QBUS6810
Statistical Learning and Data Mining

Tutorial 6 (Written Exercises)

Question 1

Show that the OLS estimator is unbiased, i.e., derive the following:

$$E\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}.$$

Treat the x values as fixed (i.e. non-random) and use the formula for the OLS estimator.

Question 2

Let $y_1, \dots, y_n$ be a sample from a distribution with the density function $p(y; \theta) = \theta y^{\theta-1}$ for $0 < y < 1$, where $\theta > 0$.

Find $\widehat{\theta}$, the maximum likelihood estimator of θ.

Compute $\widehat{\theta}$ for the sample $y_1 = 0.35$, $y_2 = 0.28$, $y_3 = 0.91$.

Question 3

Consider the following penalized least-squares estimator, called the *Ridge regression estimator* (discussed in Lecture 6):

$$\widehat{\boldsymbol{\beta}}_{\text{ridge}} = \underset{\boldsymbol{\beta}}{\text{argmin}} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\}$$

Note that OLS is a special case of Ridge, corresponding to $\lambda = 0$.

Show that if we set $\lambda = \sigma^2/\tau^2$, the ridge regression estimator is the posterior mode (i.e. the MAP estimator) in a Gaussian linear regression model with the prior on the regression coefficients under which β_j are independent $N(0, \tau^2)$, for $j = 1, ..., p$. Here we are not putting an informative prior on the intercept β_0 (this is equivalent to using a flat prior density for β_0, i.e., a density that is proportional to the constant 1).