Machine Learning Homework Week 2

Dat Nguyen Ngoc

August 2022

1 Proof that:

1.1 Gaussian distribution is normalized:

We all know that the gaussian distribution is the following:

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

We want to prove that the above expression is normalized, we have to show:

$$\int\limits_{-\infty}^{\infty} N(x|\mu,\sigma^2) = 1$$

For the sake of simplicity let us assume that the mean in equation (1) is zero Let: $I = \int\limits_{-\infty}^{\infty} exp(\frac{-x}{2\sigma^2})dx$

$$I^2 = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} exp(\frac{-x}{2\sigma^2} - \frac{-y}{2\sigma^2})dxdy$$

To integrate this expression we make the transformation from Cartesian coordinates (x, y) to polar coordinates (r, ϕ), which is defined by

$x = rcos\phi$, $y = rsin\phi$

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

The Jacobian of the change of variable is given by:

$\frac{\partial(x,y)}{\partial(r,\mu)} = rcos^2\phi + rsin^2\phi = r$

We have:

$$\begin{aligned}
I^2 &= \int_0^{2\pi}\int_0^{\infty} exp(-\frac{r^2}{2\sigma^2})rdrd\phi \\
&= 2\pi\int_0^{\infty} exp(-\frac{r^2}{2\sigma^2})rdr \\
&= 2\pi\int_0^{\infty} exp(-\frac{u}{2\sigma^2})\frac{1}{2}du \\
&= \pi[exp(\frac{-u}{2\sigma^2})(-2\sigma^2)] \\
&= 2\pi\sigma^2
\end{aligned}$$

$\Rightarrow$I = $\sqrt{2\pi\sigma^2}$
Tranform $y = x - \mu$, so that:

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

1.2 Expectation value of Gaussian distribution is μ:

Before prove the Expectation value of Gaussian distribution is μ. I will prove $\int_{-\infty}^{\infty} exp(-x^2)dx = \sqrt{\pi}$ and $\int_{-\infty}^{\infty} texp(-t^2)dt = 0$

Let:

$$\begin{aligned}
I &= \int\limits_{-\infty}^{\infty} exp(-x^2)dx \\
I^2 &= \int\limits_{-\infty}^{\infty} exp(-x^2)dx \int\limits_{-\infty}^{\infty} exp(-y^2)dy \\
&= \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} exp(-x^2-y^2)dxdy \\
&= \int\limits_{0}^{2\pi}\int\limits_{0}^{\infty} rexp(-r^2)drd\phi \\
&= \pi
\end{aligned}$$

$\Rightarrow$I $= \sqrt{\pi}$

Cause t is the odd function and $\exp(-t^2)$ is even function so $texp(-t^2)$ is odd function

$\Rightarrow \int\limits_{-\infty}^{\infty} texp(-t^2)dt = 0$

We have:

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

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

$$\begin{aligned}
&= \frac{\sigma\sqrt{2}}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} (\sqrt{2}t\sigma+\mu)exp(-t^2)dt \\
&= \frac{1}{\sqrt{\pi}}(\sqrt{2}\sigma \int\limits_{-\infty}^{\infty} texp(-t^2)dt + \mu \int\limits_{-\infty}^{\infty} exp(-t^2)dt) \\
&= \frac{1}{\sqrt{\pi}}(0+\mu\sqrt{\pi})(provedabove) \\
&= \mu
\end{aligned}$$

1.3 Variance of Gaussian distribution is $sigma^2$:

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

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

$$\begin{aligned}
&= \frac{\sigma\sqrt{2}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sqrt{2}t\sigma + \mu)^2 exp(-t^2)dt - \mu^2 \\
&= \frac{1}{\sqrt{\pi}}(\sqrt{2}\sigma \int_{-\infty}^{\infty} t^2 exp(-t^2)dt + 2\sqrt{2}\sigma\mu \int_{-\infty}^{\infty} texp(-t^2)dt + \mu^2 \int_{-\infty}^{\infty} exp(-t^2)dt) - \mu^2 \\
&= \frac{1}{\sqrt{\pi}}(\sqrt{2}\sigma \int_{-\infty}^{\infty} t^2 exp(-t^2)dt + 0 + \mu^2\sqrt{\pi}) - \mu^2 (provedabove) \\
&= \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} t^2 exp(-t^2)dt \\
&= \frac{2\sigma^2}{\sqrt{\pi}} \frac{1}{2} \int_{-\infty}^{\infty} exp(-t^2)dt \\
&= \sigma^2
\end{aligned}$$

1.4 Multivariate Gaussian distribution is normalized:

For a D-dimensional vector x, the multivariate Gaussian distribution takes the form

μ is a D-dimensional mean vector, Σ is a D x D covariance matrix, and $|\Sigma|$ denotes the det of Σ

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

Set:

$$\begin{aligned}
\Delta^2 &= -\frac{1}{2}(x-\mu)^T(\Sigma^{-1})(x-\mu) \\
&= \frac{-1}{2} x^T \Sigma^{-1} x + x^T \Sigma^{-1} \mu + const
\end{aligned}$$

It is a quadratic form of Gaussian distribution
Consider the 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 its eigenvectors 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 that:

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

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

$p(y) = \prod_{j=1}^{D} \frac{1}{2\pi\lambda_j}^{\frac{1}{2}} exp(-\frac{y_j^2}{2\lambda_j})$

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

If two sets of variables are jointly Gaussian, then the conditional distribution of one set conditioned on the other is again Gaussian. Similarly, the marginal distribution of either set is also Gaussian

2 Calculate:

2.1 The conditional of Gaussian distribution

Suppose x is a D-dimensional vector with Gaussian distribution $N(x|\mu, \Sigma)$ and that we partition x into two disjoint subsets x_a and x_b

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

We also define corresponding partitions of the mean vector μ given by

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

and of the covariance matrix Σ given by

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

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

We are looking for conditional distribution $p(x_a|x_b)$

We have

$$
\begin{aligned}
-\frac{1}{2}(x-\mu)^T(\Sigma^{-1})(x-\mu) &= -\frac{1}{2}(x-\mu)^T A(x-\mu) \\
&= -\frac{1}{2}(x_a-\mu_a)^T A_{aa}(x_a-\mu_a) \\
&\quad -\frac{1}{2}(x_a-\mu_a)^T A_{ab}(x_b-\mu_b) \\
&\quad -\frac{1}{2}(x_b-\mu_b)^T A_{ba}(x_a-\mu_a) \\
&\quad -\frac{1}{2}(x_b-\mu_b)^T A_{bb}(x_b-\mu_b) \\
&= -\frac{1}{2}x_a^T(A_{aa})^{-1}x_a + x_a^T(A_{aa}\mu_a - A_{ab}(x_b-\mu_b))
\end{aligned}
$$

Compare with Gaussian disitrubtion

$\Delta^2 = \frac{-1}{2}x^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu + const$

$\Rightarrow \Sigma_{a|b} = A_{aa}^{-1}$

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

By using Schur complement,

$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} M & -MBD^{-1} \\ -D^{-1}CMD^{-1} & D^{-1}CMBD^{-1} \end{pmatrix}$

where we have defined $M = (A - BD^{-1}C)^{-1}$

Using this definition

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

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

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

As the result:

$$
\begin{aligned}
\mu_{a|b} &= \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b-\mu_b) \\
\Sigma_{a|b} &= \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}
\end{aligned}
$$

$\Rightarrow \mathrm{p}(\mathrm{x}_a|x_b) = N(x_{a|b}|\mu_{a|b}, \Sigma_{a|b})$

2.2 The marginal of Gaussian distribution

The marginal distribution given by:

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

We need to integrate out x_b by looking the quadratic form related to x_b

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

with $m = A_{bb}\mu_b - A_{ba}(x_a - \mu_a)$

We can integrate over unnormalized Gaussian:
$\int exp[-\dfrac{1}{2}(x_b - A_{bb}^{-1}m)^T A_{bb}(x_b - A_{bb}^{-1}m)]dx_b$

The remaining term

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

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

Similarly, we have

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

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