1 Bias Variance for Ridge Regression

Recall the statistical model for ridge regression from lecture. We have a set of samples $\{x_i, y_i\}_{i=1}^n$ and **zero-mean** Gaussian noise z_i. Our model is then the following, where the rows of X are x_i:

$$Y = Xw^* + z$$

Throughout this problem, you may assume $X^T X$ is invertible. Recall both least squares estimators we studied:

$$w_{\text{OLS}} = \min_{w \in \mathbb{R}^d} \|Xw - y\|_2^2$$

$$w_{\text{Ridge}} = \min_{w \in \mathbb{R}^d} \|Xw - y\|_2^2 + \lambda \|w\|_2^2$$

1. **Write the solution** for $w_{\text{OLS}}, w_{\text{Ridge}}$. There's no need to re-derive it.

2. Let $\widehat{w} \in \mathbb{R}^d$ denote any estimator of w_*, the optimal weights. In the context of this problem, an estimator $\widehat{w} = \widehat{w}(X, y)$ is any function which takes the data X and the labels y, and computes a guess of w_*.

   Define the MSE (mean squared error) of the estimator $\widehat{w}$ as

   $$\text{MSE}(\widehat{w}) := \mathbb{E}\left[\left\|\widehat{w} - w_*\right\|_2^2\right].$$

   Above, the expectation is taken with respect to the randomness inherent in the noise z.

   Define $\widehat{\mu} := \mathbb{E}\widehat{w}$. **Show that the MSE decomposes into**

   $$\text{MSE}(\widehat{w}) = \left\|\widehat{\mu} - w_*\right\|_2^2 + \text{Tr}(\text{Cov}(\widehat{w})) .$$

   Hint: Expectation and trace commute, so $E[\text{Tr}(A)] = \text{Tr}(E[A])$ for any square matrix A.

3. **Show that**

   $$E[w_{\text{Ridge}}] = (X^\top X + \lambda I_d)^{-1} X^\top X w_* .$$

   Also compute $E[w_{\text{OLS}}]$ from your expression for $E[w_{\text{Ridge}}]$. **Which estimator is biased, and which estimator is unbiased?**

2 Independence and Multivariate Gaussians

1. For $X = [X_1, \cdots, X_n]^\top \sim \mathcal{N}(\mu, \Sigma)$, **verify that if X_i, X_j are independent (for all $i \neq j$), then Σ must be diagonal, that is, X_i, X_j are uncorrelated.**

2. Let $N = 2$, $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, and $\Sigma = \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix}$. Suppose $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}(\mu, \Sigma)$. **Show that X_1, X_2 are independent if $\beta = 0$.** Recall that two continuous random variables W, Y with joint density $f_{W,Y}$ and marginal densities f_W, f_Y are independent if $f_{W,Y}(w, y) = f_W(w) f_Y(y)$.