

PubH 8475/STAT 8056 Homework 1 (Spring 2024)

Due on Feb 19, 2024

1. Apply penalized linear or logistic regression with Ridge, LASSO, Elastic net, at least one of non-convex penalties (e.g. adaptive LASSO, SCAD and TLP), at least one of group/grouping penalties; SIS/ISIS or PCR or PLS (or sPLS); FNN (one or more hidden layers), to one of the following two data sets; your chosen methods can vary between the two datasets. (70 pts: 10 pts for each method-dataset combination.)

- NCI (or 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 CV (e.g. LOOCV or 5- or 10-fold CV) 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 CV or a test dataset to evaluate a classifier.*

The data and some information on the data are available from the Data link on the homepage of "The Elements of Statistical Learning".

2. Read Friedman et al (2007, Ann Appl Statist) and Boyd et al (2011; sections 2.3, 3.1, 6.3, 6.4). For a given dataset $\{(Y_i, X_i) : i = 1, 2, \dots, n\}$ with $X_i = (X_{i1}, \dots, X_{ip})'$, describe a coordinate-descent algorithm and an ADMM algorithm to compute the Lasso estimates:

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^n (Y_i - \sum_{j=1}^p X_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j|.$$

You can assume that all Y_i 's and X_{ij} 's are already centered at 0 (i.e. their sample means are all 0). (30 pts)

3. **(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)

4. **(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)
5. Choose one paper from each of Groups 3 and 4 (other than those in Q 2) on the course Updates page: summarize the main points of each paper and comment. (30 pts)

Please attach your computer program and relevant output.