Homework 3
Stat 215A, Fall 2024

Due: push a `homework3.pdf` file **Gradescope** by **Friday, October 18 23:59**

1 Ordinary Least Squares

Suppose that we observe our usual data matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$ and response vector $\mathbf{y} \in \mathbb{R}^n$, where n is the number of samples/observations and p is the number of features. Suppose also that $\mathbf{X}$ has rank $p < n$. Under this setting, the ordinary least squares (OLS) estimator is given by

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

1. Provide an expression for $\hat{\boldsymbol{\beta}}_{OLS}$ in terms of $\mathbf{X}$ and $\mathbf{y}$ by solving the optimization problem above. Why do we require the assumption that $\operatorname{rank}(\mathbf{X}) = p < n$?
2. Show that the OLS predictions $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}_{OLS}$ can be written as $\hat{\mathbf{y}} = \mathbf{H}\,\mathbf{y}$, where $\mathbf{H}^2 = \mathbf{H}$.
3. Prove that the residuals $\hat{\mathbf{r}} = \mathbf{y} - \hat{\mathbf{y}}$ are orthogonal to the OLS predictions $\hat{\mathbf{y}}$. Draw a picture to show what this means geometrically.

2 Miscellaneous

What was the original motivation for the development of the Ridge regression algorithm? What was the original motivation for the development of the LASSO algorithm?