We implemented the naive Bayes model according to 9.27 and 9.28 Equations of the textbook. This is a faster implementation compared to using 9.25 Equation. The reason is that in Equation 9.25, all of the computations in the equation has to be done each time we would like to classify a data point. We have to do this on all of the data points in the training set, validation set, and test set. This is costly. However, using equation 9.28, we can compute “W” and “b” only once and then use them in the classification time. In other words, when we would like to classify the data points we only have to compute $X . W + b$ which is faster than computing everything every time.

We experimented with different smoothing hyperparameters and the results are demonstrated below. We observed that the closer to zero was the smoothing value, the higher was the accuracy. The best accuracy was obtained with smoothing=0.001. On the other hand, the higher values of smoothing contributed to underfitting. In general, naive Bayes model performed as an effective classifier with very good accuracy results.

Smoothing Hyperparameter	Training Acc.	Validation Acc.	Test Acc.
0.001	99.96	98.13	98.92
0.01	99.96	97.92	99.07
0.1	99.68	97.85	98.92
1	97.88	95.76	96.27
5	86.65	87.65	85.44
10	86.65	87.65	85.44