Q. . D

m features, d order

1×m m×m m×1

Q 3.1 Part A.

∵ least-squares estimator

∴ $\theta = (X^T X)^{-1} X^T y$

$= (X^T X)^{-1} X^T (X \theta^* + \epsilon)$

(n×d d×1 n×1)

$E(\theta) = E(\theta^* + (X^T X)^{-1} X^T \epsilon)$

$= E(\theta^*) + E((X^T X)^{-1} X^T \epsilon)$ ←

$= E(\theta^*) = \theta^*$

#

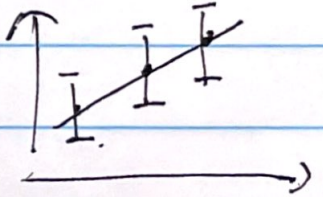
$D = \{(x^{(i)}, y^{(i)}) \mid 1 \le i \le m, x^{(i)} \in \mathbb{R}^d, y^{(i)} \in \mathbb{R}\}$

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

m×d

↓

true param



∵ ϵ is drawn from normal dist,

$E(\epsilon) = 0$