Machine Learning Homework Week 1

Trinh Thi Huong Lien

August 2022

1 Question 1

1.1 Proof that Gaussian distribution is normalised.

Firs, we all know that the the Gaussian distribution is the following:

$$p(X|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{1}$$

To proof that this distribution is normalised, we have to proof the following equation:

$$\int_{-\infty}^{\infty} p(X|\mu,\sigma^2) = 1 \tag{2}$$

Let assume that the mean in equation (1) is zero. Then we introduce the following corresponding equation:

$$I = \int_{-\infty}^{\infty} exp\left(\frac{-x^2}{2\sigma^2}\right) \tag{3}$$

dx This integral can not be calculated directly, therefore, we should take the square of both sides of the equation (3) to get:

$$I^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} exp\left(\frac{-x^2}{2\sigma^2} - \frac{y^2}{2\sigma^2}\right) dxdy \tag{4}$$

Then we change the integral to polar coordinates and below is the transformation of the variables:

$$x = rcos\theta \tag{5}$$

$$y = rsin\theta \tag{6}$$

Besides, we have $sin^2\theta + cos^2\theta = 1$, therefore $x^2 + y^2 = r^2$. We have the Jacobian matrix is given by:

$$dxdy = \begin{bmatrix} \frac{\partial(x)}{\partial(r)} & \frac{\partial(y)}{\partial(r} \\ \frac{\partial(x)}{\partial(\theta)} & \frac{\partial(y)}{\partial(\theta} \end{bmatrix} drd\theta = \begin{bmatrix} cos\theta & -rsin\theta \\ sin\theta & rcos\theta \end{bmatrix} drd\theta = (rcos^2\theta + rsin^2\theta) drd\theta = rdrd\theta \tag{7}$$

We have the limitation of x and y is from $-\infty$ to $+\infty$, so that the limitation of θ is from 0 to 2π and the limitation of r is from 0 to $+\infty$. Thus equation (4) can be rewritten as:

$$I^2 = \int_0^{2\pi}\int_0^{\infty} exp\left(\frac{-r^2}{2\sigma^2}\right) rdrd\theta = 2\pi\int_0^{\infty} exp\left(\frac{-r^2}{2\sigma^2}\right) rdr \tag{8}$$

We substitute $r^2 = u => 2rdr = du => rdr = \frac{du}{2}$, so that we have:

$$I^2 = 2\pi\int_0^{\infty} exp\left(\frac{-u}{2\sigma^2}\right)\frac{du}{2} = \pi\int_0^{\infty} exp\left(\frac{-u}{2\sigma^2}\right) du = \pi\left[exp(\frac{-u}{2\sigma^2})(-2\sigma^2)\right] = 2\pi\sigma^2 \tag{9}$$

$\Rightarrow I = \sqrt{2\pi\sigma^2}$

Finally, we have:

$$\int_{-\infty}^{\infty} p(X|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx = \frac{1}{\sqrt{2\pi\sigma^2}} \times I = 1 \tag{10}$$

1.2 Proof that mean of Gaussian distribution is μ

Firs, we all know that the the Gaussian distribution is the following:

$$p(X|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx \tag{11}$$

$$=> E[x] = \int_{-\infty}^{\infty} x \times \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx \tag{12}$$

$$=> E[x] = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} x \times exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx \tag{13}$$

Let $Z = \frac{X-\mu}{\sigma} \Rightarrow X = \mu + \sigma Z$ and $dx = \sigma dZ$, we have the following equation:

$$E(X) = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} (\mu + \sigma Z) \times exp(\frac{-Z^2}{2})dz \tag{14}$$

$$E(X) = \frac{1}{\sqrt{2\pi}}\left[\int_{-\infty}^{\infty} \mu \times exp(\frac{-Z^2}{2})dz + \int_{-\infty}^{\infty} \sigma Z \times exp(\frac{-Z^2}{2})dz\right] \tag{15}$$

We have $\mu \times exp(\frac{-Z^2}{2})dz$ is the even function since f(-z) = f(z), and $\sigma Z \times exp(\frac{-Z^2}{2})$ is the odd function since -f(z) = f(-z). Also, we have some properties. With even function, we have:

$$f(x) = f(-x) and \int_{-\infty}^{\infty} f(x)dx = 2\int_0^{\infty} f(x)dx \tag{16}$$

With odd function, we have:

$$-f(x) = f(-x) and \int_{-\infty}^{\infty} f(x)dx = 0 \tag{17}$$

Therefore, we have:

$$E(X) = \frac{2}{\sqrt{2\pi}}\mu \int_0^\infty exp(\frac{-Z^2}{2})dz \quad (18)$$

Substiture $u = \frac{Z^2}{2} \Rightarrow du = \frac{2zdz}{2} = zdz \Rightarrow dz = \frac{du}{\sqrt{2u}}$

$$E(X) = \frac{2}{\sqrt{2\pi}}\mu \int_0^\infty exp(\frac{-u}{\sqrt{2u}})du \quad (19)$$

Sorry but I can't continue to solve this problem now.

1.3 Proof that multivariate Gaussian distribution is normalised.

We have the multivariate Gaussian distribution:

$$p(x|\mu, \sigma^2) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} exp\left(\frac{-1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right) \quad (20)$$

With: μ is a D-dimensional mean vector
Σ is a D X D covariance matrix
$|\Sigma|$ is the determinant of Σ
We have:
+) Σ symmetric so eigenvalues (λ_1, λ_2,.., λ_n) are real and eigenvector (u_1, u_2, .., u_n) are orthornomal
+) Eigendecomposition decomposes a matrix

$$\Sigma = PDP^{-1} = PDP^T \quad (21)$$

With P is a matrix of eigenvectors. The form of P is:

$$P = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & ... & u_n \\ | & | & & | \end{bmatrix} \quad (22)$$

D is a diagonal matrix of eigenvalues. The form of D is:

$$D = \begin{bmatrix} \lambda_1 & 0 & .. & 0 \\ 0 & \lambda_2 & ... & 0 \\ 0 & 0 & .. & \lambda_n \end{bmatrix} \quad (23)$$

$$\Rightarrow \Sigma = \sum_{i=1}^{D} u_i \lambda_i u_i^T \quad (24)$$

$$\Rightarrow \Sigma^{-1} = \sum_{i=1}^{D} u_i \frac{1}{\lambda_i} u_i^T \quad (25)$$

So that:

$$\Delta^2 = (x-\mu)^T\Sigma^{-1}(x-\mu) = \sum_{i=1}^{D}\frac{1}{\lambda_i}(x-\mu)^T u_i u_i^T(x-\mu) = \sum_{i=1}^{D}\frac{y_i^2}{\lambda_i} \tag{26}$$

with $y_i = u_i^T(x-\mu)$ After that, we have:

$$|\Sigma| = |P||D||P^{-1}| = |D| => |\Sigma|^{1/2} = \prod_{j=1}^{D}\lambda_i^{1/2} \tag{27}$$

Finally, we have:

$$p(y) = \prod_{j=1}^{D}(\frac{1}{2\pi\lambda_j})^{1/2}e^{\frac{-y_i^2}{2\lambda_i}} \tag{28}$$

$$=> \int_{-\infty}^{\infty} p(y)dy = \prod_{j=1}^{D}\int_{-\infty}^{\infty}(\frac{1}{2\pi\lambda_j})^{1/2}e^{\frac{-y_i^2}{2\lambda_i}}dy_j = 1 \tag{29}$$

2 Question 2

2.1 Calculate conditional normal distribution

Set:

$$\Delta^2 = \frac{-1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) = \frac{-1}{2}x^T\Sigma^{-1}x + \frac{1}{2}x^T\Sigma^{-1}\mu + \frac{1}{2}\mu^T\Sigma^{-1}x - \frac{1}{2}\mu^T\Sigma^{-1}\mu \tag{30}$$

$= x^T\Sigma^{-1}\mu - \frac{1}{2}x^T\Sigma^{-1}x + const$
Suppose x is a D-dimensional vector with Gaussian distribution p(x $|\mu$, σ^2) and that we partition x into two disjoint subsets x_a and x_b: x = $\begin{pmatrix} a \\ b \end{pmatrix}$
We also define corresponding partitions of the mean vector μ given by:
$\mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix}$ and of the covariance matrix Σ given by:

$$\Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix} => A = \Sigma^{-1} = \begin{pmatrix} A_{aa} & A_{ab} \\ A_{ba} & A_{bb} \end{pmatrix} \tag{31}$$

We have μ is symmetric so Σ_{aa} and Σ_{bb} are symmetric while $\Sigma_{ab} = \Sigma_{ba}^T$
We have:

$$\Delta^2 = \frac{-1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) = \frac{-1}{2}(x-\mu)^T A(x-\mu) \tag{32}$$

$$\Delta^2 = \frac{-1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) = \frac{-1}{2}(x-\mu)^T A(x-\mu) \tag{33}$$

$$= \frac{-1}{2}\begin{pmatrix} x_a-\mu_a \\ x_b-\mu_b \end{pmatrix}^T \begin{pmatrix} A_{aa} & A_{ab} \\ A_{ba} & A_{bb} \end{pmatrix}\begin{pmatrix} x_a-\mu_a \\ x_b-\mu_b \end{pmatrix} \tag{34}$$

$$= \frac{-1}{2}(x_a - \mu_a)^T A_{aa}(x_a - \mu_a) - \frac{1}{2}(x_a - \mu_a)^T A_{ab}(x_b - \mu_b) - \frac{1}{2}(x_b - \mu_b)^T A_{ba}(x_a - \mu_a) - \frac{1}{2}(x_b - \mu_b)^T A_{bb}(x_b - \mu_b) \tag{35}$$

$$= \frac{-1}{2} x_a^T A_{aa} x_a + x_a^T (A_{aa}\mu_a - A_{ab}(x_b - \mu_b)) + const \tag{36}$$

From (30) and (36), we have:

$$\Sigma_{a|b} = A_{a_a}^{-1} and \mu_{a|b} = \Sigma_{a|b}(A_{aa}\mu_a - A_{ab}(x_b - \mu_b)) = \mu_a - A_{a_a}^{-1} A_{ab}(x_b - \mu_b) \tag{37}$$

By using Schur complement, we have: $\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} M & -MBD^{-1} \\ -D^{-1}CMD^{-1} & D^{-1}CMBD^{-1} \end{pmatrix}$

With $M = (A - BD^{-1}C)^{-1}$

$$A_{aa} = (\Sigma_{aa} - \Sigma_{ab}\Sigma_{b_b}^{-1}\Sigma_{ba})^{-1} \tag{38}$$

$$A_{ab} = -(\Sigma_{aa} - \Sigma_{ab}\Sigma_{b_b}^{-1}\Sigma_{ba})^{-1}\Sigma_{ab}\Sigma_{b_b}^{-1} \tag{39}$$

As a result,

$$\mu_{a|b} = \mu_a \Sigma_{ab}\Sigma_{b_b}^{-1}(x_b - \mu_b) \tag{40}$$

$$\Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{b_b}^{-1}\Sigma_{ba} \tag{41}$$

2.2 Calculate marginal normal distribution

The marginal normal distribution given by:

$$p(x_a) = \int p(x_a, x_b) dx_b \tag{42}$$

We need to integrate out x_b by looking te quadratic form related to x_b:

$$\Delta^2 = \frac{-1}{2}(x - \mu)^T A (x - \mu) = \frac{-1}{2} x_b^T A_{bb} x_b + x_b^T m + const \tag{43}$$

(with m = $\mathrm{A}_{bb}\mu_b - A_{ba}(x_a - \mu_a)$)

$$= \frac{-1}{2}(x_b - A_{b_b}^{-1}m)^T A_{bb}(x_b - A_{b_b}^{-1}m) + \frac{1}{2} m^T A_{b_b}^{-1} m \tag{44}$$

We can itegrate over unnormalised Gaussian
$\int exp\left(\frac{-1}{2}(x_b - A_{b_b}^{-1}m)^T A_{bb}(x_b - A_{b_b}^{-1}m) \right) dx_b$

The remaining term
$\frac{-1}{2} x_a^T (A_{aa} - A_{ab}A_{b_b}^{-1}A_{ba}) x_a + x_a^T (A_{aa} - A_{ab}A_{b_b}^{-1}A_{ba})^{-1}\mu_a + const$
Similarly, we have:

$$E[x_a] = \mu_a \tag{45}$$

$$cov[x_a] = \Sigma_{aa} \tag{46}$$