QUESTION

1

Problem: 1

We have used **Neural Network** model for classification.
A neural network is a simple combination of multiple linear classifiers inspired from human brain. A neuron is simply a linear transformation of its inputs. Stacking the output of multiple neurons in a vector forms a neural network layer. A non-linearity is applied to the output to make the output non-linear. Then multiple layers are put in a pipeline to get the neural network.

The **architecture** of our neural network is as follows-
Input (BoW representation) $\rightarrow$ A layer with 128 neurons $\rightarrow$ Relu non-linearity $\rightarrow$ Another layer with 128 neuron with regularization and relu non-linearity $\rightarrow$ output layers with 50 neurons and softmax non-linearity
The output of the neural network gives probabilities of the input belonging to the respective class, taking argmax() of the output gives the class label.

Hyperparameters:
We had to tune the whole network architecture, we first experimented with more layers, no dropout the result was overfitting, after some iterations we finalized the architecture above.
Other hyperparameters that we tuned are regularization constant in 2nd layer and droput constant in dropout layer.
The regularization constant was searched within a range of 10^{-4} to 10^{-6} as usually they are small. Whereas the dropout constant was searched for in 0.2 to 0.5.
The hyperparameters are chosen to maximize validation accuracy and minimize the difference between training loss and validation loss.

Problem 2: Advantages/Disadvantages of NN, LwP, OvA, DT

Neural Network

1. Advantages

    (a) They have a structure that is adaptive in nature.
    (b) These networks have numerical strength, which makes them capable of performing more than one function at a time.

2. Disadvantages

    (a) Neural Networks are computationally more expensive than traditional algorithms.
    (b) They are highly dependent on the data made available to them. This means that the efficiency of any neural network is directly proportional to the amount of data it receives to process.

LwP

1. Advantages

    (a) They are very simple, interpretable and lightweight model.
    (b) They can work accurately with any number of classes, and can predict accurately for classes with less prototypes

2. Disadvantages

(a) The performance in terms of predictions from the model is in general slow.

(b) The model ignores the possibility that some features of the prototype can be irrelevant for the required outcome.

OvA

1. Advantages

    (a) OvA methods usually offer very high precision values.

    (b) We can use OvA model to improve performance of large datasets.

2. Disadvantages

    (a) They are usually slow at prediction.

    (b) A possible downside of this approach is that it requires one model to be created for each class.

    (c) It is challenging to deal with large datasets having many numbers of class instances.

Decision Trees

1. Advantages

    (a) Requires less effort for data preparation during pre-processing.

    (b) Missing values in the data also do NOT affect the process of building a decision tree to any considerable extent.

2. Disadvantages

    (a) We cannot use it if the size of the data is too big as the nodes will grow high in number, which will lead to over-fitting.

    (b) This method involves a higher time to train the model.