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:

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

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

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

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

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

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:

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

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

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:

$$\cos^2\theta + \sin^2\theta = 1$$
$$x^2 + y^2 = r^2$$

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

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

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:

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

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:

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} xf(x)\mathrm{d}x \\
&= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} xe^{-\frac{(x-\mu)^2}{2\sigma^2}}\mathrm{d}x
\end{aligned}$$

Let t = $\frac{x-\mu}{\sqrt{2}\sigma}$
=> $\mathrm{d}t = \frac{1}{\sqrt{2}\sigma}\mathrm{d}x$
=> $\sqrt{2}\sigma\mathrm{d}t = \mathrm{d}x$

So:

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

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 Varience as Expectation of Square minus Square of expactation:

$$var(X) = \int\limits_{-\infty}^{\infty} x^2 f_X(x) \mathrm{d}x - (E(X))^2$$

So:

$$
\begin{aligned}
var(X) &= \frac{1}{\sigma\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} x^2 \exp{(-\frac{(x-\mu)^2}{2\sigma^2})}\mathrm{d}x - \mu^2 \\
Let t = \frac{x-\mu}{\sqrt{2}\sigma} : & \\
&= \frac{\sqrt{2}\sigma}{\sigma\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} (\sqrt{2}\sigma t+\mu)^2 \exp{(-t^2)}\mathrm{d}t - \mu^2 \\
&= \frac{1}{\sqrt{\pi}}(2\sigma^2\int\limits_{-\infty}^{\infty} t^2\exp{(-t^2)}\mathrm{d}t + 2\sqrt{2}\sigma\mu\int\limits_{-\infty}^{\infty} t\exp{(-t^2)}\mathrm{d}t + \mu^2\int\limits_{-\infty}^{\infty}\exp{(-t^2)}\mathrm{d}t) - \mu^2 \\
&= \frac{1}{\sqrt{\pi}}(2\sigma^2\int\limits_{-\infty}^{\infty} t^2\exp{(-t^2)}\mathrm{d}t + 2\sqrt{2}\sigma\mu[-\frac{1}{2}\exp{(-t^2)}] + \mu^2\sqrt{\pi}) - \mu^2 \\
&= \frac{1}{\sqrt{\pi}}(2\sigma^2\int\limits_{-\infty}^{\infty} t^2\exp{(-t^2)}\mathrm{d}t + 2\sqrt{2}\sigma\mu\cdot 0) + \mu^2 - \mu^2 \\
&= \frac{2\sigma^2}{\sqrt{\pi}}\int\limits_{-\infty}^{\infty} t^2\exp{(-t^2)}\mathrm{d}t \\
&= \frac{2\sigma^2}{\sqrt{\pi}}([-\frac{t}{2}\exp{-t^2}] + \frac{1}{2}\int\limits_{-\infty}^{\infty}\exp{(-t^2)}\mathrm{d}t) \\
&= \frac{2\sigma^2}{\sqrt{\pi}}\cdot\frac{1}{2}\int\limits_{-\infty}^{\infty}\exp{(-t^2)}\mathrm{d}t \\
&= \frac{2\sigma^2\sqrt{\pi}}{2\sqrt{\pi}} \\
&= \sigma^2
\end{aligned}
$$

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

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

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

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

Vit cho Long Nguyn

2 Problem 2

Let $Y\ N(\mu;\sum)$. COnsider the partitioning μ and Y into:

$$\mu = \begin{vmatrix}\mu_1\\ \mu_2\end{vmatrix}$$
$$\sum = \begin{vmatrix}y_1\\ y_2\end{vmatrix}$$

With a similar partition of $\sum$ into:

$$\begin{vmatrix}\sum\limits_{11} & \sum\limits_{12}\\ \sum\limits_{21} & \sum\limits_{22}\end{vmatrix}$$

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

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

and covarience matrix:

$$\overline{\sum} = \sum_{11} - \sum_{12}\sum_{22}^{-1}\sum_{21}$$

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

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

Where:
$\sum_{11}^{*} = \sum_{*}^{-1}$

$\sum_{12}^{*} = -\sum_{*}^{-1} \sum_{12} \sum_{22}^{-1}$
$\sum_{21}^{*} = -\sum_{22}^{-1} \sum_{21} \sum_{*}^{-1}$
$\sum_{22}^{*} = \sum_{22}^{-1} + \sum_{22}^{-1} \sum_{21} \sum_{*}^{-1} \sum_{12} \sum_{22}^{-1}$
We have:

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

Where:
Conditional part: $(y_1-\mu_*)^T \sum_{*}^{-1} (y_1-\mu_*)$
Marginal part: $(y_2-\mu_2)^T \sum_{22}^{-1} (y_2-\mu_2)$