COMS 474

Homework 4

Haadi Majeed

Spring 2022

Problem 1

[24 points total (5,5,3,3,4,4)]
Suppose you use lasso fit to a linear model for a data set. Let $\beta^*(\lambda)$ denote the lasso solution for a specific λ (i.e. the coeffcient vector you get for that λ).

Provide explanations for your answers to the following questions.

A

Describe how the training MSE changes as a fucntion of λ, including $\lambda = 0$ and as $\lambda \to \infty$

B

Describe how the hold-out MSE changes as a fucntion of λ, including $\lambda = 0$ and as $\lambda \to \infty$

C

Describe $\beta^*(0)$.

D

Describe what happens to $\beta^*(0)$ as λ grows.

E

If you used ridge regression instead of lasso, explain how your answers to (a).-(d). would differ.

F

We discussed the "constrained form" of lasso, with a constraint of the form

$$\sum_{i=1}^{p} |\beta_j| \leq t$$

Which value, or limiting value, of t coresponds to $\lambda = 0$ and which corresponds to $\lambda \to \infty$

Problem 2

[15 points]
You have already seen formulas for the best intercept in linear models when there are no features $p = 0$ and a single feature $p = 1$. You will now look at what happens with p features when we center the data.
Recall that "centering" a feature means subtracting its mean. For example, if the sample values for feature X_4 are $\begin{bmatrix} 5 \\ 0 \\ 1 \end{bmatrix}$, which has a mean of 2,
we could replace it with $\begin{bmatrix} 5-2 \\ 0-2 \\ 1-2 \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \\ -1 \end{bmatrix}$ which has a mean of 0. Thus if feature X_4 is centred, then $\sum_{i=1}^{n} X_4(i) = 0$.

What is the value of the intercept β_0^* in the ordinary least squares solution, i.e.

$$(\beta_0^*, \beta_1^*, \dots, \beta_p^*) = \underset{(\beta_0^*, \beta_1^*, \dots, \beta_p^*) \epsilon \mathbb{R}^{p+1}}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \left(Y(i) - \beta_0 \sum_{j=1}^{p} \beta_j X_j(i) \right)^2$$

when the features $X_1, \dots, X_p$ are all centered? (e.g. $\sum_{i=1}^{n} X_j(i) = 0$ for $j = 1, \dots, p$). (You do not need to use a second derivative test or solve for $\beta_1^*, \dots, \beta_p^*$, just use the first derivative test $0 = \frac{\partial}{\partial \beta_0}$ MSE).