A picture containing knife

Description automatically generated

Module 4 Assignment

Neural Network

ALY 6020 – Predictive Analysis

Instructor – Justin Grosz

Name -Ashish Kothari NUID- 001360980

**Introduction**

In this assignment we are using Digit recognizer where it consists of images drawn by hand using grey scale and we are trying to see how well our model can recognize digits correctly. This data set consists of 785 variables and 42000 observations. The dataset has label which tells us what digit will be printed. There are many pixel values with 0 value and many carry meaningful data. We are going to see how accurately the hand drawn numbers from 0 to 9 are oriented using Neural Network and KNN model. We are going to build both models and compare their accuracies to check which model is more fit.

**Data Analysis**

1. **Data Cleaning:**

In this we are loading the dataset and we are checking first for outliers. When we run the dataset and check for outlier, we observe that there are only two outliers in pixel779 column. Rest there are no outliers in any of the variables. So, we drop the outliers and remove those two rows out of the 42000 observations. We also check the datatype of the variables to check if there are any changes needed. There is no need to change any datatype value as all values are in integers

1. **Checking for missing values:**

We use the dataset to check if there are any null values in the dataset. After running the code for missing values, we observe that the dataset had no null or NA values in the dataset. Now we have cleaned the dataset and it is ready for further analysis. We will begin the analysis.

1. **Neural Network:**

We use neural network model as the dataset size is huge and dataset works better with neural network model to give out faster predictions depending on nu. After cleaning the data, we start building our first model Neural network. In this we begin with normalizing the dataset for quick learning and leading to fast convergence as there data imbalance and high variance in the dataset. For normalizing we divide the test and train dataset by 255 as the pixel value is between 0 -255. After normalizing we use MLPclassifier to build a model and it by default uses relu. We build relu as there are no negative values in the dataset and relu has value range from 0 to maximum value. So, relu is the best method we can use for this dataset. We build the neural network model using hidden layer size and random state we can change and check the change in accuracy. We check for hidden layer size 1 and random state 1 we get the accuracy of 25%. So, we increase both values and check again. Now the hidden layer size value is 10 and random state is 18 we observe that accuracy is 81%. So, we increase the hidden layer size to get better model. We choose 30 and the accuracy of the model is 93%. Therefore, we observe that by increasing the hidden layer size we get better and accurate model. This shows that 93% of the images drawn by hand where recognized correctly. Also, if we perform neural network model without normalization the accuracy is less as compared to normalizing it.

Table

Description automatically generated

1. **KNN model:**

After getting 93% accuracy for neural network model we build another model to compare with neural network model. We build KNN model because it is good for forecasting, text mining and many other reasons. So, we are trying to predict what hand drawn images will be recognized digitally. So here we can use KNN model. KNN model provides us with better predictions. We built KNN model for k=1 and checked that the accuracy was 95.4%. We changed the k value to 2 and the accuracy decreased. So, we tried one more value for k= 5 and accuracy of the model was again 95.2%. Therefore, we can keep the k value as 1.

Text

Description automatically generated

1. **Comparison between two models:**

We compare the two models based on their runtime. We observe that KNN model took around 2:45 minutes to run whereas Neural network model took hardly 19 seconds to run. So, we use neural network model to achieve faster results. That is why neural network is used more widely to make predictions where we require more faster results like auto driving.

**Conclusion:**

In this module we observe that pixels with 0 value are more important compared to pixels with values >0 because pixel value with 0 tells us that there will be no digit representation which is more important and pixel values with >0 tells us what shade will be there. The higher number represents darker shade. So, by knowing what spaces are empty (pixel =0) it will help us to predict the number more easily and accurately.