

# Task 2 - Bernoulli naive Bayes Classification

## The Algorithm

The algorithm first uses binarization on the training and test data using the threshold = 1. It then calculates the rows of the training, test and training class matrix and the column of the training data.

The first for loop calculates the likelihood probabilities for each class when  $b = 0$  and  $b = 1$ . The second for loop is a nested for loop which goes through each row of the test data and calculates the probability using naive bayes. It then finds the maximum and the class is predicted.

## Standard output

When running `my_bnn_system.m` the following table illustrates what is displayed to standard output:

Time taken	2.682200 seconds
N	7800
Nerrs	2915
acc	0.6263

## Effect of Threshold on Classification accuracy

To do this we first try out a few threshold values when the threshold is 1, 2, 3, 5, 10 and 15. The results to standard output for each value is as shown below:

Threshold	Time taken	N	Nerrs	Acc
1	2.682200 seconds	7800	2915	0.6263
2	2.645789 seconds	7800	2899	0.6283

3	2.702660 seconds	7800	2877	0.6312
5	2.668105 seconds	7800	2855	0.6340
10	2.621291 seconds	7800	2825	0.6378
15	2.826131 seconds	7800	2825	0.6385
20	2.493489 seconds	7800	2834	0.6367
25	2.680490 seconds	7800	2815	0.6391

From this we can see as we increase the threshold value, the accuracy of predicting classes increases although it is only by a small increase. We can also see this from the value of Nerrs decreasing which means that the prediction of classes that are correct is increasing as the threshold increases.