CMP 466 – Machine Learning and Data Mining- Spring 2021

Team Project- Assignment 2

Please submit a report including the following by the deadline:

1- [10 points] Read your data into your program. Save it in 2 parallel matrices, one for features (X), and one for labels (y).

2- [10 points] Perform k-fold cross validation on the data with k= 5 and/or 10. The classification results should be reported for each fold (points 3 and 4 below).

3- [20 points] Apply linear support vector machine classifier (LinearSVC) on your dataset and report the training and testing accuracy using the model described in points 2 above.

4- [40 points] Apply support vector machine classifier (SVC) with linear and non-linear kernels and report the training and testing accuracy using the model described in points 2 above. Try all the non-linear kernels provided in sklearn with different values of hyper parameters (C and gamma for RBF kernel, degree for polynomial kernel, etc.).

5- [20 points] In a table, summarize your results from points 3 and 4 above for each of the classifiers and the different values of their hyper parameters. Discuss your results.

Submission: Please submit your assignment to iLearn by the deadline. Late submissions will be penalized according to the syllabus late policy. No submissions will be accepted after four days of the deadline.