I am using a class-based approach to implement the Naïve Bayes classifier. The train function reads the contents of the train file line wise, since each patient’s record is stored in a single line and store them in the variable “patients\_record”. The read data is used to compute the class priors for both the classes. The rows containing data of the normal and abnormal patients are separated to make computations easier. Then we remove the first column from the normal and abnormal rows, leaving us with only the medical tests’ data. Since we have already divided the data into classes, we do not need the first column anymore.

Then we calculate the conditional probabilities for each medical test, from test 1 till test 22, for four different cases: a) The test passed given the patient was normal, b) The test failed given the patient was normal, c) The test passed given the patient was abnormal, d) The test failed given the patient was abnormal. We store these probabilities in row form.

For the test part, we read the patients’ records in the same way, and store the test data in a single np array. Ye extract the first column from the data as our “y\_true”; the truth values. We get the likelihood of normal and abnormal. This is done using the likelihood function.

The likelihood function takes in a row, i.e a patient’s record, and appends the probability of the test predicting normal and abnormal in 2 separate arrays. At this point we have an array carrying the probabilities of the tests passing for each test for the given patient, and an array carrying the probabilities of the tests failing for each test for the given patient. We multiply all the probabilities of each test passing, which gives us the likelihood of the patient being normal, and the probabilities of each test failing, which give us the likelihood of the patient being abnormal.

The two likelihoods are returned to the test function for each patient, and we multiply these likelihoods with their corresponding class priors. If likelihood of patient being multiplied with the prior of the normal class is greater than the same product for the abnormal class, the prediction is “Normal” and vice versa. We then send our predictions and the “y\_true” extracted earlier to our accuracy function, which returns the accuracy of our test.