The Elements of statistical learning

Ex 3.2

Ex. 3.2 Given data on two variables X and Y, consider fitting a cubic polynomial regression model $f(X) = \sum_{j=0}^{3} \beta_j X^j$. In addition to plotting the fitted curve, you would like a 95% confidence band about the curve. Consider the following two approaches:

1. At each point x_0, form a 95% confidence interval for the linear function $a^T\beta = \sum_{j=0}^{3} \beta_j x_0^j$.
2. Form a 95% confidence set for β as in (3.15), which in turn generates confidence intervals for $f(x_0)$.

How do these approaches differ? Which band is likely to be wider? Conduct a small simulation experiment to compare the two methods.

1 ère méthode:

Approche "point par point" avec x_0

$$\hat{Y}_0 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

$$\hat{Y}_0 = X_0 \hat{\beta}$$

Where:

$$X_0 = [1, x_0, x_0^2, x_0^3]$$

$$\hat{\beta} = [\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3]^T$$

donc

$$\mathrm{Var}(\hat{Y}_0) = X_0 \,\mathrm{Var}(\hat{\beta})\, X_0^T$$

$$= X_0 \underbrace{(X^T X)^{-1}} X_0^T$$

Matrice Var-Covar
des coeffs

Vu qu'on veut un intervalle à 95%:

$$\boxed{\hat{Y}_0 \pm 1{,}96 \sqrt{X_0 (X^T X)^{-1} X_0^T}}$$

2ème méthode:

1ère étape:

$Y = X\beta + \varepsilon$ where $\varepsilon \sim \mathcal{N}(0, \sigma^2)$

2ème étape: Estimateur des moindres carrés, $\hat{\beta}$:

$\hat{\beta} = (X^T X)^{-1} X^T Y$ where $\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2 (X^T X)^{-1})$

3ème étape: standardisation

$$\frac{(\hat{\beta} - \beta)}{\sigma} \sim \mathcal{N}(0, (X^T X)^{-1})$$

$$(X^T X)^{(1/2)} (\hat{\beta} - \beta) \sim \mathcal{N}(0, 1)$$

4ème étape: Chi-Carré

Si $Z \sim \mathcal{N}(0, 1)$ de dimension p, alors

$Z^T Z \sim \chi^2_p$

$$\bullet \left[\frac{(X^TX)^{1/2}(\hat{\beta}-\beta)}{\sigma}\right]^T \left[\frac{(X^TX)^{1/2}(\hat{\beta}-\beta)}{\sigma}\right] \sim \chi^2_4$$

$$= (\hat{\beta}-\beta)^T X^TX(\hat{\beta}-\beta)/\sigma^2 \sim \chi^2_4$$

95 % de confiance $\alpha = 0,05$

$$P\left((\hat{\beta}-\beta)^T X^TX(\hat{\beta}-\beta)/\sigma^2 \leq \chi^2_{4,0.05}\right) = 0,95$$

$$\boxed{C_\beta = \{\beta \mid (\hat{\beta}-\beta)^T X^TX(\hat{\beta}-\beta) \leq \sigma^2 \chi^2_{4,0.05}\}}$$

\- région de confiance de la 1ère méthode va donc être plus large que la 2ème.

Pour le code voir mon github