National Economics University, Vietnam
Faculty of Mathematics Economics
Data Science in Economics and Business
Machine Learning 1

Homework Week 2: Gaussian Distribution

Student: Nguyễn Anh Tú - ID: 11207333

1 Problem 1. Proof that:

(a) Gaussian distribution is normalized

(b) Expectation of Gaussian distribution is μ (mean)

(c) Variance of Gaussian distribution is σ^2 (variance)

(d) Multivariate Gaussian distribution is normalized

Solution.

(a) Normalization of Univariate Gaussian distribution is given by:

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

$$\Longleftrightarrow \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) dx = 1$$

We will first prove the base case when the mean equals to zero ($\mu = 0$), which means that:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2\sigma^2}x^2\right) dx = 1$$

$$\Longleftrightarrow \int_{-\infty}^{\infty} \exp\left(\frac{-1}{2\sigma^2}x^2\right) dx = \sqrt{2\pi\sigma^2}$$

Let:

$$I = \int_{-\infty}^{\infty} \exp\left(\frac{-1}{2\sigma^2}x^2\right) dx$$

Then we will take the square of both side:

$$I^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(\frac{-1}{2\sigma^2}x^2\right) \exp\left(\frac{-1}{2\sigma^2}y^2\right) dxdy$$

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

To conduct the integration, we will make the transformation from Cartesian coordinates (x, y) to **polar coordinates** (r, θ) by assuming:

$$x = r\cos\theta$$
$$y = r\sin\theta$$

where r and θ are arbitrary number and angle. By the trigonometric identity we also have:

$$\begin{aligned}\cos^2\theta + \sin^2\theta &= 1\\ x^2 + y^2 &= r^2\end{aligned}$$

While transforming integrals between two coordinate systems, we also note that the Jacobian the change of variables is given by:

$$\begin{aligned}dxdy &= |J|drd\theta\\ &= \begin{vmatrix}\frac{\partial(x)}{\partial(r)} & \frac{\partial(x)}{\partial(\theta)}\\ \frac{\partial(y)}{\partial(r)} & \frac{\partial(y)}{\partial(\theta)}\end{vmatrix}\\ &= \begin{vmatrix}\cos\theta & -r\sin\theta\\ \sin\theta & r\cos\theta\end{vmatrix}\\ &= r\cos^2\theta + r\sin^2\theta\\ &= r\\ &\implies dxdy = rdrd\theta\end{aligned}$$

Substituting the above results to the expression of I then:

$$\begin{aligned}I^2 &= \int_0^{2\pi}\int_0^{\infty} \exp\left(-\frac{r^2}{2\sigma^2}\right) rdrd\theta\\ &= 2\pi\int_0^{\infty}\exp\left(-\frac{r^2}{2\sigma^2}\right) rdr\\ &= 2\pi\pi\int_0^{\infty}\exp\exp\left(-\frac{r^2}{2\sigma^2}\right)\frac{d(r^2)}{2}\\ &= \pi\left[\exp\left(-\frac{r^2}{2\sigma^2}\right)(-2\sigma^2)\right]_0^{\infty}\\ &= 2\pi\sigma^2\end{aligned}$$

Now we have $I = \sqrt{2\pi\sigma^2}$, to prove the case when mean is non zero, we suppose $t = x - \mu$ so that:

$$\begin{aligned}\int_{-\infty}^{\infty} p(x|\mu,\sigma^2)dx &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\exp\left(-\frac{t^2}{2\sigma^2}\right)dt\\ &= \frac{I}{\sqrt{2\pi\sigma^2}}\\ &= \frac{\sqrt{2\pi\sigma^2}}{\sqrt{2\pi\sigma^2}} = 1 \quad \square\end{aligned}$$

(b) The formula of expected value of continuous random variable is given by:

$$E(X) = \int_{-\infty}^{\infty} xf_X(x)dx$$

Then the expectation of Univariate Gaussian Distribution is:

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \end{aligned}$$

To simplify the equation, let $t = \frac{x-\mu}{\sqrt{2\sigma^2}}$. Then $dt = \frac{dx}{\sqrt{2\sigma^2}}$ and $x = t\sqrt{2\sigma^2} + \mu$.

Substituting t in $E(X)$:

$$\begin{aligned} E(X) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (t\sqrt{2\sigma^2} + \mu) \exp(-t^2)\sqrt{2\sigma^2} dt \\ &= \frac{\sqrt{2\sigma^2}}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (t\sqrt{2\sigma^2} + \mu) \exp(-t^2) dt \\ &= \frac{1}{\sqrt{\pi}} \left(\sqrt{2\sigma^2} \int_{-\infty}^{\infty} t \exp(-t^2) dt + \mu \int_{-\infty}^{\infty} \exp(-t^2) dt \right) \end{aligned}$$

Let $A = \sqrt{2\sigma^2} \int_{-\infty}^{\infty} t \exp(-t^2) dt$ and $B = \mu \int_{-\infty}^{\infty} \exp(-t^2) dt$. Then:

$$\begin{aligned} A &= \sqrt{2\sigma^2} \int_{-\infty}^{\infty} \exp(-t^2) t d(t) \\ &= \sqrt{2\sigma^2} \int_{-\infty}^{\infty} -\frac{1}{2} \exp(-t^2) d(-t^2) \\ &= \sqrt{2\sigma^2} \left[-\frac{1}{2} \exp(-t^2) \right]_{-\infty}^{\infty} \\ &= 0 \ (1) \end{aligned}$$

$$\begin{aligned} B^2 &= \mu^2 \int_{-\infty}^{\infty} \exp(-t^2) dt \int_{-\infty}^{\infty} \exp(-u^2) du \\ &= \mu^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(-t^2 - u^2) du dt \end{aligned}$$

At this step, we will calculate B using polar transformation just like how the expression of I was calculated in part **(a)** of this problem. The final result is $B = \mu\sqrt{\pi}$ (2)

From (1) and (2) we have:

$$E(X) = \frac{\mu\sqrt{\pi}}{\sqrt{pi}} = \mu \quad \square$$

(c) By definition, the formula of the variance in Gaussian distribution is given by:

$$\begin{aligned}
V(X) &= E[(X - E(X))^2] \\
&= \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x) dx \\
&= \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\
&= \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \\
&= \int_{-\infty}^{\infty} (\sqrt{2\sigma^2}x)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\sqrt{2\sigma^2}x)^2}{2\sigma^2}\right) d(\sqrt{2\sigma^2}x) \\
&= \sigma^2 \frac{2}{\sqrt{\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2} dx \\
&= \sigma^2 \frac{4}{\sqrt{\pi}} \int_{0}^{\infty} x^2 e^{-x^2} dx
\end{aligned}$$

Let $t = x^2 \Rightarrow x = \sqrt{t}$ and $dt = 2xdx = 2\sqrt{t}dx = (2\sqrt{t})^{-1}dt$. Substituting to $V(X)$:

$$V(X) = \sigma^2 \frac{4}{\sqrt{\pi}} \int_0^{\infty} te^{-t}(2\sqrt{t})^{-1}dt = \sigma^2 \frac{2}{\sqrt{\pi}} \int_0^{\infty} t^{1/2}e^{-t}dt$$

Note that $\int_0^{\infty} t^{1/2}e^{-t}dt = \int_0^{\infty} t^{\frac{3}{2}-1}e^{-t}dt = \Gamma(\frac{3}{2}) = \frac{\sqrt{\pi}}{2}$ (Gamma function)

The result of $\Gamma(\frac{3}{2})$ can be calculated manually using Legendre duplication formula, a property of Gamma function that allows the calculation of half-integer entry. Substituting to $V(X)$:

$$V(X) = \sigma^2 \frac{2}{\sqrt{\pi}} \frac{\sqrt{\pi}}{2} = \sigma^2 \quad \square$$

(d) Multivariate Gaussian distribution is normalized

We have

$$\begin{aligned}
\Delta^2 &= (x-\mu)^T \Sigma^{-1} (x-\mu) \\
&= \sum_{i=1}^{D} \frac{1}{\lambda_i} (x-\mu)^T (x-\mu) \\
&= \sum_{i=1}^{D} \frac{y_i^2}{\lambda_i}
\end{aligned}$$

with $y_i = u_i^T(x-\mu)$ We also have $|\Sigma|^{\frac{1}{2}} = \prod_{i=1}^{D} \lambda_i^{\frac{1}{2}}$. For a D-dimensional vector x, the multivariate Gaussian distribution takes the form

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

We replace $y_i = u_i^T(x-\mu)$ into the equation, we have

$$\begin{aligned} p(y) &= \frac{1}{(2\pi)^{\frac{D}{2}}\left(\prod_{i=1}^{D}\lambda_i\right)^{\frac{1}{2}}}\exp\left(-\frac{1}{2}\sum_{i=1}^{D}\frac{y_i^2}{\lambda_i}\right) \\ &= \frac{1}{(2\pi)^{\frac{D}{2}}\left(\prod_{i=1}^{D}\lambda_i\right)^{\frac{1}{2}}}\prod_{i=1}^{D}\exp\left(-\frac{1}{2}\frac{y_i^2}{\lambda_i}\right) \\ &= \prod_{j=1}^{D}\frac{1}{(2\pi\lambda_i)^{\frac{1}{2}}}\exp\left(-\frac{y_j^2}{2\lambda_j}\right) \\ \implies \int_{-\infty}^{\infty} p(y)dy &= \prod_{j=1}^{D}\int_{-\infty}^{\infty}\frac{1}{(2\pi\lambda_i)^{\frac{1}{2}}}\exp\left(-\frac{y_j^2}{2\lambda_j}\right)dy_j \\ &= 1 \end{aligned}$$

2 Problem 2. Calculate:

(a) The marginal of Gaussian distribution

(b) The conditional of Gaussian distribution

Solution. Before jumping in the solution for each part, there will be some useful results that would help later. Given $x \in \mathbb{R}^n$:

$$\int p(x;\mu;\Sigma)dx = \int_{-\infty}^{\infty}...\int_{-\infty}^{\infty}p(x;\mu;\Sigma)dx_1...dx_n = 1 \tag{1}$$

$$\int x_i p(x;\mu;\sigma^2)dx = \mu_i \tag{2}$$

$$\int (x_i-\mu_i)(x_j-\mu_j)p(x;\mu,\sigma^2)dx = \Sigma_{ij} \tag{3}$$

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} M^{-1} & -M^{-1}BD^{-1} \\ -D^{-1}CM^{-1} & D^{-1}+D^{-1}CM^{-1}BD^{-1} \end{bmatrix} \text{ where } M = A - BD^{-1}C \tag{4}$$

(Schur complement)

(a) Suppose that

$$\begin{bmatrix} x_A \\ x_B \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_A \\ \mu_B \end{bmatrix}, \begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{bmatrix}\right)$$

where $x_A \in \mathbb{R}^m$, $x_B \in \mathbb{R}^n$. The marginal PDF for x_A (calculation of marginal distribution for x_B is similar) is:

$$\begin{aligned} p(x_A) &= \int p(x_A, x_B;\mu,\Sigma)dx_B \\ &= \frac{1}{(2\pi)^{\frac{m+n}{2}}|\Sigma|^{1/2}}\int\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)dx_B \\ &= \frac{1}{(2\pi)^{\frac{m+n}{2}}\begin{vmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{vmatrix}^{1/2}}\int\exp\left(-\frac{1}{2}\begin{bmatrix} x_A-\mu_A \\ x_B-\mu_B \end{bmatrix}^T\begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{bmatrix}^{-1}\begin{bmatrix} x_A-\mu_A \\ x_B-\mu_B \end{bmatrix}\right) \end{aligned}$$

Denote:

$$V = \begin{bmatrix} V_{AA} & V_{AB} \\ V_{BA} & V_{BB} \end{bmatrix} = \Sigma^{-1}$$

Let $Z = \frac{1}{(2\pi)^{\frac{m+n}{2}} |\Sigma|^{1/2}}$ be a constant that its value does not depend on x_A. Substituting to the marginal distribution expression:

$$\begin{aligned} p(x_A) = \frac{1}{Z} \int \exp \left(-\frac{1}{2}(x_A - \mu_A)^T V_{AA}(x_A - \mu_A) - \frac{1}{2}(x_A - \mu_A)^T V_{AB}(x_B - \mu_B) \right) dx_B \\ \cdot \int \exp \left(-\frac{1}{2}(x_B - \mu_B)^T V_{BA}(x_A - \mu_A) - \frac{1}{2}(x_B - \mu_B)^T V_{BB}(x_B - \mu_B) \right) dx_B \end{aligned}$$

Here we will apply a mathematical trick known as "completion of squares" to transform $p(X_A)$ to an expression including quadratic form. Consider the quadratic function $z^T Az + b^T z + c$ where A is a symmetric, non singular matrix. Then:

$$\frac{1}{2} z^T Az + b^T z + c = \frac{1}{2}(z + A^{-1}b)^T A(z + A^{-1}b) + c - \frac{1}{2} b^T A^{-1} b$$

In our case:

$$\begin{aligned} z &= x_B - \mu_B \\ A &= V_{BB} \\ b &= V_{BA}(x_A - \mu_A) \\ c &= \frac{1}{2}(x_A - \mu_A)^T V_{AA}(x_A - \mu_A) \end{aligned}$$

Then:

$$\begin{aligned} p(x_A) &= \frac{1}{Z} \exp \left(-\frac{1}{2}(x_A - \mu_A)^T V_{AA}(x_A - \mu_A) + \frac{1}{2}(x_A - \mu_A)^T V_{AB} V_{BB}^{-1} V_{BA}(x_A - \mu_A) \right) \\ &\cdot \int \exp \left[-\frac{1}{2} \left(x_B - \mu_B + V_{BB}^{-1} V_{BA}(x_A - \mu_A) \right)^T V_{BB} \left(x_B - \mu_B + V_{BB}^{-1} V_{BA}(x_A - \mu_A) \right) \right] dx_B \\ &= \frac{1}{Z} \exp \left(-\frac{1}{2}(x_A - \mu_A)^T (V_{AA} - V_{AB} V_{BB}^{-1} V_{BA})(x_A - \mu_A) \right) \\ &\cdot \int \exp \left[-\frac{1}{2} \left(x_B - \mu_B \right)^T V_{BB} \left(x_B - \mu_B \right) \right] dx_B \end{aligned}$$

Recall the result **(1)** in the useful result section I have listed at the beginning of this problem:

$$p(x) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \int \exp \left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu) \right) dx = 1$$

Then:

$$\frac{1}{(2\pi)^{n/2} |V_{BB}|^{1/2}} \int \exp \left[-\frac{1}{2} \left(x_B - \mu_B \right)^T V_{BB} \left(x_B - \mu_B \right) \right] dx_B = 1$$

Now we have eliminated the integral part in the expression of $p(x_A)$, the remainder will be:

$$p(x_A) = \frac{1}{Z} (2\pi)^{n/2} |V_{BB}|^{1/2} \exp \left(-\frac{1}{2}(x_A - \mu_A)^T (V_{AA} - V_{AB} V_{BB}^{-1} V_{BA})(x_A - \mu_A) \right)$$

Applying Schur complement knowing that:

$$\begin{aligned}\begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{bmatrix}^{-1} &= \begin{bmatrix} V_{AA} & V_{AB} \\ V_{BA} & V_{BB} \end{bmatrix} \\ &= \begin{bmatrix} (V_{AA} - V_{AB}V_{BB}^{-1}V_{BA})^{-1} & V_{AB}V_{BB}^{-1}V_{BA})^{-1}V_{AB}V_{BB}^{-1} \\ -V_{BB}V_{BA}(V_{AA} - V_{AB}V_{BB}^{-1}V_{BA})^{-1} & (V_{BB} - VBAV_{AA}^{-1}V_{AB})^{-1} \end{bmatrix}\end{aligned}$$

Then we can see that $(V_{AA} - V_{AB}VBB^{-1}V_{BA})^{-1} = \Sigma_{AA}$. The formula for marginal distribution $p(x_A)$ will be:

$$\mathbf{p(x_A) = \frac{1}{Z}(2\pi)^{n/2}|V_{BB}|^{1/2}} \exp\left(-\mathbf{\frac{1}{2}(x_A - \mu_A)^T\Sigma_{AA}(x_A - \mu_A)}\right)$$

(b) Suppose that

$$\begin{bmatrix} x_A \\ x_B \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_A \\ \mu_B \end{bmatrix}, \begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{bmatrix}\right)$$

where $x_A \in \mathbb{R}^m$, $x_B \in \mathbb{R}^n$. The conditional PDF $p(x_A|x_B)$ is:

$$\begin{aligned}p(x_A|x_B) &= \frac{p(x_A, x_B; \mu, \sigma)}{p(x_A)} \\ &= \frac{p(x_A, x_B; \mu, \sigma)}{\int p(x_A, x_B; \mu, \sigma)dx_A}\end{aligned}$$

Note that the integral in the denominator does not depend on x_A because it is the marginal distribution over x_B. Then, to simplify, we will denote M including all the factors that does not depend on x_A and the denominator as well. The expression for $p(x_A|x_B)$ will be:

$$\begin{aligned}p(x_A|x_B) &= \frac{1}{M}\exp(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)) \\ &= \frac{1}{M}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)dx_B \\ &= \exp\left(-\frac{1}{2}\begin{bmatrix} x_A - \mu_A \\ x_B - \mu_B \end{bmatrix}^T \begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{bmatrix}^{-1} \begin{bmatrix} x_A - \mu_A \\ x_B - \mu_B \end{bmatrix}\right) \\ &= \frac{1}{M}\exp\left(-\frac{1}{2}(x_A - \mu_A)^T V_{AA}(x_A - \mu_A) - \frac{1}{2}(x_A - \mu_A)^T V_{AB}(x_B - \mu_B)\right) \\ &\cdot \exp\left(-\frac{1}{2}(x_B - \mu_B)^T V_{BA}(x_A - \mu_A) - \frac{1}{2}(x_B - \mu_B)^T V_{BB}(x_B - \mu_B)\right)\end{aligned}$$

Applying complement square again, in this case:

$$\begin{aligned}z &= x_A - \mu_A \\ A &= V_{AA} \\ b &= V_{AB}(x_B - \mu_B) \\ c &= \frac{1}{2}(x_B - \mu_B)^T V_{BB}(x_B - \mu_B)\end{aligned}$$

Then the new expression for $p(x_A|x_B)$ will be:

$$p(x_A|x_B) = \frac{1}{M}\exp\left[-\frac{1}{2}(x_A - \mu_A + V_{AA}^{-1}V_{AB}(x_A - \mu_A))^T V_{AA}(x_A - \mu_A + V_{AA}^{-1}V_{AB}(x_B - \mu_B))\right]$$
$$\cdot \exp\left[-\frac{1}{2}(x_B - \mu_B)^T V_{BB}(x_B - \mu_B) + \frac{1}{2}(x_B - \mu_B)^T V_{BA}V_{AA}^{-1}V_{AB}(x_B - \mu_B)\right] \ (*)$$

The second exp factor in (*) does not depend on x_A so we can include it and M in a normalization constant M'. Then:

$$p(x_A|x_B) = \frac{1}{M'}\exp\left[-\frac{1}{2}(x_A - \mu_A + V_{AA}^{-1}V_{AB}(x_A - \mu_A))^T V_{AA}(x_A - \mu_A + V_{AA}^{-1}V_{AB}(x_B - \mu_B))\right]$$

Applying Schur complement knowing that:

$$\begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{bmatrix}^{-1} = \begin{bmatrix} V_{AA} & V_{AB} \\ V_{BA} & V_{BB} \end{bmatrix}$$
$$= \begin{bmatrix} (V_{AA} - V_{AB}V_{BB}^{-1}V_{BA})^{-1} & V_{AB}V_{BB}^{-1}V_{BA})^{-1}V_{AB}V_{BB}^{-1} \\ -V_{BB}V_{BA}(V_{AA} - V_{AB}V_{BB}^{-1}V_{BA})^{-1} & (V_{BB} - VBAV_{AA}^{-1}V_{AB})^{-1} \end{bmatrix}$$

Then we have:

$$\mu_{A|B} = \mu_A - V_{AA}^{-1}V_{AB}(x_B - \mu_B) = \mu_A + \Sigma_{AB}\Sigma_{BB}(x_B - \mu_B)$$

Conversely, we can also derive $V_{AA}^{-1} = \Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA})^{-1}$ then

$$\Sigma_{A|B} = V_{AA}^{-1} = \Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA})^{-1}$$

$$\Rightarrow p(x_A|x_B) \sim \mathbb{N}(\mu_{A|B}; \Sigma_{A|B})$$

with mean and variance are calculated above.