Solution 2.5

Training data: $\underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$, $\quad \underset{\sim}{\varepsilon} \sim N(0, \sigma^2 I)$

For a test point x_0 we want to predict

$$y_0 = x_0^T \beta + \underbrace{\varepsilon_0}_{N(0,\sigma^2)}$$

single observation in $\mathbb{R}$

We use to predict y_0: $\quad \hat{y}_0 = x_0^T \hat{\beta}$

Then $EPE(x_0) = E\left[(y_0 - \hat{y}_0)^2\right]$

Here $\hat{y}_0 = x_0^T \hat{\beta} = x_0^T (X^T X)^{-1} X^T y$

$$= x_0^T (X^T X)^{-1} X^T (X\beta + \varepsilon)$$

$$= x_0^T \beta + x_0^T (X^T X)^{-1} X^T \varepsilon$$

so – by taking expectation over the random training data and the random new y_0, we get

$$EPE(x_0) = E\left[(y_0 - \hat{y}_0)^2\right]$$

$$= E\left[\left(\varepsilon_0 - x_0^T (X^T X)^{-1} X^T \varepsilon\right)^2\right]$$

$$= Var(\varepsilon_0) + Var(x_0^T(X^TX)^{-1}X^T\varepsilon)$$

(since $E(\varepsilon_0)=0$ and $E(\varepsilon)=0$ and ε, X are independent ~~&~~ is there no bias-term)

$$= \sigma^2 + E_X[Var(x_0^T(X^TX)^{-1}X^T\varepsilon \mid X)]$$

$$+ Var_X[\underbrace{E(x_0^T(X^TX)^{-1}X^T\varepsilon \mid X)}_{0}]$$

[using $Var\, Z = E[Var(Z|X)] + Var[E(Z|X)]$]

$$= \sigma^2 + E_X[x_0^T(X^TX)^{-1}X^T\sigma^2 I \cdot X(X^TX)^{-1}x_0]$$

Using $Var(AZ) = A\,Cov(Z)\,A^T$ if A is a matrix of constants and Z is a random vector

$$= \sigma^2 + \sigma^2 E_X[x_0^T(X^TX)^{-1}x_0]$$

$$= \sigma^2 + \sigma^2 x_0^T E_X[(X^TX)^{-1}]x_0 \qquad (*)$$

which is (2.27).

Now, as $N \to \infty$ we have $\frac{1}{N}(X^TX)^{-1} \to Cov(X)$

if we assume that $E(X) = 0$

[which corresponds to assuming centered covariates]

Hence (*) $\approx \sigma^2 + \sigma^2 x_0{}^T \operatorname{Cov}(X)^{-1} x_0 N^{-1}$

Finally, taking expectation over the point x_0 (which we have not done above), we use that

$$\operatorname{Cov}(x_0) = \operatorname{Cov}(X),$$

and get

$$E_{x_0} \, EPE(x_0) = E_{x_0}\left[x_0{}^T \operatorname{Cov}(X)^{-1} x_0\right] \cdot \frac{\sigma^2}{N} + \sigma^2$$

$$= E_{x_0}\left[\operatorname{trace}\left(x_0 x_0{}^T \operatorname{Cov}(X)^{-1}\right)\right] \cdot \frac{\sigma^2}{N} + \sigma^2$$

$$= \operatorname{trace}\Big[\underbrace{\overbrace{E(x_0 x_0{}^T)}^{\operatorname{Cov}(x_0)} \operatorname{Cov}(X)^{-1}}_{I}\Big] \cdot \frac{\sigma^2}{N} + \sigma^2$$

$$= \operatorname{trace}[I] \cdot \frac{\sigma^2}{N} + \sigma^2$$

$$= \sigma^2 \cdot \frac{p}{N} + \sigma^2$$ which is hence <u>linear</u> in p