Ex1:
Multivariate Gaussian Distribution:
Proof that Multivariate Gaussian Distribution is normalize:

Proof:
First, we have the PDF of the Gaussian Distribution is:

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

The Gaussian Distribution is normalize

$$\Longleftrightarrow \int_{-\infty}^{+\infty} p\left(x \mid \mu, \sigma^2\right)=1$$

Where μ is a D-dimensional mean vector Σ is a D xD covariance matrix $|\Sigma|$ denotes the determinant of Σ
Set

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

Consider eigenvalues and eigenvvectors of Σ we have:

$$\Sigma u_i=\lambda_i u_i, i=1, \ldots, D$$

Because Σ is a real, symmetric matrix $\rightarrow$ its eigenvalues will be real and its eigenvectors form an orthnormal set. Proof: 1. its eigenvalues will be real Example:

$$\left(\begin{array}{cc}\sigma_1^2 & \sigma_{1,2} \\ \sigma_{2,1} & \sigma_2^2\end{array}\right)$$

The equation to find the eigenvalues is :

$$\begin{gathered}\left(\sigma_1^2-\lambda\right) \times\left(\sigma_2^2-\lambda\right)-\left(\sigma_{1,2}\right)^2=0 \\ \Longleftrightarrow\left(\sigma_1^2-\lambda\right) \times\left(\sigma_2^2-\lambda\right)=\left(\sigma_{1,2}\right)^2\end{gathered}$$

$\Longrightarrow$ With $\lambda=\lambda_1$:

$$\begin{gathered}\left(\begin{array}{cc}\sigma_1^2-\lambda_1 & \operatorname{cov}\left(\sigma_{1,2}\right) \\ \operatorname{cov}\left(\sigma_{1,2}\right) & \sigma_2^2-\lambda_1\end{array}\right) \\ \left(\sigma_1^2-\lambda_1\right) x_1+\left(\sigma_{1,2}\right) x_2=0(1) \\ \left(\sigma_{1,2}\right) x_1+\left(\sigma_2^2-\lambda_1\right) x_2=0(2)\end{gathered}$$

From (1) we have:

$$\begin{gathered}x_1=\frac{-y \times \operatorname{cov}\left(\sigma_{1,2}\right)}{\sigma_1^2-\lambda_1} \\ x_2=x_2\end{gathered}$$

So the eigenvector in this case is:

$$\left(\frac{-\operatorname{cov}\left(\sigma_{1,2}\right)}{\sigma_1^2-\lambda_1}\right)$$

With $\lambda=\lambda_2$:

$$\begin{gathered}\left(\begin{array}{cc}\sigma_1^2-\lambda_2 & \operatorname{cov}\left(\sigma_{1,2}\right) \\ \operatorname{cov}\left(\sigma_{1,2}\right) & \sigma_2^2-\lambda_2\end{array}\right) \\ \left(\sigma_1^2-\lambda_2\right) x_1+\left(\sigma_{1,2}\right) x_2=0(3) \\ \left(\sigma_{1,2}\right) x_1+\left(\sigma_2^2-\lambda_2\right) x_2=0(4)\end{gathered}$$

From (3) we have:

$$x_1 = \frac{-y \times \operatorname{cov}(\sigma_{1,2})}{\sigma_1^2 - \lambda_2}$$
$$\longrightarrow x_2 = x_2$$

So the eigenvector in this case is:

$$\left(\frac{-\operatorname{cov}(\sigma_{1,2})}{\sigma_1^2 - \lambda_2} \right)$$

And:

$$\begin{pmatrix} \frac{-\operatorname{cov}(\sigma_{1,2})}{\sigma_1^2-\lambda_2} \\ 1 \end{pmatrix}^T \times \left(\frac{-\operatorname{cov}(\sigma_{1,2})}{\sigma_1^2 - \lambda_2} \right) = 1$$

So its eigenvectors form an orthnormal set.

$$\Sigma = \Sigma_{i=1}^D \lambda_i u_i (u_i)^T \longrightarrow \Sigma^{-1} = \Sigma_{i=1}^D \frac{1}{\lambda i} u_i u_i^T$$

So that:

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

Let:

$$y_i = u_i^T (x-\mu)$$
$$\longrightarrow \Delta^2 = \sum_{i=1}^D \frac{y_i^2}{\lambda_i}$$
$$|\Sigma|^{1/2} = \prod_{j=1}^D \lambda_j^{1/2}$$

Now, we have:

$$p\left(x \mid \mu, \sigma^2\right) = p(y) = \prod_{j=1}^D \frac{1}{(2\pi\lambda_j)^{1/2}} e^{\frac{-(y_j)^2}{2\lambda_j}}$$
$$\Longleftrightarrow \int_{-\infty}^{+\infty} p(y)dy = \prod_{j=1}^D \int_{-\infty}^{+\infty} \frac{1}{(2\pi\lambda_j)^{1/2}} e^{\frac{-(y_j)^2}{2\lambda_j}}$$

But in the last homework we have proofed that:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-y^2} 2dy = 1$$

So:

$$\int_{-\infty}^{+\infty} \frac{1}{(2\pi\lambda_j)^{1/2}} e^{\frac{-(y_j)^2}{2\lambda_j}} = 1$$
$$\Longleftrightarrow \prod_{j=1}^D \int_{-\infty}^{+\infty} \frac{1}{(2\pi\lambda_j)^{1/2}} e^{\frac{-(y_j)^2}{2\lambda_j}} = 1$$

Ex2:

Calculate marginal normal distribution

Proof:
Let:

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

Where: and So, the dimension of

$$\begin{aligned}
x^T\Sigma^{-1}\mu &= 1\times D \otimes D\times D \otimes D\times 1 \\
&= 1\times 1
\end{aligned}$$

$\rightarrow x^T\Sigma^{-1}\mu$ equal a numeric value

$$\begin{aligned}
\Rightarrow x^T\Sigma^{-1}\mu &= \left(x^T\Sigma^{-1}\mu\right)^T \\
&= \mu^T\Sigma^{-1}x
\end{aligned}$$

Therefore, equation (2) can be rewritten as

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

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

$$\Sigma=\begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$$

Let

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

We are looking for conditional distribution $p\left(x_a \mid x_b\right)$. We have:

$$\begin{aligned}
\Delta^2 &= -\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) \\
&= -\frac{1}{2}(x-\mu)^TA(x-\mu) \\
&= -\frac{1}{2}\begin{bmatrix} x_a-\mu_a \\ x_b-\mu_b \end{bmatrix}^T\begin{bmatrix} A_{aa} & A_{ab} \\ A_{ba} & A_{bb} \end{bmatrix}\begin{bmatrix} x_a-\mu_a \\ x_b-\mu_b \end{bmatrix} \\
&= -\frac{1}{2}\left(x_a-\mu_a\right)^TA_{aa}\left(x_a-\mu_a\right)-\frac{1}{2}\left(x_a-\mu_a\right)^TA_{ab}\left(x_b-\mu_b\right) \\
&\quad -\frac{1}{2}\left(x_b-\mu_b\right)^TA_{ba}\left(x_a-\mu_a\right)-\frac{1}{2}\left(x_b-\mu_b\right)^TA_{bb}\left(x_b-\mu_b\right) \\
&= -\frac{1}{2}x_a^TA_{aa}x_a+x_a^T\left[A_{aa}\mu_a-A_{ab}\left(x_b-\mu_b\right)\right]+const
\end{aligned}$$

Compare with Gaussian distribution

$$\Delta^2 = -\frac{1}{2}x^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu + const$$
$$\rightarrow \begin{cases} -\frac{1}{2}x^T\Sigma^{-1}x = -\frac{1}{2}x_a^T A_{aa}x_a \\ x^T\Sigma^{-1}\mu = x_a^T\left[A_{aa}\mu_a - A_{ab}\left(x_b-\mu_b\right)\right] \end{cases}$$
$$\Leftrightarrow \begin{cases} \Sigma^{-1} = A_{aa} \\ \Sigma^{-1}\mu = A_{aa}\mu_a - A_{ab}\left(x_b-\mu_b\right) \end{cases}$$
$$\Leftrightarrow \begin{cases} \Sigma^{-1} = A_{aa} \\ \mu = \mu_a - A_{aa}^{-1}A_{ab}\left(x_b-\mu_b\right) \end{cases}$$

By using Schur complement,

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} M & -MBD^{-1} \\ -D^{-1}CMD^{-1} & D^{-1}CMBD^{-1} \end{pmatrix}, \text{ with } M = \left(A - BD^{-1}C\right)^{-1}$$

As a result,

$$\begin{cases} \mu_{a|b} = \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}\left(x_b-\mu_b\right) \\ \Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba} \end{cases}$$
$$\Rightarrow p\left(x_a \mid x_b\right) = \mathcal{N}\left(x_{a|b} \mid \mu_{a|b}, \Sigma_{a|b}\right)$$

Ex3:
Calculate conditional normal distribution

Proof:
The marginal distribution given by:

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

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

$$\Delta^2 = -\frac{1}{2}(x-\mu)^T A(x-\mu)$$
$$= -\frac{1}{2}x_b^T A_{bb}x_b + x_b^T m + \text{ const } (\text{ with } m = A_{bb}\mu_b - A_{ba}\left(x_a-\mu_a\right))$$
$$= -\frac{1}{2}\left(x_b - A_{bb}^{-1}m\right)^T A_{bb}\left(x_b - A_{bb}^{-1}m\right) + \frac{1}{2}m^T A_{bb}^{-1}m$$

Integrate over unnormalized Gaussian:

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

The remaining term:

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

Similarly we have:

$$\mathbb{E}\left[x_a\right] = \mu_a$$
$$\operatorname{cov}\left[x_a\right] = \Sigma_{aa}$$
$$\Rightarrow p\left(x_a\right) = \mathcal{N}\left(x_a \mid \mu_a, \Sigma_{aa}\right)$$