DASC 521 - Introduction to Machine Learning

Emre Küçük – 0071236

First I imported necessary libraries, then read data from csv files. After that, I changed label data set's column name as another integer and concatenated images and labels data sets.

After that I simply divided total data set into two as train and test data sets. Train data set contains first 200 data points and test contains last 200.

Then I derived mean, deviation parameters and prior probabilities for each class using training data set.

After deriving parameters and necessary probabilities, I used parametric classification algorithm in order to create a classification model. For a feasible approach, I used logarithms of Gaussian probability density function for a feasible calculation and created a score function for each class. Then I calculated scores for each data point in each class.

For each data point, there is a higher score. We labeled data points with scores, then measured accuracy by using a confusion matrix.

On first look accuracy rate might seem to be high but one of labels have very little amount of data points so using is data set is not very healthy.