PubH 7475/8475 Homework 4 (Spring 2021)

Due on March 17 (or 24 or 31), 2021

1. Apply a SVM and a feed-forward neural network (FNN) to ONE of the two following data sets. For the SVM, you can choose a fixed kernel (or select by sample splitting or CV). For the FNN, you can use only one hidden layers (or more layers if you like); the number of the hidden units can be fixed (or selected by sample splitting or CV). To save computing time, if you like, you can also do variable screening first. (15*2=30 pts)

   - NCI microarray data: there are $p = 6830$ predictors (i.e. genes). By ignoring a few classes with only few samples, we only consider 5 CNS, 9 renal, 7 breast, 9 NSCLC, 8 melanoma, 6 ovarian, 6 leukemia and 7 colon tumor samples. The predictors are in a file called Data, and the class labels in Info. *Use LOOCV to evaluate a classifier.*

     This dataset is one of the three used by Dudoit et al (JASA, 2002, p.77-87) to evaluate several classification methods.

   - Spam data: there are $p = 57$ variables (in the Data file) to distinguish two classes, spam (coded as 1) and email (coded as 0). There are total 1813 spams and 2788 emails. As done in the textbook (p.262-263), we take a random subset with 3065 observations as a training set, and the remaining ones as a test set (as indicated in the Indicator file). *Use the test set to evaluate a classifier.*

     You may want to save your random seed so that in the future you can use the same training/test data to evaluate other methods.

   The data and some information on the data are available from the Data link on the textbook homepage.

2. **(FNN vs CNN)** In an CNN with an input dimension of 3×3, the input layer is followed by a convolution layer with 2 kernels of size 2×2 (with stride =1 and no padding). Draw the corresponding (and equivalent) FNN architecture; please mark out clearly which weight parameters are shared. (20 pts)

3. **(CNN with the MNIST data)** Play with the example CNN R/Keras code (or your own code) by changing a few tuning/hyper parameters you like, such as the number of kernels, kernel size, other aspects of the CNN architecture, learning rate, batch size, SGD or its variant, etc. and show how the test results change. (30 pts)

4. **(8000)** Choose two papers from the lists given under Weeks 6-8 on the course Updates page: summarize the main points of each paper and comment. (20 pts)

Please attach your computer program and relevant output.