IST 707 Applied Machine Learning
HW6: Naïve Bayes and Decision Tree for Digit Recognition
Name: Jainish Savaliya

➔ In this particular classification problem out got is to recognize the digit from 0 to 9 for the dataset and after a glance in the dataset we can see the labels and pixel values so basically the image of the handwritten digit is converted into a numerical form by the lightness and the darkness of the writing so according to the darkness we will predict the shape of the image and predict the number.

➔ Here we have already given two separate datasets so for preprocessing we will only factorize the label variable instead of the whole dataset because it contains too many pixels and numerical values.

➔ Also, we will slice the train set to first 1000 rows only so that the model will take less time to run.

➔ For the Decision Tree model we ran 5 fold cross validation the dataset so that the model can use whole dataset and then I trained the model and it gave the accuracy of 27.34 percent which is less.

➔ So, I tuned the model using search grid which will create the data frame called search grid that includes a sequence of values for the 'cp' parameter ranging from 0.01 to 0.50 in increments of 0.01. This is likely used for the tunning of the model

➔ After tuning the model, I got the accuracy of 60.75% which is more than double of the model which was untuned.

➔ For the Naïve Bayes model we ran 5 repeated cross validation on the data set so that the model can use whole data and trained model with 5 fold CV and it gave the accuracy of 46.23%

➔ So to tune the model I used grid expansion function with the parameters user kernel, Laplace and adjust which give the data different dimensions, adding small values to reduce 0 probabilities and adjust it to control the overfitting respectively.

➔ In this case I was getting the same accuracy for the dataset before and after tunning with is 46.25% which I think because of the slicing of the data. I ran this model in python and I gave the higher accuracy which was also higher than Decision Tree model.

➔ After comparing both the algorithms naïve bayes reached higher accuracy and If I compare the running time than I can say decision tree is faster because it only requires one pass through the data during training, while Naive Bayes requires multiple passes to compute conditional probabilities. Additionally, decision trees can quickly make predictions for new instances once the tree is built, while Naive Bayes requires computing probabilities for each feature in the instance.