Machine Learning 1: Homework week 2

Mai Xuan Bach
ID 11200489

Ngày 24 tháng 8 năm 2022

1 Problem 1. Proof that:

a. Gaussian distribution is normalized:

Have to prove:

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

Substitute $y = \frac{x-\mu}{\sigma} \iff dy = \frac{dx}{\sigma} \iff dx = \sigma\, dy$

It means: $\sigma\int_{-\infty}^{+\infty} e^{\frac{-y^2}{2}}\,dy = \sqrt{2\pi\sigma^2} \iff \int_{-\infty}^{+\infty} e^{\frac{-y^2}{2}}\,dy = \sqrt{2\pi} = I$

We compute:

$$
\begin{aligned}
I^2 &= \int_{-\infty}^{+\infty} e^{\frac{-y^2}{2}}\,dy \int_{-\infty}^{+\infty} e^{\frac{-y^2}{2}}\,dy && (1)\\
&= \int_{-\infty}^{+\infty} e^{\frac{-y^2}{2}}\,dy \int_{-\infty}^{+\infty} e^{\frac{-x^2}{2}}\,dx && (2)\\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{\frac{-(x^2+y^2)}{2}}\,dy\,dx && (3)
\end{aligned}
$$

With: $r^2 = x^2 + y^2$

$x = rcos\theta$

$y = rsin\theta$

$$
\begin{aligned}
I^2 &= \int_0^{2\pi}\int_0^{+\infty} e^{\frac{-r^2}{2}} r dr d\theta &\quad (4)\\
&= \int_0^{+\infty}\int_0^{2\pi} e^{\frac{-r^2}{2}} r d\theta dr &\quad (5)\\
&= \int_0^{+\infty} e^{\frac{-r^2}{2}} r 2\pi dr &\quad (6)\\
&= -2\pi\int_0^{+\infty} e^{\frac{-r^2}{2}} d(\frac{-r^2}{2}) &\quad (7)\\
&= -2\pi * 0 - (-2\pi) = 2\pi &\quad (8)
\end{aligned}
$$

So that, $I = \sqrt{2\pi}$, i.e Gaussian distribution is normalized

b. Expectation of Gaussian distribution is mu (mean)

Have to prove: $E[x] = \int_{-\infty}^{+\infty} xp(x|\mu,\sigma^2)dx = \mu$

That means: $\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{+\infty} xe^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \mu$

Or: $\int_{-\infty}^{+\infty} xe^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \mu\sqrt{2\pi\sigma^2}$

$$
\begin{aligned}
I &= \int_{-\infty}^{+\infty} xe^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx &\quad (9)\\
& &\quad (10)
\end{aligned}
$$

Let $t = \frac{x-\mu}{\sigma} \Longleftrightarrow dt = \frac{1}{\sigma}dx \Longleftrightarrow dx = \sigma dt$

$$
\begin{aligned}
I &= \int_{-\infty}^{+\infty} (\sigma t+\mu)e^{\frac{-t^2}{2}}\sigma\, dt &\quad (11)\\
&= \sigma\left(\int_{-\infty}^{+\infty} \sigma t e^{\frac{-t^2}{2}} dt + \mu\int_{-\infty}^{+\infty} e^{\frac{-t^2}{2}} dt\right) &\quad (12)\\
&= \sigma\left(\sigma\int_{-\infty}^{+\infty} t e^{\frac{-t^2}{2}} dt + \mu\int_{-\infty}^{+\infty} e^{\frac{-t^2}{2}} dt\right) &\quad (13)\\
&= \sigma\left(-\sigma\int_{-\infty}^{+\infty} e^{\frac{-t^2}{2}} d(\frac{-t^2}{2}) + \mu\sqrt{2\pi}\right) &\quad (14)\\
&= \sigma\left(0 - 0 + \mu\sqrt{2\pi}\right) &\quad (15)\\
&= \mu\sqrt{2\pi\sigma^2} &\quad (16)
\end{aligned}
$$

So that, Expectation of Gaussian distribution is μ.

c. Variance of Gaussian distribution is σ^2 (variance)

Have to prove: $Var[x] = \int_{-\infty}^{+\infty}(x-\mu)^2 p(x|\mu,\sigma^2)dx = \sigma^2$

It means: $\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (x-\mu)^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \sigma^2$

Or: $\int_{-\infty}^{+\infty} (x-\mu)^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \sigma^2\sqrt{2\pi\sigma^2} = \sigma^3\sqrt{2\pi} = I$

Let $t = \frac{x-\mu}{\sigma} \Longleftrightarrow dt = \frac{1}{\sigma}dx \Longleftrightarrow dx = \sigma dt$

$$\begin{aligned}
I &= \sigma \int_{-\infty}^{+\infty} \sigma^2.t^2.e^{\frac{-t^2}{2}}\, dt && (17)\\
&= \sigma^3 \int_{-\infty}^{+\infty} .t^2.e^{\frac{-t^2}{2}}\, dt && (18)\\
&= \sigma^3 \int_{-\infty}^{+\infty} .t.t.e^{\frac{-t^2}{2}}\, dt && (19)\\
&= \sigma^3 \left(-\int_{-\infty}^{+\infty} .t.e^{\frac{-t^2}{2}}\, d(\frac{-t^2}{2}) \right) && (20)\\
&= \sigma^3 \left(-\int_{-\infty}^{+\infty} .t.\, d(e^{\frac{-t^2}{2}}) \right) && (21)
\end{aligned}$$

By integration by parts:

$$\begin{aligned}
I &= -\sigma^3 \left(t.e^{\frac{-t^2}{2}} |_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} e^{\frac{-t^2}{2}}\, dt \right) && (22)\\
&= -\sigma^3 \left(0 - 0 - \sqrt{2\pi} \right) && (23)\\
&= \sigma^3\sqrt{2\pi} && (24)
\end{aligned}$$

So that, Variance of Gaussian distribution is σ^2 (variance)

d. Multivariate Gaussian distribution is normalized

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

Set

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

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 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,

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

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

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

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

2 Problem 2. Calculate

a. The conditional of Gaussian distribution:

Suppose x is a D-dimensional vector with Gaussian distribution $\mathcal{N}(x|\mu,\Sigma)$ 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 Σ_{aa} and Σ_{bb} are symmetric while $\Sigma_{ab} = \Sigma_{ba}^T$

We have,

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

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

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

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

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

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

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

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

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

Compare with the form:

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

We got:

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

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

Using Schur complement: $\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} M & -MBD^{-1} \\ -D^{-1}CMD^{-1} & D^{-1}CMBD^{-1} \end{pmatrix}$ with $M = (A-BD^{-1}C)^{-1}$

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

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

$\Rightarrow \Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}.\Sigma_{bb}^{-1}\Sigma_{ba}$

$\mu_{a|b} = \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b - \mu_b)$

b. The marginal of Gaussian distribution:

The marginal distribution given by

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

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

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

Combining this term with the remaining terms that depend on x_a, we obtain:

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

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

Again, comparison give us:

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

Noted that:

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

and

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

We got: $E(x_a) = \mu_a$

$cov(x_a) = \Sigma_{aa}$