Project 2

Due: 11:59pm, March 4.

1   Introduction

In this project, you will further practise your model evaluation skills. you need to build the Multi-layer Perceptron (MLP) model for a specific dataset. Then you should evaluate the model using different techniques.

Breast Cancer Wisconsin Data Set   In this project, you should use the Breast Cancer Wisconsin Data Set. You may download the dataset here: http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29. Detailed data description can also be found in this site.

Learning Curves.   Plot the Learning curves for your model. Based on your graph, detect overfitting and determine model parameters in a MLP. For example, width of hidden layer.

Support Vector Machine.   Build a Support Vector Machine (SVM) for the same dataset. Compare the results of MLP and SVM. Show the differences you find (for example, running time, accuracy).

5-fold Cross-validation.   Just like the first project, use 5-fold cross-validation to evaluate your model result. Randomly divide the training data D into 5 groups with the same size. During each iteration $i \in \{1, \ldots, 5\}$, select $D - D_i$ and D_i as training data and testing data independently. Finally, output the averaged testing accuracy as the output accuracy for a specific predicting model.

2   Submission

You only need to generate one **pdf** file which can be directly generated by colab notebook and then submit on Canvas before the due date. Fail to do so will make your final grade deducted. In the report, you should specify your model details when necessary. Try to write your code clearly so that someone else reading the code can understand it without significant effort (i.e. structure it and put enough documentation). The final report will be judged based on the clarity of the report.

3   Collaboration

Note that this is an **independent** project, which means you are not allowed to make a group. However, discussion is allowed. If you have discussed with someone or got any help from others, you need to clearly specify their names in acknowledgement.