ECE 532, Spring 2019
Homework #1
Due: Tue, Feb 5 @ 2:30 pm

Problem 1. We showed in class that when A is a symmetric matrix, we have $\nabla_\beta(\beta^T A\beta) = 2A\beta$. What is $\nabla_\beta(\beta^T A\beta)$ when A is not necessarily a symmetric matrix? State and derive a general formula.

Problem 2. Consider the matrices $X = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 2 \\ 2 & 3 & 5 \end{bmatrix}$ and $y = \begin{bmatrix} 0 \\ 4 \\ 10 \end{bmatrix}$.

(a) Compute $X^T X$, and verify that $\text{rank}(X^T X) < 3$, so $X^T X$ is not invertible.

(b) What is $\min_\beta \|y - X\beta\|_2^2$? Show that multiple minimizers exist.

Problem 3. Suppose $p = 3$, and we are performing linear regression to predict BMI with β_1 corresponding to the intercept, β_2 corresponding to height, and β_3 corresponding to weight. We want to perform a variant of ridge regression, where only the coefficients β_2 and β_3 are encouraged to be close to 0. Furthermore, we think that β_2 (the coefficient of weight) should be smaller than β_3, so we solve the problem

$$\hat{\beta} = \arg\min_\beta \left\{ \|y - X\beta\|_2^2 + \lambda\beta_2^2 + \frac{\lambda}{2}\beta_3^2 \right\}.$$

What is the generative model interpretation of this objective function? (What is the prior distribution $p(\beta)$?)

Problem 4.

(a) Suppose $x_i \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}\right)$. Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, and define $y_i = Ax_i$. What is the mean of y_i? What is the covariance matrix of y_i?

(b) Suppose instead that $x_i \sim N\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}\right)$. Let $y_i = Ax_i$, where A is defined as in (i). What are the mean and covariance of y_i?

Problem 5.

(a) Generate $n = 100$ observations from the linear model $y_i = x_i^T\beta + \epsilon_i$, where $\beta = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$, the ϵ_i's are i.i.d. draws from a $N(0, 12)$ distribution, and the x_i's are i.i.d. draws from a $N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 10 \end{bmatrix}\right)$ distribution. Compute $\widehat{\beta}_{OLS}$, and construct a 90% confidence interval for $\widehat{\beta}_1$. Does the interval cover β_1?

(b) Keeping X fixed, repeat the simulation in (a) 500 times by regenerating ϵ. What percentage of the time do the confidence intervals cover β_1?

(c) Still keeping X fixed, repeat part (b), this time generating the ϵ_i's from a Uniform$[-6, 6]$ distribution. What percentage of the time do the confidence intervals cover β_1?

Problem 6. The following question uses the dataset `bikeshare.csv`, collected from 2011–2012 from the Capital Bikeshare system in Washington, D.C.

(a) Perform OLS regression of cnt (total count of rented bikes) against holiday (1 if holiday, 0 otherwise), temp (temperature in Celsius), hum (humidity), and windspeed (wind speed). Do not forget to include an intercept! What is $\widehat{\beta}_{OLS}$? Are the coefficients what you would expect?

(b) Now perform ridge regression for various values of $\lambda : \{0.1, 1, 10, 100, 1000\}$ and report the values of $\widehat{\beta}$. What happens as λ increases?