ST451 Bayesian Machine Learning
Week 3

Exercises

1. Let $y = (y_1, \ldots, y_{10})$ be a sample of independent and identically distributed (iid) random variables from the Normal distribution with unknown mean θ and known variance 1. The sample mean $\bar{y}$ is recorded to be 0.3.

   (a) Suppose that we are interested in only two values of θ s, i.e. 0 and 1, and no prior knowledge is available. Assign a suitable prior on θ, justifying your choice, and derive the corresponding posterior. Consider the hypotheses $H_0 : \theta = 0$ and $H_1 : \theta = 1$, calculate the Bayes factor for comparing H_1 in reference to H_0 and interpret its value.

   Answer: If we assume $\pi(\theta_0) = \pi(\theta_1)$ the Bayes factor reduces to the ratio of the likelihoods evaluated at these points.

   The likelihood function can be written as

$$
\begin{aligned}
f(x|\theta) &= \prod_{i=1}^{n} (2\pi)^{-1/2} \exp\left\{-\frac{1}{2}(y_i - \theta)^2\right\} \propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(y_i - \theta)^2\right\} \\
&= \exp\left(-\frac{1}{2}\sum_{i=1}^{n} y_i^2 + \theta \sum_{i=1}^{n} y_i - \frac{n}{2}\theta^2\right) \propto \exp\left(\theta \sum_{i=1}^{n} y_i - \frac{n}{2}\theta^2\right)
\end{aligned}
$$

   For $\theta = 0$ we get $f(x|\theta = 0) = 1$ whereas for $\theta = 1$ we get $\exp(n/2 - n\bar{y}) = \exp(-2)$. Hence $B_{01} = \exp(2) \approx 7.38$, so we conclude that there exists substantial evidence in favour of H_0.

   (b) Now consider the hypotheses $H_0 : \theta \leq 0$ and $H_1 : \theta > 0$. Calculate the Bayes factor for comparing H_1 in reference to H_0 and interpret its value.

   Answer: We will assign the unit information prior $N(0, 1)$ (given that $\sigma^2 = 1$). The justification for this is that we want to add some information pointing towards 0 but only of the magnitude of one observation. Other options like setting the prior variance to a large value or using the Jeffreys prior who is improper are also justified for this problem as both hypotheses contain intervals of θ so there is no fear of the Jeffreys-Lindley paradox.

   With $N(0, 1)$ as the prior for θ the posterior was shown in the slides to be $\mathrm{N}(\mu_n, \tau_n^2)$ with

$$
\mu_n = \frac{n}{n+1}\bar{y} = \frac{3}{11}, \quad \tau_n^2 = \tfrac{1}{n+1} = \tfrac{1}{11}
$$

   We can now compute posterior and prior probabilities of $\theta > 0$ and $\theta \leq 0$ based on the $N(3/11, 1/11)$ and $N(0, 1)$. Under $N(0, 1)$ both hypotheses have equal probability, so it suffices to calculate

$$
\pi(H_0|y) = \pi(\theta \leq 0|y) = \pi\left(\frac{\theta - 3/11}{1/\sqrt{11}} \leq -\frac{3/11}{1\sqrt{11}}|y\right) = \Phi(-3/\sqrt{11}) = 0.183
$$

   The Bayes factor in favour of the alternative is then $B_{10} = (1 - 0.183)/0.183 = 4.47$ so we conclude that there exists substantial evidence in favour of H_1 (against H_0)

   (c) Finally, consider the hypotheses $H_0 : \theta = 0$ and $H_1 : \theta \neq 0$. Can you calculate the Bayes factor in this case? Justify your answers.

   Answer: As also shown in the lecture slides the Bayes factor in this case is equal to

$$
B_{01} = \frac{\exp\left(-n\bar{y}^2/2\right)}{\int_{-\infty}^{+\infty} \exp\{-n(\bar{y} - \theta)^2/2\}(2\pi)^{-1/2}\exp(-\theta^2/2)\,\mathrm{d}\theta}
$$

where we used the unit information prior of the previous part. Further calculations yield

$$B_{01} = \frac{(2\pi)^{1/2}\exp\left(-n\bar{y}^2/2\right)\exp\left(\frac{n}{2}\bar{y}^2\right)}{\int \exp\left(-\frac{1}{2/n}(-2\theta\bar{y}+\theta^2\frac{n+1}{n})\right)d\theta}$$
$$= \frac{(2\pi)^{1/2}}{\int \exp\left(-\frac{\theta^2-2\theta\frac{n\bar{y}}{n+1}}{2\frac{1}{n+1}}\right)d\theta}$$

Generally speaking if $\theta \sim N(\mu, \sigma^2)$ we know that

$$\int (2\pi\sigma^2)^{-1/2}\exp\left(-\frac{\theta^2-2\theta\mu+\mu^2}{2\sigma^2}\right)d\theta = 1 \text{ or else}$$
$$\int \exp\left(-\frac{\theta^2-2\theta\mu}{2\sigma^2}\right)d\theta = (2\pi\sigma^2)^{1/2}\exp\left(\frac{\mu^2}{2\sigma^2}\right)$$

Substituting in the above for $\mu = \frac{n\bar{y}}{n+1}$, $\sigma^2 = \frac{1}{n+1}$ gives

$$B_{01} = \frac{(2\pi)^{1/2}}{(2\pi)^{1/2}(n+1)^{-1/2}\exp\left(\frac{1}{2}\frac{n^2\bar{y}^2}{(n+1)^2}(n+1)\right)} = (n+1)^{1/2}\exp\left(-\frac{n^2\bar{y}^2}{2(n+1)}\right) = 2.203$$

Hence we have evidence in favour of H_0 but this is considered as poor.

2. Load the dataset ‘Boston’ from the scikit-learn library in Python. Fit a linear regression with the price as dependent variable and all the other variables as predictors. Provide the output using both the Bayesian and the maximum likelihood approaches.

   Answer: Code from the computer class can be used directly.