Math 7339 Machine Learning and Statistical Learning 2

Homework 1.

Problem 1. Assume the data $\mathscr{D} = (X, \vec{y})$ are centered. That is, $E(X) = \vec{0}$ and $E(y) = 0$. Here, $X \in \mathbb{R}^{N\times d}$ is fixed data matrix. Assume there data was drawn from $y = \vec{\theta}^T \vec{x} + \epsilon$ and $\epsilon \sim Normal(0, \sigma^2)$. The Ridge Regression estimate for $\vec{\theta}$ is given by

$$\hat{\vec{\theta}} = (X^T X + \lambda I)^{-1} X^T \vec{y}$$

Question 1: Find the **Expected value** $E_{\mathscr{D}}(\hat{\vec{\theta}})$ of the Ridge Regression $\hat{\vec{\theta}}$ over data $\mathscr{D} = (X, \vec{y})$.

Question 2: Is $\hat{\vec{\theta}}$ an **unbiased** estimator for $\vec{\theta}$?

Question 3: Find the **Covariance matrix** $Cov(\hat{\vec{\theta}})$ of the Ridge Regression $\hat{\vec{\theta}}$ over data $\mathscr{D} = (X, \vec{y})$.

Question 4: (Bias-Variance Trade-off decomposition) Decompose the **Expected Prediction Error** at a fixed test point $\vec{x}^{(0)}$,

$$EPE(\vec{x}^{(0)}) := E_{D,y^{(0)}}\left[\left(y^{(0)} - \hat{y}^{(0)}\right)^2\right]$$

as irreducible error, $bias^2$ and variance. (Use the results in class, but write each final term concretely using X, λ, σ and $\vec{\theta}$.)

Problem 2.

Assume the data $\mathscr{D} = (X, \vec{y})$ are centered. That is, $E(X) = \vec{0}$ and $E(y) = 0$. Here, $X \in \mathbb{R}^{N\times d}$ is fixed data matrix with rank d. Assume the data was drawn from $y = h(\vec{x}) + \epsilon = \vec{\theta}^T \vec{x} + \epsilon$ and $\epsilon \sim Normal(0, \sigma^2)$

An estimate of $h(\vec{x})$ is given by

$$\hat{h}(\vec{x}) = \vec{\beta}^T X \vec{x}$$

where, the parameter vector $\vec{\beta} \in \mathbb{R}^N$.

Define the cost function as

$$J(\vec{\beta}) := \sum_{j=1}^{N} (\hat{h}(\vec{x}) - y^{(i)})^2 + \lambda \|X^T \vec{\beta}\|^2$$

that is

$$J(\vec{\beta}) := (XX^T\vec{\beta} - \vec{y})^T (XX^T\vec{\beta} - \vec{y}) + \vec{\beta}^T X \lambda X^T \vec{\beta}$$

Question 1. Find the solution $\hat{\vec{\beta}}$ that minimize the cost function. (Hint, write down the matrix form of $J(\vec{\beta})$, then use the matrix calculus to solve $\nabla J = 0$.)

Question 2. Find the **degree of freedom** of the model $\hat{h}(\vec{x})$.