CS771: Introduction to Machine Learning
Assignment 2

Group Name : Perceptron

Problem 1 :

- To solve this problem, we have used One-Vs-All(OVA) method. It becomes prolonged when many classes are involved, though it is highly accurate. In our problem, we do not have many classes (as problems have millions of classes as well), which diminishes the disadvantage of this method.

- OVA method is a heuristic approach to solve multiclassification problems by dividing them into multiple sets of binary classification problems. The binary classifier is trained in one go, and predictions are made by the model which is the most confident about its prediction. The main advantage of this method is that it separates all the classes with good accuracy and provides a robust model.

- We also tried to do some pre-processing (standard scalar, min-max scalar, Etc.) in the features, but it turns out that accuracy dropped significantly, due to which we went ahead with the given features without any processing.

- We have used logistic regression classifier in our model over many values of hyperparameter C and at last we have trained our model at C = 10.

- Table shown below contains values of prec and mprec across a few values of C. To find these values, we have trained our model in 100% data and then tested on the same data, so these are relative values using which we came up with the best hyperparameter value.

- We also tried to oversample our data using a library like imblearn, which did increase mprec values. However, prec values show a significant dip when compared to previous values.

- We have added probabilities of classes 33, 36, and 38 to '0' in the predict.py file to get the correct classes when accessed by their index.

Preprint. Under review.

Hyperparameter	Prec@1	Prec@3	Prec@5	mprec@1	mprec@3	mprec@5
C = 1	81.6	94.4	97.4	56.28	82.88	91.36
C = 2	82.4	95.1	97.6	60.33	86.09	94.29
C = 4	83.0	95.6	97.9	63.5	88.83	95.14
C = 6	83.4	95.7	97.9	64.66	89.44	95.58
C = 8	83.4	95.8	97.9	65.88	90.26	95.58
C = 10	83.7	95.9	97.9	66.96	90.43	96.01

Problem 2 :

Advantages :

- One-Vs-All(OVA) is very precise compare to other methods. This fact is also shown in the table at the end of this answer. We have compared ova with decision tree and KNN.
- Size of model trained by this method is comparatively smaller than the other methods, specially models trained by neural network are in mega bytes.
- It is far less complex than other methods like decision tree, neural network
- Methods like decision trees, artificial neural network overfits on data while logistic regression does not overfit on data as it uses regularizer.

Disadvantages :

- If the data is not linearly separable then this method may not work while methods like neural network, decision tree etc will work on linearly not separable data as well.
- When there is very complex relationship between data we need more powerful methods like neural network.
- It does not train very well with rare classes present in the dataset.
- With increasing number of classes, it slows down.

Methods	Prec@1	Prec@3	Prec@5	mprec@1	mprec@3	mprec@5
OVA	83.7	95.9	97.9	66.96	90.43	96.01
KNN	57.8	80.8	89.7	21.09	42.94	59.16
Decision Tree	73.9	79.1	79.7	49.64	54.77	57.76