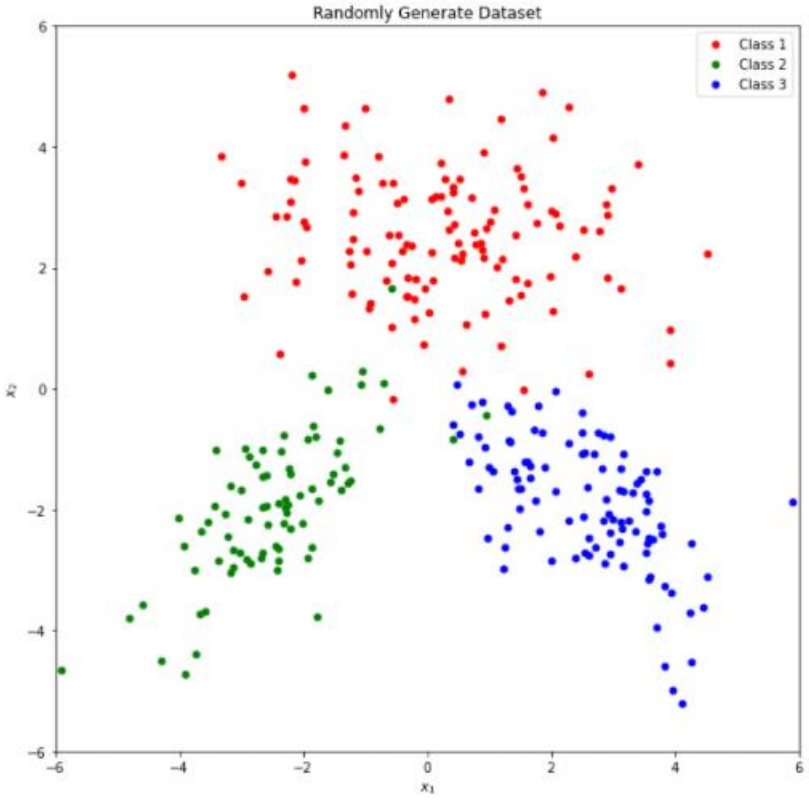

Figure 1: Randomly generated data points

After dataset is generated, learning parameters and functions are created. As provided, learning parameter that is called as eta in the script is 0.01. Acceptable loss is determined as 0.001 that is labeled as epsilon. Weights in the model are randomly initialized between -0.01 and 0.01 where shape of w is 2x3 and shape of w0 is 1x3. When learning parameters are defined, required functions are generated to make learning process. Sigmoid function (2) is used in the model.

$$\sigma(a) = \frac{1}{1+e^{-a}} \quad (2)$$

After the prediction with sigmoid function, weights are updated by Gradiend Descent method and its result is implemented as functions of change in weights. Equation 3 represents change in w and Equation 4 represents change w0. It is the learning part of the model. These equations are below where η is learning rate, N is size of sample set:

$$\Delta\omega = \eta \sum_{i=1}^{N}(y_i - \hat{y}_i)\, x_i \quad (3)$$

$$\Delta\omega_0 = \eta \sum_{i=1}^{N}(y_i - \hat{y}_i) \quad (4)$$

In the iteration part, weights are updated and learning process is checked by minimizing sum squared error.

$$Error = \frac{1}{2}\sum_{i=1}^{N}\sum_{c=1}^{K}(y_{ic} - \widehat{y_{ic}})^2 \quad (5)$$

Finally, optimum weights are below that are first 2x3 matrix is w and final row is w0:

```
[[ 0.42854028 -2.1834883   2.07150796]
 [ 3.98895367 -1.68764395 -2.19526022]]
[[-1.41906338 -2.63345208 -3.45798382]]
```

Figure 2: Weight matrix

When the optimum weights are obtained, all iteration results are provided in Figure 3:

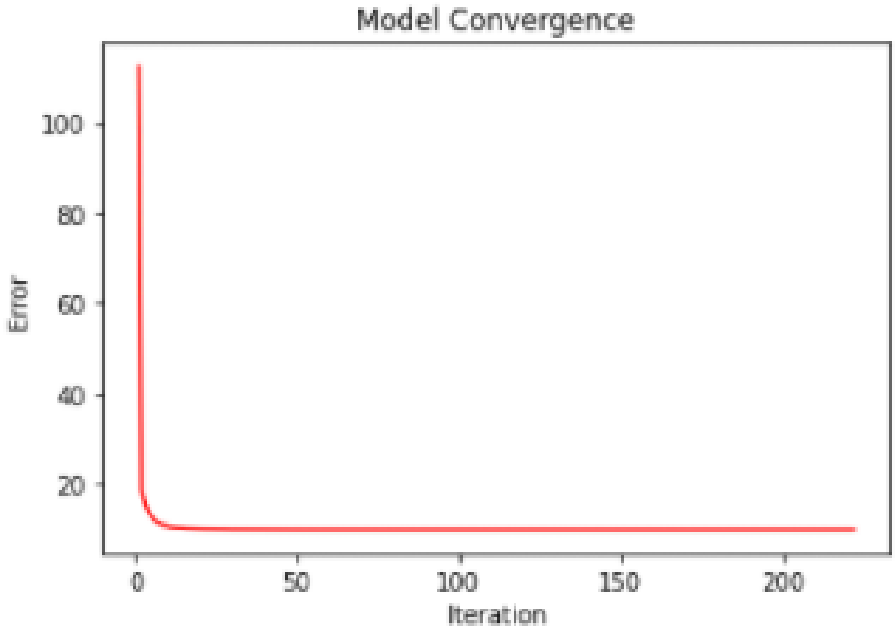

Figure 3: Convergence Plot

The confusion matrix of optimum model:

y_truth	1	2	3
y_pred			
1	115	2	1
2	2	76	0
3	3	2	99

Figure 4: Confusion matrix

The model works as in Figure 5 that represents decision boundaries:

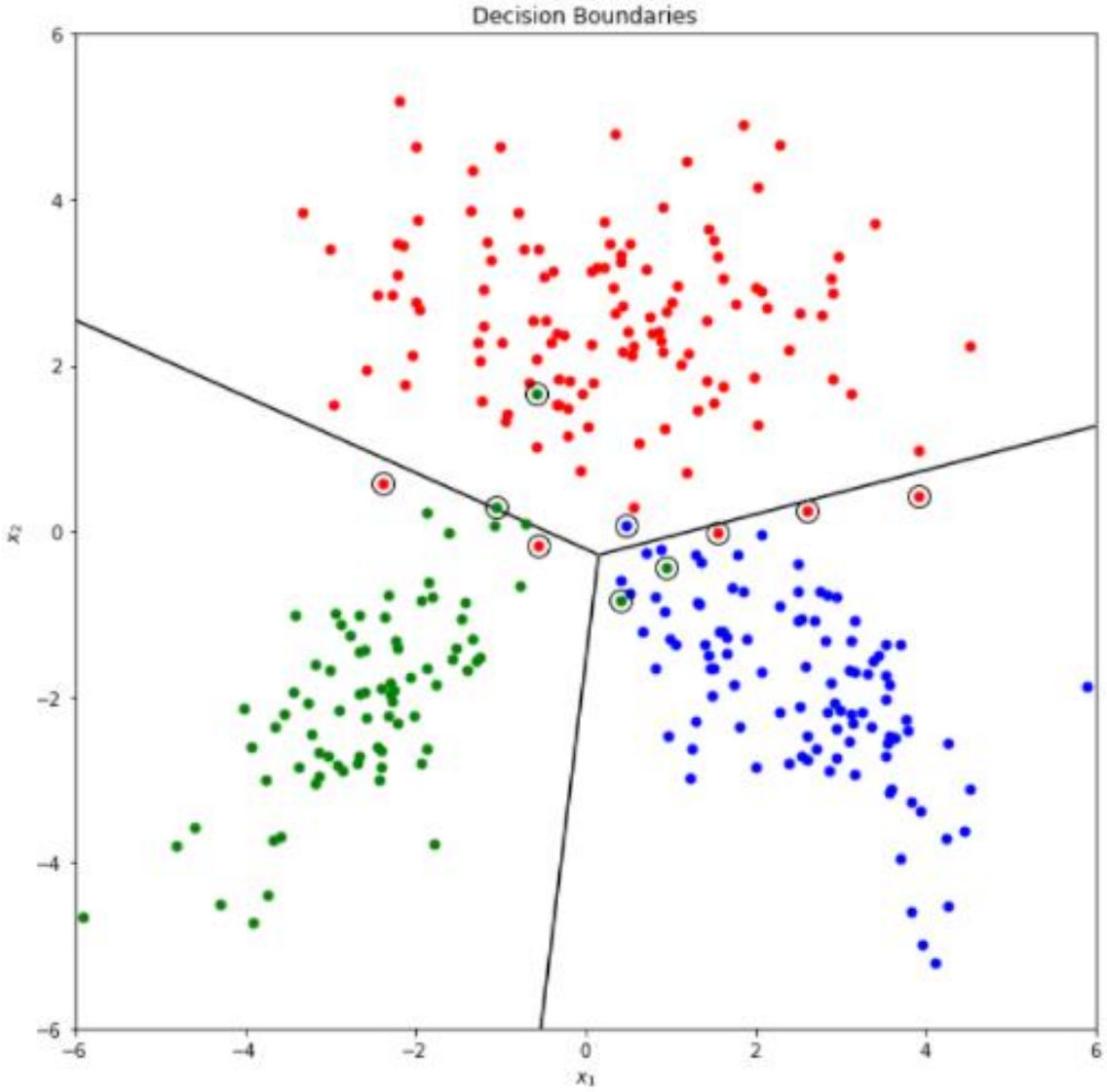

Figure 5: Decision Boundaries