Machine Learning

Tran Kha Uyen - 11207413

24th August, 2022

1 Prove that:

1.1 Gaussian distribution is normalized

The Gaussian Distribution is defined by:

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

Assume $\mu = 0$

$$I = \int_{-\infty}^{\infty} e^{\frac{-1}{2\sigma^2}\cdot x^2}dx$$

Take the square of both sides. When multiplying the 2 integrals, we need to change 1 of variables of intergration to y

$$I = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{\frac{-1}{2\sigma^2}\cdot x^2 - \frac{-1}{2\sigma^2}\cdot y^2}dxdy$$

Transform from Cartesian coordinates (x, y) to Polar coordinates (r, θ)

$$x = r \cdot cos\theta$$

$$y = r \cdot sin\theta$$

Given that $cos^2\theta + sin^2\theta = 1$

$$\begin{aligned} r \cdot (cos^2\theta + sin^2\theta) &= r \\ r \cdot cos^2\theta + r \cdot sin^2\theta &= r \\ r^2 \cdot cos^2\theta + r^2 \cdot sin^2\theta &= r^2 \\ \rightarrow x^2 + y^2 &= r^2 \end{aligned}$$

The Jacobian of the change of variables is given by: $\frac{\partial(x,y)}{\partial(r,\theta)} =$

$$\begin{vmatrix} \frac{\partial(x)}{\partial(r)} & \frac{\partial(x)}{\partial(\theta)} \\ \frac{\partial(y)}{\partial(r)} & \frac{\partial(y)}{\partial(\theta)} \end{vmatrix} = \begin{vmatrix} cos\theta & -rsin\theta \\ sin\theta & rcos\theta \end{vmatrix} = rcos^2\theta + rsin^2\theta = r$$

$$\rightarrow dxdy = r \cdot drd\theta$$

Hence,

$$I^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{\frac{-r^2}{2\sigma}} \cdot rdrd\theta = 2\pi\int_{-\infty}^{\infty} e^{\frac{-u}{2\sigma}} \cdot \frac{1}{2}du = 2\pi\sigma^2$$

where $r^2 = u$

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

Finally, we make the transformation $y = x - \mu$

$$\int_{-\infty}^{\infty} N(x|\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-y^2}{2\sigma^2}} dy = \frac{I}{\sigma\sqrt{2\pi}} = \frac{\sigma\sqrt{2\pi}}{\sigma\sqrt{2\pi}} = 1$$

$\rightarrow So, Gaussian distribution is normalized$

1.2 Expectation of Gaussian distribution is mu (mean)

From the definition of Gaussian Distribution, x has the probability density function (pdf) as following:

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

From the definition of expected value of a continuous random variable:

$$E[x] = \int_{-\infty}^{\infty} x^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

$$\rightarrow E[x] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$$

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

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

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

Hence,

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

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

$$\frac{1}{\pi} \cdot \mu\sqrt{\pi} = \mu$$

1.3 Variance of Gaussian distribution is sigma$^2(variance)$

$$Var[x] = E[x^2] - E[x]^2 = E[x^2] - \mu^2$$

To determine variance of Gaussian distribution, we have to find $E[x^2]$

$$E[x^2] = \int_{-\infty}^{\infty} x \dot{f}_x(x) dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{\frac{-y^2}{2\sigma^2}} dx$$

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

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

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

Hence,

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

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

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

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

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

So, $Var(x) = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$

1.4 Multivariate Gaussian distribution is normalized

Consider eigenvalues and eigenvectors of Σ

$$\Sigma u_i = \lambda_i u_i (i = 1, ..., D)$$

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

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

So the quadratic form becomes

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

where $y_i = u_i^T(x-\mu)$

And

$$|\Sigma|^{\frac{1}{2}} = \prod_{j=1}^{D}\lambda_j^{\frac{1}{2}}$$

Therefore we have

$$p(y) = \prod_{j=1}^{D}\frac{1}{2\pi\lambda_j}^{\frac{1}{2}}e^{\frac{-y_j^2}{2\lambda_j}} \Rightarrow \int_{-\infty}^{\infty}p(y)dy = \prod_{j=1}^{D}\int_{-\infty}^{\infty}\frac{1}{2\pi\lambda_j}^{\frac{1}{2}}e^{\frac{-y_j^2}{2\lambda_j}}dy_j$$

Substitute $t_j = \frac{y_j}{\sqrt{2\lambda_j}} \rightarrow dt_j = \frac{-1}{\sqrt{2\lambda_j}}dy_j$ So

$$= \prod_{j=1}^{D}\int_{-\infty}^{\infty}\frac{1}{2\pi\lambda_j}^{\frac{1}{2}}e^{-t^2}\sqrt{2\lambda_j}dt_j = \prod_{j=1}^{D}\int_{-\infty}^{\infty}\frac{1}{\pi}^{\frac{1}{2}}e^{-t^2}dt_j = \prod_{j=1}^{D}\frac{1}{\pi}^{\frac{1}{2}}\int_{-\infty}^{\infty}e^{-t^2}dt_j = \prod_{j=1}^{D}\frac{1}{\sqrt{\pi}}\sqrt{\pi} = 1$$

Therefore, Multivariate Gaussian distribution is normalized

2 Calculate:

2.1 The conditional of Gaussian distribution

Suppose $x \in R^D withGaussiandistributionN(x|\mu,\Sigma)andthatwepartitionxintotwodisjointsubsetsx_a andx_b$

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

Partitions of the mean μ vector and of the covariance matrix Σ

$$\mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix}$$

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

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

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

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

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

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

2.2 The marginal of Gaussian distribution

The margianl distribution given by

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

We need to integrate outxbby looking the quadratic form related to x_b

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

where $m = A_{bb}\mu_b - A_{ba}(x_a-\mu_a)$
We can integrate over unnormalized Gaussian

$$\int e^{\frac{-1}{2}(x_b - A_{bb}^{-1}m)^T A_{bb}(x_b - A_{bb}^{-1}m)}dx_b$$

The reamining term

$$-\frac{1}{2}x_a^T(A_{aa} - A_{ab}A_{bb}^{-1}A_{ba})x_a + x_a^T(A_{aa} - A_{ab}A_{bb}^{-1}A_{ba})^{-1}\mu_a + const$$

Similarly, we have

$$E[x_a] = \mu_a$$

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

$$\Rightarrow p(x_a) = N(x_a|\mu_a, \Sigma_{aa})$$