DMW C2 Assignment 1 : k -Times Markov Sampling for SVMC

Group Members

IIT2018111 Jishan Singh

IIT2018141 Rajat Sharma

IIT2018148 Fahad Ali

IIT2018151 Bhavya Jain

IIT2018157 Yatharth Parate

Abstract :-

In doing this assignment we have used algorithm 2 of the given research paper i.e. SVMC Algorithm Based on k Times Markov Sampling for Unbalanced Training Samples, because the datasets which were given to us were of large size, multi class and unbalanced in nature.

Methodology :-

First of all N randomly selected data points were taken from the training dataset and were trained by SVM classifier to obtain a basic learning model. After that one variable was created for counting the number of samples. Then two entries were chosen randomly from the training set and a ratio was calculated for taking further decisions. Then based on that ratio the decision of including or excluding one entry was being made.

If the entry was selected then we checked whether the limit of the maximum samples was reached or not and if it wasn’t then we added that entry in our samples and simply went back to restart from the place where two entries were chosen for choosing one more. If the entry was not selected then we simply kept on selecting random entries one by one and calculated the ratio until some number of them was rejected. If the limit of the maximum samples was reached then we trained those samples by SVM classifier and obtained another learning model which replaced the previous one.

In this manner we repeated the above mentioned process k times in order to do the sampling multiple times for training the *markov chain samples,* which were formed because of the decision parameter and selection/rejection of randomly selected data points from the training dataset.

⍺ = *e*-*l(f, z1)/e-l(f, z2)*

where “⍺” is the decision parameter and *“l(f, z)”*  is the hinge loss function corresponding to the entry point “*z”* and learning model *“f”*.

The code is given in the .ipynb file and the results are provided in the excel file. There are four .ipynb files i.e. one for each of the four kernels that are linear, RBF, polynomial and sigmoid.

Conclusion :-

The accuracy for linear kernel was found to be 28%, for RBF it was 22%, for polynomial (X2) it was 22% and for sigmoid it was 4%.

Reference :-

k-Times Markov Sampling for SVMC, Bin Zou, Chen Xu, Yang Lu, Yuan Yan Tang, Fellow, IEEE, Jie Xu, and Xinge You, Senior Member, IEEE