BIM3008-Assignment2

Junyang Deng (120090791)

November 2, 2022

1. Write the log likelihood for a multinomial sample and show equation (4.6).

$$\hat{p}_i = \frac{\sum_t x_i^t}{N}$$

Answer:
In a multinomial sample, the outcome of a random event is one of K mutually exclusive and exhaustive states, each of which has a probability of occuring p_i with $\sum_{i=1}^K p_i$. Let $x_1, x_2, ..., x_K$ be indicators where $x_i = 1$ if the outcome is state i and 0 otherwise.
In one experiment,

$$P(x_1, x_2, \ldots, x_K) = \prod_{i=1}^{K} p_i^{x_i}$$

We do N such independent experiments with outcomes $\mathscr{X} = \{\boldsymbol{x}^t\}_{t=1}^N$ where $\sum_i x_i^t = 1$ and

$$x_i^t = \begin{cases} 1 & \text{if experiment } t \text{ chooses state } i \\ 0 & \text{otherwise} \end{cases}$$

The constraint is $\sum_i p_i = 1$.

We add the constraint as a Lagrange term and maximize the formula below.

$$\begin{aligned}
J(p_i) &= \sum_i \sum_t x_i^t \log p_i + \lambda \left(1 - \sum_i p_i\right) \\
\frac{\partial J(p_i)}{p_i} &= \sum_t \frac{x_i^t}{p_i} - \lambda = 0 \\
\lambda &= \sum_t \frac{x_i^t}{p_i} \Rightarrow p_i \lambda = \sum_t x_i^t \\
\sum_i p_i \lambda &= \sum_i \sum_t x_i^t \Rightarrow \lambda = \sum_t \sum_i x_i^t \\
p_i &= \frac{\sum_t x_i^t}{\sum_t \sum_i x_i^t} = \frac{\sum_t x_i^t}{N}, \text{ since} \sum_i x_i^t = 1
\end{aligned}$$

2. Write the code that generates a normal sample with given μ and σ, and the code that calculates m and s from the sample. Do the same using the Bayes' estimator assuming a prior distribution for μ.

Answer:

3. Assume a linear model and then add 0-mean Gaussian noise to generate a sample. Divide your sample into two as training and validation sets. Use linear regression using the training half. Compute error on the validation set. Do the same for polynomials of degrees 2 and 3 as well.

Answer:

4. When the training set is small, the contribution of variance to error may be more than that of bias and in such a case, we may prefer a simple model even though we know that it is too simple for the task. Can you give an example?

Answer:

5. Generate a sample from a multivariate normal density $N(\mu, \Sigma)$, calculate m and S, and compare them with μ and Σ. Check how your estimates change as the sample size changes.

   Answer:

6. Generate samples from two multivariate normal densities$N(\mu_i, \Sigma_i)$, i = 1, 2, and calculate the Bayes' optimal discriminant for the four cases in table 5.1.

   Answer:

7. In figure 6.11, we see a synthetic two-dimensional data where LDA does a better job than PCA. Draw a similar dataset where PCA and LDA find the same good direction. Draw another where neither PCA nor LDA find a good direction.

   Answer: