

Solution for HW2

Tianyi Luo
Ehsan Amid

1.3 Suppose that we have three coloured boxes r (red), b (blue), and g (green). Box r contains 3 apples, 4 oranges, and 3 limes, box b contains 1 apple, 1 orange, and 0 limes, and box g contains 3 apples, 3 oranges, and 4 limes. If a box is chosen at random with probabilities $p(r) = 0.2$, $p(b) = 0.2$, $p(g) = 0.6$, and a piece of fruit is removed from the box (with equal probability of selecting any of the items in the box), then what is the probability of selecting an apple? If we observe that the selected fruit is in fact an orange, what is the probability that it came from the green box?

Solution:

(1): Let we define the following events:

A: An apple is selected; B1: Red box is selected; B2: Blue box is selected; B3: Green box is selected.

Based on the definitions of the these events, sum rule and product rule, the marginal probability of of A is as following:

$$\begin{aligned} P(A) &= P(A|B1) * P(B1) + P(A|B2) * P(B2) + P(A|B3) * P(B3) \\ &= \frac{3}{10} * 0.2 + \frac{1}{2} * 0.2 + \frac{3}{10} * 0.6 = 0.06 + 0.1 + 0.18 = 0.34 \end{aligned}$$

(2): Let we define the following events:

C: An orange is selected; B1: Red box is selected; B2: Blue box is selected; B3: Green box is selected.

According to the definitions of the these events, bayes' theorem and product rule, the probability of C is as following:

$$\begin{aligned} P(B3|C) &= \frac{P(C|B3) * P(B3)}{P(C|B1) * P(B1) + P(C|B2) * P(B2) + P(C|B3) * P(B3)} \\ &= \frac{\frac{3}{10} * 0.6}{\frac{4}{10} * 0.2 + \frac{1}{2} * 0.2 + \frac{3}{10} * 0.6} = \frac{0.18}{0.36} = 0.5 \end{aligned}$$

1.9 Show that the mode (i.e. the maximum) of the Gaussian distribution (1.46) is given by μ . Similarly, show that the mode of the multivariate Gaussian (1.52) is given by μ .

Solution:

(1): If $X \sim \mathcal{N}(\mu, \sigma^2)$, its likelihood function is as following:

$$\begin{aligned} L(\mu, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \\ &= (2\pi)^{-n/2} (\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2} \end{aligned}$$

We take \ln on both sides:

$$\ln(L) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

The extremum values can be achieved through setting the derivative with respect to μ equals to 0:

$$\frac{\partial}{\partial \mu} \ln L = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0$$

So we can get: $\mu_{ML} = \frac{1}{n} \sum_{i=1}^n x_i = \mu$.

(2):

Similar to (1), the likelihood function of multivariate case is as following:

$$\begin{aligned} L(\mu, \Sigma) &= \prod_{i=1}^n \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} e^{-\frac{1}{2}(x_i - \mu)^T \Sigma^{-1}(x_i - \mu)} \\ &= \frac{1}{(2\pi)^{Dn/2}} \frac{1}{|\Sigma|^{n/2}} e^{-\frac{1}{2} \sum_{i=1}^n (x_i - \mu)^T \Sigma^{-1}(x_i - \mu)} \end{aligned}$$

We take \ln on both sides:

$$\ln(L) = -\frac{Dn}{2} \ln(2\pi) - \frac{n}{2} |\Sigma| - \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^T (\Sigma)^{-1} (x_i - \mu)$$

We need to utilize a conclusion of matrix derivative:

$$\frac{\partial x^T A x}{\partial x} = (A + A^T)X$$

The extremum values can be achieved through setting the derivative with respect to μ equals to 0:

$$\begin{aligned} \frac{\partial}{\partial \mu} \ln L &= \frac{1}{2} \sum_{i=1}^n ((\Sigma)^{-1}(x_i - \mu) + ((\Sigma)^{-1})^T(x_i - \mu)) \\ &= \sum_{i=1}^n (\Sigma)^{-1}(x_i - \mu) = 0 \\ &\text{because } (\Sigma)^{-1} \text{ is symmetry and we have } (\Sigma)^{-1} = ((\Sigma)^{-1})^T. \end{aligned}$$

So we can get: $\mu_{ML} = \frac{1}{n} \sum_{i=1}^n x_i = \mu$.

1.11 By setting the derivatives of the log likelihood function (1.54) with respect to μ and σ^2 equal to zero, verify the results (1.55) and (1.56).

Solution:

If $X \sim \mathcal{N}(\mu, \sigma^2)$, its likelihood function is as following:

$$\begin{aligned} L(\mu, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \\ &= (2\pi)^{-n/2} (\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2} \end{aligned}$$

We take \ln on both sides:

$$\ln(L) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

The extremum values can be achieved through setting the derivative with respect to μ equals to 0:

$$\begin{cases} \frac{\partial}{\partial \mu} \ln L = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0 \\ \frac{\partial}{\partial \sigma^2} \ln L = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 = 0 \end{cases}$$

Solving the first equation, we get: $\mu_{ML} = \frac{1}{n} \sum_{i=1}^n x_i$, take it into the second equation, we have: $\sigma_{ML}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_{ML})^2$
Thus, we have:

$$\mu_{ML} = \frac{1}{N} \sum_{n=1}^N x_n$$

$$\sigma_{ML}^2 = \frac{1}{N} \sum_{n=1}^N (x_n - \mu_{ML})^2$$

As required.

1.29 Consider an M-state discrete random variable x , and use Jensen's inequality in the form (1.115) to show that the entropy of its distribution $p(x)$ satisfied $H[x] \leq \ln M$

Solution:

Since the random variable x contains M states, so the entropy expression of its distribution $p(x)$ is as following:

$$H(x) = - \sum_{i=1}^M p(x_i) \ln p(x_i) = \sum_{i=1}^M p(x_i) \ln p(x_i)^{-1}$$

And based on Jensen's inequality:

$$\text{convex function } f(x) \text{ satisfies : } f\left(\sum_{i=1}^M \lambda_i x_i\right) \leq \sum_{i=1}^M \lambda_i f(x_i)$$

$$\text{concave function } f(x) \text{ satisfies : } \sum_{i=1}^M \lambda_i f(x_i) \leq f\left(\sum_{i=1}^M \lambda_i x_i\right)$$

Because \ln is a concave function, we have:

$$H(x) = \sum_{i=1}^M p(x_i) \ln p(x_i)^{-1} \leq \ln\left(\sum_{i=1}^M p(x_i) p(x_i)^{-1}\right) = \ln M$$

1.30 Evaluate the Kullback-Leibler divergence (1.113) between two Gaussians $p(x) = \mathcal{N}(x|\mu, \sigma^2)$ and $q(x) = \mathcal{N}(x|m, s^2)$.

Solution:

Based on (1.113), the expression of Kullback-leibler divergence is as following:

$$KL(p||q) = - \int p(x) \ln q(x) dx - \left(- \int p(x) \ln p(x) dx \right)$$

$$= \int p(x) \ln p(x) dx - \int p(x) \ln q(x) dx$$

Based on the expression of Gaussian distribution introduced in the textbook, we have

$$\begin{aligned}\mathcal{N}(x|\mu, \sigma^2) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\ \mathcal{N}(x|m, s^2) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2s^2}}\end{aligned}$$

Based on (1.104) and (1.110),

$$\int p(x) \ln p(x) dx = -H[x] = -\left(\frac{1}{2}\{1 + \ln(2\pi\sigma^2)\}\right)$$

Based on Exercices 1.7, (1.106) and some conclusions about mean and variance of Gaussian distribution, we have:

$$\begin{aligned}\int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) dx &= 1 \\ \int_{-\infty}^{\infty} x \mathcal{N}(x|\mu, \sigma^2) dx &= \mu \\ \int_{-\infty}^{\infty} x^2 \mathcal{N}(x|\mu, \sigma^2) dx &= \mu^2 + \sigma^2\end{aligned}$$

$$\begin{aligned}\int p(x) \ln q(x) dx &= -\frac{1}{2} \int \mathcal{N}(x|\mu, \sigma^2) \left(\ln(2\pi s^2) + \frac{(x-m)^2}{s^2} \right) dx \\ &= -\frac{1}{2} (\ln(2\pi s^2) \int \mathcal{N}(x|\mu, \sigma^2) dx + \frac{1}{s^2} \int \mathcal{N}(x|\mu, \sigma^2) (x^2 - 2xm + m^2) dx) \\ &= -\frac{1}{2} (\ln(2\pi s^2) + \frac{1}{s^2} (\mu^2 + \sigma^2 - 2\mu m + m^2))\end{aligned}$$

So we have

$$\begin{aligned}KL(p||q) &= \int p(x) \ln p(x) dx - \int p(x) \ln q(x) dx \\ &= -\left(\frac{1}{2}\{1 + \ln(2\pi\sigma^2)\}\right) + \frac{1}{2} (\ln(2\pi s^2) + \frac{1}{s^2} (\mu^2 + \sigma^2 - 2\mu m + m^2)) \\ &= -\frac{1}{2} (-1 + 2 \ln(\frac{s}{\sigma}) + \frac{1}{s^2} (\mu^2 + \sigma^2 - 2\mu m + m^2))\end{aligned}$$

1.31 Consider two variables x and y having joint distribution $p(x,y)$. Show that the differential entropy of this pair of variables satisfies

$$H[x, y] \leq H[x] + H[y]$$

with equality if, and only if, x and y are statistically independent.

Solution:

Based on (1.112), we have:

$$H[x, y] = H[y|x] + H[x]$$

So we only need to prove:

$$H[y|x] \leq H[y]$$

According to (1.121), we have:

$$I[x, y] = H[x] - H[x|y] = H[y] - H[y|x]$$

And the textbook told us, "From the properties of the Kullback-Leibler divergence, we see that $I(x, y) \geq 0$ with equality if, and only if, x and y are independent." As required.

1.40 By applying Jensen's inequality (1.115) with $f(x) = \ln x$, show that the arithmetic mean of a set of real numbers is never less than their geometrical mean.

Solution:

The definitions of arithmetic mean and the geometrical mean are as following:

$$arith_{mean} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$geome_{mean} = \sqrt[n]{\prod_{i=1}^n x_i}$$

And we add \ln to two sides and can have:

$$\ln arith_{mean} = \ln\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \ln\left(\sum_{i=1}^n \frac{1}{n} x_i\right)$$

$$\ln geome_{mean} = \frac{1}{n} \sum_{i=1}^n \ln x_i = \sum_{i=1}^n \frac{1}{n} \ln x_i$$

And based on Jensen's inequality:

$$\text{convex function } f(x) \text{ satisfies : } f\left(\sum_{i=1}^M \lambda_i x_i\right) \leq \sum_{i=1}^M \lambda_i f(x_i)$$

$$\text{concave function } f(x) \text{ satisfies : } \sum_{i=1}^M \lambda_i f(x_i) \leq f\left(\sum_{i=1}^M \lambda_i x_i\right)$$

Because \ln is a concave function, we have:

$$\ln geome_{mean} = \sum_{i=1}^n \frac{1}{n} \ln x_i \leq \ln\left(\sum_{i=1}^n \frac{1}{n} x_i\right) = \ln arith_{mean}$$

$$\Leftrightarrow geome_{mean} \leq arith_{mean}$$