HW2 - Gaussian distribution

Phương Quỳnh Chi

August 24, 2022

Exercise 1

Proof that:

a) Gaussian distribution is normalized

$p(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}.e^{\frac{-(x-\mu)^2}{2\sigma^2}}$

To proof that above expression is normalized, we have to show that:

$$\int\limits_{-\infty}^{\infty} p(x|\mu,\sigma^2)dx = 1 \Leftrightarrow \int\limits_{-\infty}^{\infty} e^{\frac{-(x-\mu)^2}{2\sigma^2}} = \sqrt{2\pi\sigma^2} \qquad (1)$$

Assume that $\mu = 0 \Rightarrow \int\limits_{-\infty}^{\infty} e^{\frac{-x^2}{2\sigma^2}} \sqrt{2\pi\sigma^2}$

Let $I = \int\limits_{-\infty}^{\infty} e^{\frac{-x^2}{2\sigma^2}} dx$

$$\begin{aligned} I^2 &= \int\limits_{-\infty}^{\infty} e^{\frac{-x^2}{2\sigma^2}} dx. \int\limits_{-\infty}^{\infty} e^{\frac{-y^2}{2\sigma^2}} dy \\ &= \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} e^{\frac{-(x^2+y^2)}{2\sigma^2}} dxdy \end{aligned}$$

To evaluate the above expression, we make the transformation to polar coordinates (r,θ), that is $x = r.cos\theta, y = r.sin\theta$. And using the trigonometric identity $cos^2\theta + sin^2\theta = 1$. So $x^2 + y^2 = r^2$. Thus:

$$\begin{aligned} I^2 &= \int\limits_0^{2\pi} \int\limits_0^{\infty} e^{\frac{-r^2}{2\sigma^2}} rdrd\theta \\ &= 2\pi \int\limits_0^{\infty} e^{\frac{-r^2}{2\sigma^2}} rdr \\ &= 2\pi \int\limits_0^{\infty} e^{\frac{-u}{2\sigma^2}} \frac{1}{2}du \qquad (r^2 = u) \\ &= \pi[(e^{\frac{-u}{2\sigma^2}})(-2\sigma^2)]_0^{\infty} \\ &= 2\pi\sigma^2 \end{aligned}$$

$\Rightarrow I = \sqrt{2\pi\sigma^2}$satisfy (1) $\Rightarrow$ Proved

b) Expectation of Gaussian distribution is mean (μ)

$E(X) = xf(x)dx$ (expected value of continuous random variable)
So:

$$E(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} xe^{\frac{-(x-\mu^2)}{2\sigma^2}} dx \quad \text{substituting t} = \frac{x-\mu}{\sqrt{2}\sigma} \rightarrow x = \sqrt{2}\sigma t + \mu$$

$$= \frac{\sqrt{2}\sigma}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sqrt{2}\sigma t + \mu)e^{-t^2} dt$$

$$= \frac{1}{\pi}\left(\sqrt{2}\sigma \int_{-\infty}^{\infty} te^{-t^2} dt + \mu \int_{-\infty}^{\infty} e^{-t^2} dt\right)$$

$$= \frac{1}{\pi}\left(\sqrt{2}\sigma \left[\frac{-1}{2}e^{-t^2}\right]_{-\infty}^{\infty} + \mu\sqrt{\pi}\right)$$

$$= \frac{\mu\sqrt{\pi}}{\sqrt{\pi}}$$

$$= \mu \rightarrow \text{Proved}$$

c) Variance of Gaussian distribution is σ^2

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

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x^2 e^{\frac{-(x-\mu^2)}{2\sigma^2}} dx - \mu^2 \quad \text{substituting t} = \frac{x-\mu}{\sqrt{2}\sigma} \rightarrow x = \sqrt{2}\sigma t + \mu$$

$$= \frac{\sqrt{2}\sigma}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sqrt{2}\sigma t + \mu)^2 . e^{-t^2} dt - \mu^2$$

$$= \frac{1}{\pi} \int_{-\infty}^{\infty} (2\sigma^2 t^2 + 2\sqrt{2}\sigma\mu t + \mu^2).e^{-t^2} dt - \mu^2$$

$$= \frac{1}{\pi}\left(2\sigma^2 \int_{-\infty}^{\infty} t^2 e^{-t^2} dt + 2\sqrt{2}\sigma\mu \int_{-\infty}^{\infty} te^{-t^2} dt + \mu^2 \int_{-\infty}^{\infty} e^{-t^2} dt\right) - \mu^2$$

$$= \frac{1}{\pi}\left(2\sigma^2 \int_{-\infty}^{\infty} t^2 e^{-t^2} dt + 2\sqrt{2}\sigma\mu \left[\frac{-1}{2}e^{-t^2}\right]_{-\infty}^{\infty} + \mu^2\sqrt{\pi}\right) - \mu^2$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} t^2 e^{-t^2} dt + \frac{1}{\sqrt{\pi}} 2\sqrt{2}\sigma\mu.0 + \frac{1}{\sqrt{\pi}}\mu^2\sqrt{\pi} - \mu^2$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} t^2 e^{-t^2} dt$$

$$= \frac{2\sigma^2}{\sqrt{\pi}}\left(\left[\frac{-t}{2}e^{-t^2}\right]_{-\infty}^{\infty} + \frac{1}{2}\int\limits_{-\infty}^{\infty} e^{-t^2}dt\right)$$

$$= \frac{2\sigma^2}{\sqrt{\pi}}.\frac{1}{2}\int\limits_{-\infty}^{\infty} e^{-t^2}dt$$

$$= \frac{2\sigma^2\sqrt{\pi}}{2\sqrt{\pi}}$$

$= \sigma^2 \rightarrow$ Proved

d) Multivariate Gaussian distribution is normalized

The multivariate Gaussian distribution has the form

$p(x|\mu,\sigma^2) = \frac{1}{(2\pi)^{\frac{D}{2}}|\Sigma|^{\frac{1}{2}}}.e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$

Because Σ is a real, symmetric matrix, its eigenvalues will be real and its eigenvectors form an orthonormal set

$\Sigma = \sum\limits_{i=1}^{D}\lambda_i u_i u_i^T \Rightarrow \Sigma^{-1} = \sum\limits_{i=1}^{D}\frac{1}{\lambda_i}u_i u_i^T$

So that

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

$$= \sum_{i=1}^{D}\frac{y_i^2}{\lambda_i}, \text{with} y_i = u_i^T(x-\mu)$$

$|\Sigma|^{1/2} = \prod\limits_{j=1}^{D}\lambda_j^{1/2}$

$p(y) = \prod\limits_{j=1}^{D}\frac{1}{2\pi\lambda_j}^{1/2}.e^{-\frac{y_j^2}{2\lambda_j}}$

$\Rightarrow \int\limits_{-\infty}^{\infty} p(y)dy = \prod\limits_{j=1}^{D}\int\limits_{-\infty}^{\infty}\frac{1}{2\pi\lambda_j}^{1/2}.e^{-\frac{y_j^2}{2\lambda_j}}dy_j = 1 \rightarrow$ the multivariate Gaussian is normalized

Exercise 2

Calculate:

a) The conditional of Gaussian distribution

Suppose x is a D-dimensional vector with Gaussian distribution $\mathcal{N}(x|\mu,\Sigma)$

$\rightarrow x = \begin{pmatrix} x_a \\ x_b \end{pmatrix}$

Partitions of the mean vector: $\mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix}$

The covariance matrix: $\Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix} \rightarrow A = \Sigma^{-1} = \begin{pmatrix} A_{aa} & A_{ab} \\ A_{ba} & A_{bb} \end{pmatrix}$

We want to calculate the conditional distribution $p(x_a|x_b)$

We have:

$-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)$
$= -\frac{1}{2}(x-\mu)^T A(x-\mu)$
$= -\frac{1}{2}(x_a-\mu_a)^T A_{aa}(x_a-\mu_a) - \frac{1}{2}(x_a-\mu_a)^T A_{ab}(x_b-\mu_b) - \frac{1}{2}(x_b-\mu_b)^T A_{ba}(x_a-\mu_a) - \frac{1}{2}(x_b-\mu_b)^T A_{bb}(x_b-\mu_b)$
$= -\frac{1}{2}x_a^T A_{aa}^{-1}x_a + x_a^T(A_{aa}\mu_a - A_{ab}(x_b-\mu_b)) + const \qquad (2)$

It is quadratic form of x_a hence conditional distribution $p(x_a|x_b)$ will be Gaussian, because this distribution is characterized by its mean and its variance. Compare with Gaussian distribution $\Delta^2 = -\frac{1}{2}x^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu + const$
$\Sigma_{a|b} = A_{aa}^{-1}$
$\mu_{a|b} = \Sigma_{a|b}(A_{aa}\mu_a - A_{ab}(x_b-\mu_b)) = \mu_a - A_{aa}^{-1}A_{ab}(x_b-\mu_b)$

By using Schur complement,
$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} M & -MBD^{-1} \\ -D^{-1}CM & D^{-1}+CMBD^{-1} \end{pmatrix}$
We define $M = (A - BD^{-1}C)^{-1}$

$\Rightarrow A_{aa} = (\Sigma_{aa} - \Sigma_{ab}\Sigma_{ab}^{-1}\Sigma_{ba})^{-1}$
$A_{ab} = -(\Sigma_{ab}^{-1}\Sigma_{ba})^{-1}\Sigma_{ab}\Sigma_{bb}^{-1}$

As a result
$\mu_{a|b} = \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b-\mu_b)$
$\Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$
$\Rightarrow p(x_a|x_b) = \mathcal{N}(x_{a|b}|\mu_{a|b}, \Sigma_{a|b})$

b) The marginal of Gaussian distribution
The marginal distribution given by

$$p(x_a) = \int p(x_a, x_b)dx_b$$

Picking out those terms that involves x_b, we have
$-\frac{1}{2}x_b^T A_{bb}x_b + x_b^T m = -\frac{1}{2}(x_b - A_{bb}^{-1}m)^T A_{bb}(x_b - A_{bb}^{-1}m) + \frac{1}{2}m^T A_{bb}^{-1}m$
with $m = A_{bb}\mu_b - A_{ba}(x_a-\mu_a)$
The integration over x_b will take the form
$\int e^{-\frac{1}{2}(x_b - A_{bb}^{-1}m)^T A_{bb}(x_b - A_{aa}^{-1}m)}dx_b$
Combining this term with the term (2) from part a) that depend on x_a, we obtain:
$-\frac{1}{2}x_a^T(A_{aa} - A_{ab}A_{bb}^{-1}A_{ba})x_a + x_a^T(A_{aa} - A_{ab}A_{bb}^{-1}A_{ba})^{-1}\mu_a + const$
So, the covariance of the marginal distribution of $p(x_a)$ is given by

$$\Sigma_{aa} = (A_{aa} - A_{ab}A_{bb}^{-1}A_{ba})^{-1}$$

The mean is given by

$$\mu_a = \Sigma_a(A_{aa} - A_{ab}A_{bb}^{-1}A_{ba})\mu_a$$

As the result, the marginal distribution has the mean and covariance given by

$$
\begin{aligned}
E[x_a] &= \mu_a \\
cov[x_a] &= \Sigma_{aa} \\
\Rightarrow p(x_a) &= \mathcal{N}(x_a|\mu_a, \Sigma_{aa})
\end{aligned}
$$