Statistical Learning - Milan, Fall2019

Homework problem 11

Leaving Out Lemma for ridge regression

Recall that the Leaving Out Lemma requires two conditions under the fixed-X, iid error assumption:

- Linear model: $\hat{\mathbf{y}} = S(X)\mathbf{y}$ in training.
- For any $1 \leq i_0 \leq n$, define a pseudo training dataset $X, \tilde{\mathbf{y}}$ with the same X as our training data, and $\tilde{y}_j = y_j$ for $j \neq i_0$ and:
$$\tilde{y}_{i_0} = \hat{y}_{i_0}^{(-i_0)},$$
where the superscript $(-i_0)$ indicates the model built on $n-1$ observations, leaving out i_0. Then we require:
$$\hat{\tilde{y}}_{i_0} = (S\tilde{\mathbf{y}})_{i_0} = \hat{y}_{i_0}^{(-i_0)}.$$

Under these conditions we proved that:

$$(y_{i_0} - \hat{y}_{i_0}^{(-i_0)}) = \frac{(y_{i_0} - \hat{y}_{i_0})}{1 - S_{i_0 i_0}}.$$

In this problem we will examine the applicability of this result to penalized ridge regression, where:

$$\hat{\beta} = \arg\min_{\beta} \|\mathbf{y} - X\beta\|^2 + \lambda\|\beta\|^2.$$

1. Write $S(X)$ for penalized ridge regression (more accurately, we can write it $S(X, \lambda)$ to emphasize that it can depend on λ as well).
2. Prove that penalized ridge complies with the second condition above, that is, it fulfills the Leaving out Lemma.
Hint: As we did for least squares in class, concentrate on the optimization objective and prove that its optimizer should be the same for $\tilde{\mathbf{y}}$ as for the data without observation i_0.
3. How does the size of the diagonal elements $S(X, \lambda)_{ii}$ affect the (optimism) gap between training and leave-one-out squared error for ridge regression? Consequently, how do you expect $S(X, \lambda)_{ii}$ to behave as a function of λ? Increasing, decreasing, or not changing in a clear manner?
4. Prove your previous claim. If you fail to prove it, supply well designed simulations for partial credit.
Hint: You may find it useful to consider arguments on non-negative-definiteness and differentiation.
5. Moving from penalized ridge regression to penalized lasso, do you expect the Lemma to hold for this problem as well? No need for formal proof, but a relevant and accurate argument is required.