Buğra Sipahioğlu

October 12, 2019

Mehmet Gönen, Comp 421

**Homework 01 Report: Naive Bayes’ Classifier**

The homework has been splitted into six parts: importing libraries, importing data, splitting data, parameter estimation, parameter classification and confusion matrix. I will explain first five of them since confusion matrix is simply a sequence of predicates.

**Importing Libraries & Data**

Pandas library has been used for data extraction. Since pandas library returns data frame and these data frames are hard to manipulate, they had to be converted into arrays. In my case, these arrays were NumPy arrays for extensive NumPy library usage. For the data import, I’ve used sklearn’s train\_test\_split function to extract all the train and test data with values and labels can be done by only one line. Since the binary labels were given as ones and twos, I’ve converted them to ones and zeros such that female class corresponds to zeroth class.

**Splitting Data**

In order to calculate means and standard deviations of features with respect to their classes, I’ve created a dictionary called “gender\_dict”. First tuple corresponds to all the female training features, and second one holds male training features.

**Parameter Estimation**

Means, standard deviations and prior probabilities were calculated with respect to class labels. These operations were held column-wise, i.e. means\_0 matrix has one row and 4096 columns.

**Parametric Classification**

First of all, I’ve calculated the log-likelihoods of classes with the assumption that data would be distributed as gaussian variables. Then I’ve calculated joint probabilities as vectorized calculations.