Problem 11

An estimator of an unknown parameter is called unbiased if its expected value equals the true value of the parameter. Here, you will prove that the least-squares estimate given by the normal equation for linear regression is an unbiased estimate of the true parameter θ^*. We first assume that the data:

$$D = \{x^{(i)}, y^{(i)} | 1 \geq i \geq m; x^{(i)} \in \mathbb{R}^d; y^{(i)} \in \mathbb{R}\}$$

comes from the linear model:

$$y^{(i)} = \theta^T x^{(i)} + \epsilon^{(i)}$$

where each $e^{(i)}$ is an independent random variable drawn from a normal distribution with zero mean and variance σ^2 . When considering the bias of an estimator, we treat the input $x^{(i)}$'s as fixed but arbitrary, and the true parameter vector θ^* as fixed but unknown. Expectations are taken over possible realizations of the output values $y^{(i)}$'s.

Part A

Show that $E[\theta] = \theta^*$ for the least squares estimator.

Solution

The goal is to show that $E[\hat{\theta}] = \theta^*$. This is $\hat{\theta}$:

$$\begin{aligned}
\hat{\theta} &= (X^T X)^{-1} X^T Y \\
E[\hat{\theta}] &= E[(X^T X)^{-1} X^T Y] \\
E[\hat{\theta}] &= (X^T X)^{-1} X^T E[Y]
\end{aligned}$$

Since X and Y are fixed (but random), we can move them outside the expectation. Since $\epsilon^{(i)} \sim \mathcal{N}(0, \sigma^2)$:

$$E[Y] = E[\theta^T X + \epsilon] = \theta^T X + E[\epsilon] = \theta^T X$$

Substituting for $E[Y]$:

$$\begin{aligned}
E[\hat{\theta}] &= (X^T X)^{-1} X^T \theta^T X \\
E[\hat{\theta}] &= (X^T X)^{-1} X^T X \theta \\
E[\hat{\theta}] &= \theta
\end{aligned}$$

Part B

Show that the variance of the least squares estimator is $Var(\theta) = (X^T X)^{-1} \sigma^2$.

Solution

The goal is to show $Var(\theta) = (X^T X)^{-1} \sigma^2$. This is $\hat{\theta}$:

$$\begin{aligned}
\hat{\theta} &= (X^T X)^{-1} X^T Y \\
Var[\hat{\theta}] &= Var[(X^T X)^{-1} X^T Y] \\
Var[\hat{\theta}] &= (X^T X)^{-1} X^T Var[Y]((X^T X)^{-1} X^T)^T \\
Var[\hat{\theta}] &= (X^T X)^{-1} X^T Var[Y] X (X^T X)^{-1}
\end{aligned}$$

Since X and Y are fixed (but random), we can move them outside the expectation. Since $\epsilon^{(i)} \sim \mathcal{N}(0, \sigma^2)$:

$$Var(Y) = Var(\theta^T X + \epsilon) = Var(\epsilon) = \sigma^2$$

Substituting for $E[Y]$:

$$Var[\hat{\theta}] = (X^T X)^{-1} X^T \sigma^2 X (X^T X)^{-1}$$
$$Var[\hat{\theta}] = (X^T X)^{-1} \sigma^2$$