ST451 Bayesian Machine Learning

Exercises

1. Let $y = (y_1, \dots, y_n)$ be a random sample from a $\mathrm{N}(\theta, \sigma^2)$ distribution with σ^2 known.

   (a) Show that the likelihood is proportional to

   $$f(y|\theta) \propto \exp\left(-\frac{n(\bar{y}-\theta)^2 + (n-1)S^2}{2\sigma^2}\right).$$

   where $\bar{x}$ is the sample mean and S^2 is the sample variance

   $$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2.$$

   Hence the likelihood simplifies to

   $$f(y|\theta) \propto \exp\left(-\frac{(\theta-\bar{y})^2}{2\frac{\sigma^2}{n}}\right)$$

   (b) Set the prior for θ to be $\mathrm{N}(\mu, \tau^2\sigma^2)$ and derive its posterior distribution. (You can use the above result)

2. Suppose that $y_i \sim N(\mu, 1)$ for $i = 1, \dots, n$ and that the y_i's are independent.

   (a) Show that the sample mean estimator $\hat{\mu}_1 = \frac{1}{n}\sum_i^n y_i$ is obtained from minimising the least squares criterion

   $$\hat{\mu}_1 = \underset{\mu}{\operatorname{argmin}} \sum_{i=1}^{n}(y_i - \mu)^2,$$

   and that $\hat{\mu}_1$ an unbiased estimator of μ. Also find the variance of $\hat{\mu}_1$.

   (b) Consider adding a penalty term to the least squares criterion, and therefore using the estimator that minimises

   $$\hat{\mu}_2 = \underset{\mu}{\operatorname{argmin}} \sum_{i=1}^{n}(y_i - \mu)^2 + \lambda\mu^2$$

   for the mean, where λ is a non-negative tuning parameter. Derive $\hat{\mu}_2$, find it bias and show that its variance is lower than that of $\hat{\mu}_1$

   (c) Find a Bayes estimator assuming the $\mathrm{N}(0, 1/\lambda)$ as prior for μ. Compare with your answer in the previous part.

3. Load the dataset 'diabetes' from the scikit-learn library in Python. Fit a linear regression and a ridge regression model and assess their predictive performance by splitting the data into a training and test dataset.