

MACHINE LEARNING

Nghiem Thi Ngoc Minh
MSV 11202545

HW week2

1 Problem 1

Proff that:

a) Gaussian distribution is normalized

Normalization of Univariate Gaussian distribution is given by:

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

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

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

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:

$$\begin{aligned} I^2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(\frac{-1}{2\sigma^2} x^2\right) \exp\left(\frac{-1}{2\sigma^2} y^2\right) dx dy \\ \Leftrightarrow I^2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) dx dy \end{aligned}$$

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

$$\begin{aligned} x &= r \cos \theta \\ y &= r \sin \theta \end{aligned}$$

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:

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

Substituting the above results to the expression of I then:

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

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:

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

b) Expectation of Gaussian distribution is: μ

We have probability density function:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The expect value:

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

$$\text{Let } t = \frac{x-\mu}{\sqrt{2}\sigma}$$

$$\implies dt = \frac{1}{\sqrt{2}\sigma} dx$$

$$\implies \sqrt{2}\sigma dt = dx$$

So:

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

c) Variance of Gaussian distribution is σ^2

We have the probability density function of Gaussian distribution is:

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

From the Variance as Expectation of Square minus Square of expectation:

$$var(X) = \int_{-\infty}^{\infty} x^2 f_X(x) dx - (E(X))^2$$

So:

$$\begin{aligned}
\text{var}(X) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx - \mu^2 \\
\text{Let } t &= \frac{x-\mu}{\sqrt{2}\sigma} : \\
&= \frac{\sqrt{2}\sigma}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sqrt{2}\sigma t + \mu)^2 \exp(-t^2) dt - \mu^2 \\
&= \frac{1}{\sqrt{\pi}} \left(2\sigma^2 \int_{-\infty}^{\infty} t^2 \exp(-t^2) dt + 2\sqrt{2}\sigma\mu \int_{-\infty}^{\infty} t \exp(-t^2) dt + \mu^2 \int_{-\infty}^{\infty} \exp(-t^2) dt \right) - \mu^2 \\
&= \frac{1}{\sqrt{\pi}} \left(2\sigma^2 \int_{-\infty}^{\infty} t^2 \exp(-t^2) dt + 2\sqrt{2}\sigma\mu \left[-\frac{1}{2} \exp(-t^2)\right] + \mu^2 \sqrt{\pi} \right) - \mu^2 \\
&= \frac{1}{\sqrt{\pi}} \left(2\sigma^2 \int_{-\infty}^{\infty} t^2 \exp(-t^2) dt + 2\sqrt{2}\sigma\mu \cdot 0 \right) + \mu^2 - \mu^2 \\
&= \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} t^2 \exp(-t^2) dt \\
&= \frac{2\sigma^2}{\sqrt{\pi}} \left(\left[-\frac{t}{2} \exp(-t^2)\right] + \frac{1}{2} \int_{-\infty}^{\infty} \exp(-t^2) dt \right) \\
&= \frac{2\sigma^2}{\sqrt{\pi}} \cdot \frac{1}{2} \int_{-\infty}^{\infty} \exp(-t^2) dt \\
&= \frac{2\sigma^2 \sqrt{\pi}}{2\sqrt{\pi}} \\
&= \sigma^2
\end{aligned}$$

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

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

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

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

2 Problem 2

Let $Y \sim N(\mu; \Sigma)$. Consider the partitioning μ and Σ into:

$$\begin{aligned}
\mu &= \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \\
\Sigma &= \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}
\end{aligned}$$

With a similar partition of Σ into:

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

Then $(y_1 | y_2 = a)$, the conditional distribution of first partition given the second is $N(\bar{\mu}, \bar{\Sigma})$ with mean:

$$\bar{\mu} = \mu_1 + \sum_{12}^{-1} \sum_{22} (a - \mu_2)$$

and covariance matrix:

$$\bar{\Sigma} = \Sigma_{11} - \sum_{12}^{-1} \sum_{22} \sum_{21}$$

Use the blockwise inversion formula to write the inverse-variance matrix as:

$$\sum^{-1} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \Sigma_{11}^* & \Sigma_{12}^* \\ \Sigma_{21}^* & \Sigma_{22}^* \end{bmatrix}$$

Where:

$$\Sigma_{11}^* = \Sigma_{11}^{-1}$$

$$\begin{aligned}
\sum_{12}^* &= -\sum_{*}^{-1} \sum_{12} \sum_{22}^{-1} \\
\sum_{21}^* &= -\sum_{22}^{-1} \sum_{21} \sum_{*}^{-1} \\
\sum_{22}^* &= \sum_{22}^{-1} + \sum_{22}^{-1} \sum_{21} \sum_{*}^{-1} \sum_{12} \sum_{22}^{-1}
\end{aligned}$$

We have:

$$\begin{aligned}
(y - \mu)^T \sum_{*}^{-1} (y - \mu) &= \begin{bmatrix} y_1 - \mu_1 \\ y_2 - \mu_2 \end{bmatrix}^T \begin{bmatrix} \sum_{11}^* & \sum_{12}^* \\ \sum_{21}^* & \sum_{22}^* \end{bmatrix} \begin{bmatrix} y_1 - \mu_1 \\ y_2 - \mu_2 \end{bmatrix} \\
&= (y_1 - \mu_1)^T \sum_{11}^* (y_1 - \mu_1) + (y_1 - \mu_1)^T \sum_{12}^* (y_2 - \mu_2) + (y_2 - \mu_2)^T \sum_{21}^* (y_1 - \mu_1) + (y_2 - \mu_2)^T \sum_{22}^* (y_2 - \mu_2) \\
&= (y_1 - \mu_1)^T \sum_{*}^{-1} (y_1 - \mu_1) - (y_1 - \mu_1)^T \sum_{*}^{-1} \sum_{12} \sum_{22}^{-1} (y_2 - \mu_2) - (y_2 - \mu_2)^T \sum_{22}^{-1} \sum_{21} \sum_{*}^{-1} \sum_{12} \sum_{22}^{-1} (y_2 - \mu_2) \\
&= (y_1 - (\mu_1 + \sum_{12} \sum_{22}^{-1} (y_2 - \mu_2)))^T \sum_{*}^{-1} (y_1 - (\mu_1 + \sum_{12} \sum_{22}^{-1} (y_2 - \mu_2))) + (y_2 - \mu_2)^T \sum_{22}^{-1} (y_2 - \mu_2) \\
&= (y_1 - \mu_*)^T \sum_{*}^{-1} (y_1 - \mu_*) + (y_2 - \mu_2)^T \sum_{22}^{-1} (y_2 - \mu_2)
\end{aligned}$$

Where:

Conditional part: $(y_1 - \mu_*)^T \sum_{*}^{-1} (y_1 - \mu_*)$

Marginal part: $(y_2 - \mu_2)^T \sum_{22}^{-1} (y_2 - \mu_2)$