Homework 3
Stat 215A, Fall 2019

Due: provide a hard copy at the beginning of the lab on Friday October 25th or push a `homework3.pdf` file to the `lab3/` folder of your `stat-215-a` GitHub repo by Thursday October 24 11:59pm

1 Bias-Variance Tradeoff: OLS vs Ridge Regression

Consider the regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{y} \in \mathbb{R}^n$, $\mathbf{X} \in \mathbb{R}^{n\times p}$ has rank $p < n$, $\boldsymbol{\beta} \in \mathbb{R}^p$, and $\boldsymbol{\epsilon} \in \mathbb{R}^n$ is random noise with $\mathbb{E}[\boldsymbol{\epsilon}] = \mathbf{0}$ and $\mathrm{Var}(\boldsymbol{\epsilon}) = \sigma^2\,\mathbf{I}$. For the sake of simplicity, assume that $\mathbf{X}$ is fixed (not random) and is centered so that each column of $\mathbf{X}$ has mean 0. Also, assume that $\mathbf{y}$ has been centered to have mean 0.

Recall that OLS solves

$$\hat{\boldsymbol{\beta}}^{OLS} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\ ||\,\mathbf{y} - \mathbf{X}\,\boldsymbol{\beta}\,||_2^2, \tag{1.1}$$

for which we know the solution to be given by

$$\hat{\boldsymbol{\beta}}^{OLS} = (\mathbf{X}^\top\,\mathbf{X})^{-1}\,\mathbf{X}^\top\,\mathbf{y}\,. \tag{1.2}$$

On the other hand, for a given constant $\lambda > 0$, ridge regression solves the following optimization problem:

$$\hat{\boldsymbol{\beta}}^{R} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\ ||\,\mathbf{y} - \mathbf{X}\,\boldsymbol{\beta}\,||_2^2 + \lambda||\,\boldsymbol{\beta}\,||_2^2. \tag{1.3}$$

We know that the ridge solution can be written in closed-form as

$$\hat{\boldsymbol{\beta}}^{R} = (\mathbf{X}^\top\,\mathbf{X} + \lambda\,\mathbf{I})^{-1}\,\mathbf{X}^\top\,\mathbf{y}\,. \tag{1.4}$$

Fix some $\lambda > 0$ for the remainder of the problem.

1. Under the given model and assumptions, compute $\mathbb{E}\left[\hat{\boldsymbol{\beta}}^{OLS}\right]$. Is $\hat{\boldsymbol{\beta}}^{OLS}$ biased for $\boldsymbol{\beta}$?
2. Under the given model and assumptions, compute $\mathbb{E}\left[\hat{\boldsymbol{\beta}}^{R}\right]$. Is $\hat{\boldsymbol{\beta}}^{R}$ biased for $\boldsymbol{\beta}$?
3. Under the given model and assumptions, compute $\mathrm{Var}\left(\hat{\boldsymbol{\beta}}^{OLS}\right)$.
4. Show that

   $$\mathrm{tr}\left(\mathrm{Var}\left(\hat{\boldsymbol{\beta}}^{OLS}\right)\right) = \sigma^2 \sum_{j=1}^{p} \frac{1}{d_j^2},$$

   where $d_1, \ldots, d_p$ are the singular values of $\mathbf{X}$. (Hint: $\mathrm{tr}(\mathbf{A}\,\mathbf{B}) = \mathrm{tr}(\mathbf{B}\,\mathbf{A})$).
5. Under the given model and assumptions, compute $\mathrm{Var}\left(\hat{\boldsymbol{\beta}}^{R}\right)$.

6. Show that

$$\mathrm{tr}\left(\mathrm{Var}\left(\hat{\boldsymbol{\beta}}^{R}\right)\right) = \sigma^2 \sum_{j=1}^{p} \frac{d_j^2}{(d_j^2+\lambda)^2},$$

where again $d_1, \ldots, d_p$ are the singular values of $\mathbf{X}$.

7. Conclude that

$$\mathrm{tr}\left(\mathrm{Var}\left(\hat{\boldsymbol{\beta}}^{OLS}\right)\right) > \mathrm{tr}\left(\mathrm{Var}\left(\hat{\boldsymbol{\beta}}^{R}\right)\right).$$

8. Explain, in words, what this problem illustrates in terms of the bias-variance tradeoff between OLS and ridge regression.