

**ECE 532, Spring 2019**  
Homework #2  
Due: Tue, Feb 19 @ 2:30 pm

**Problem 1.** The dataset `polydata.csv` contains a set of ordered pairs  $\{(x_i, y_i)\}_{i=1}^n$ , which we will fit using polynomial regression. In other words, we want to find coefficients  $(\beta_0, \dots, \beta_d)$  that minimize the sum of squared residuals

$$\min_{\{\beta_0, \dots, \beta_d\}} \sum_{i=1}^n \left( y_i - (\beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_d x_i^d) \right)^2,$$

where  $d$  is the degree of the polynomial.

- (a) For a fixed value of  $d$ , we can think of this as solving an OLS problem where the data matrix  $X$  is of the form

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^d \\ 1 & x_2 & x_2^2 & \cdots & x_2^d \\ \vdots & & & & \\ 1 & x_n & x_n^2 & \cdots & x_n^d \end{bmatrix}.$$

Perform OLS using  $d = 1, 2, 3, 4, 5$ . Report your coefficients  $\hat{\beta}$  for each value of  $d$ .

- (b) Visualize your answers in (a) by creating a scatter plot of the data, and then overlaying the polynomial approximations  $\hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 + \dots + \hat{\beta}_d x^d$  for different values of  $d$  on the same plot.
- (c) Now use 5-fold CV to find the best value of  $d$  based on the data. Does the answer agree with what you would expect by simply inspecting the plot in (b)?

**Problem 2.**

- (a) The file `brca.csv` contains gene expression data from the The Cancer Genome Atlas (TCGA) project. The data are collected from 500 patients, and the final column records expression levels of the BRCA1

gene, which is known to increase the risk of early onset breast cancer. We will use the expression level of the BRCA1 gene as the response variable and the remaining 17,322 genes as the predictors.

Perform 5-fold CV on the grid of values  $\lambda \in [0, 0.1]$  with spacing 0.01. (For ease of grading, we have already randomly permuted the rows, so you should simply divide the dataset into groups of the first 100, next 100, etc., and then apply CV.)

Plot the curve of the average MSE as a function of  $\lambda$ . What is the optimal value of  $\lambda$ ? Also plot the sparsity (i.e., number of nonzero components of  $\hat{\beta}$ ) as a function of  $\lambda$ .

- (b) Now perform 10-fold CV (again, do not repermute the data—just separate it into chunks of 50). Plot the MSE and sparsity as a function of  $\lambda$ . How does the optimal choice of  $\lambda$  according to 5-fold CV compare with the optimal choice of  $\lambda$  according to 10-fold CV?
- (c) Now use the built-in argument ‘CV’ and the function `lassoPlot` and compare the result with your answers in (a) and (b). Why (if at all) might your answers be different?

### Problem 3.

- (a) The file `brca_reduced.csv` contains a subset of columns of the `brca` dataset corresponding to 112 genes that are identified as important from the Lasso. Perform OLS regression, and report the MSE (don’t forget to include an intercept!). Also perform OLS on the dataset `brca_noisy.csv`, which has been constructed from the `brca_reduced` dataset by randomly changing 10% of the values in the BRCA1 response column by the value 10. Also compute the maximum absolute difference between components of the fitted  $\hat{\beta}$ ’s (also known as the  $\ell_\infty$ -norm) for the clean and contaminated datasets.
- (b) Using the Matlab package `robustfit`, perform robust regression with a Huber loss on the clean and contaminated data sets. Compute the MSE in both cases, as well as the  $\ell_\infty$ -norm between the  $\hat{\beta}$ ’s. Compare your answers to those obtained in part (a).

- (c) Now repeat part (b) with a few more robust losses (Cauchy, Talwar, Welsch) and compare the values of MSE and  $\ell_\infty$ -norm distance for these losses.

**Problem 4.**

- (a) Suppose we were to apply a gradient descent algorithm to optimize the OLS objective:  $\min_{\beta} \|y - X\beta\|_2^2$ . Write down an explicit formula for the iterate  $\beta^t$  in terms of  $\beta^{t-1}$ ,  $X$ ,  $y$ , and the stepsize  $\eta$ .
- (b) Suppose we instead applied the Newton-Raphson algorithm. What is the formula for  $\beta^t$ ? Simplify as much as possible.