Multivariate normal distribution Or Gaussian Distribution

$$P(x) := N_x(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$x \sim N_x(\mu, \sigma^2)$$
$$E[x] = \mu$$
$$Var[x] = \sigma^2$$

Motivation: Central limit Theorem

$x_1, x_2 \cdots x_n$ are iid R.V. such that

$E[x_i] = \mu \quad Var[x_i] = \sigma^2$

$$\frac{1}{\sqrt{n}} \sum_i (x_i - \mu) \xrightarrow{d} N(0, \sigma^2)$$

example: $y = g(x) + \underset{\sim N}{\epsilon} \Rightarrow e_1 + e_2 + \cdots e_n$

Multivariate Generalization:

$$P(x) := N_x(\mu; \Sigma)$$
$$= \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left(-\frac{(x-\mu)^T \Sigma^{-1} (x-\mu)}{2}\right)$$

$\mu \in R^d, \ \Sigma \in R^{d \times d}$

$\Sigma > 0$, P.D.

Geometry of MVN

$-\underline{(x-\mu)^T\Sigma^{-1}(x-\mu)}$ determines geometry.

Mahalanobis distance.
It's a rotated and shifted Euclidean geometry.

$-\underline{(x-\mu)^T U} D U^T(x-\mu)$ $\quad U\in R^{d\times d}$ ort. m

$D\in R^{d\times d}$ diag.

diag$(D) > 0$

$y = U^T(x-\mu)$

$y\in R^d$, $\quad y^T=(x-\mu)^T U$

$\sum_i y_i^2 D_i$

$$P(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\exp\left(-\frac{y^T D y}{2}\right)$$

$$= \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\prod_i \exp\left[-\frac{y_i^2}{2D_i}\right]$$

$1/D_i$ is i-th diagonal elem. of D.

$$=\prod_i \frac{1}{(2\pi)^{d/2} D_i^{1/2}}\exp\left[-\frac{y_i^2}{2D_i}\right]$$

$$P(y) = P(x)\,|Jac_y(Uy+\mu)| = \prod_i N_y(0, D_i)$$

Product of d- univariate normal.

$\boxed{\text{Essentially, MVN under a linear coordinate transform is a product of } d \text{ univariate normal.}}$

Normalization of MVN

$P(y) = \prod_i N_y(0, D_i)$

so, $\int P(y)\,dy = \int \prod_i N_y(0, D_i)\,dy$

$= \prod_i \int N_y(0, D_i)\,dy$

$= 1$

MVN is normalized under this new coord. sys.

Moments of MVN.

x is a MVN with P.D.F. $N_x(\mu, \Sigma)$

$$E[x] = \int \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \cdot \exp\left(-(x-\mu)^T\Sigma^{-1}(x-\mu)/2\right) x\,dx$$

$$= \int \cdots\cdots \exp(-z^T\Sigma^{-1}z)(z+\mu)\,dz$$

$$= \left(\int \cdots\cdots \exp\cdots dz\right)\mu + \underbrace{\int \cdots \exp(\cdots) z\,dz}_{\substack{0 \\ \text{symmetry}}}$$

$$= \mu \cdot 1$$

$$\int_{-\infty}^{0} \cdots \exp(\cdots) z_i\,dz_i$$

$$= \int_{0}^{\infty} \cdots \exp(\cdots) - z_i d$$

$$E[xx^T] = \int \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{(x-\mu)^T\Sigma^{-1}(x-\mu)}{2}\right) x \cdot x_i^T$$

$$= \int \cdots \exp\left(-\frac{z^T\Sigma^{-1}z}{2}\right)\underbrace{(z+\mu)(z+\mu)^T}_{zz^T+\mu z^T+z\mu^T+\mu\mu^T}dz$$

$E[\mu z^T] = E[z\mu^T] = 0 \quad E[\mu\mu^T] = \mu\mu^T$

$E[\mu z^T] = E[z\mu^T] = 0 \quad E[\mu\mu^T] = \mu\mu^T$

$E[zz^T]$?

$$= \int \cdots \exp\left(-\frac{z^T \Sigma^{-1} z}{2}\right) z z^T dz$$

$$= \int \cdots \exp\left(-\frac{z^T U D U^T z}{2}\right) z z^T dz$$

$$z = (y^T U^T)^T = Uy$$

$$= \int \cdots \exp\left(-\frac{y^T U^T U D U^T U y}{2}\right) U y y^T U^T dy$$

$$= \int \cdots \exp\left(-\frac{y^T D y}{2}\right) \sum_{ij} u_i y_i y_j u_j^T$$

$$= \sum_{ij} u_i u_j^T \int \cdots \exp\left(-\frac{y^T D y}{2}\right) y_i y_j \, dy$$

$$\forall i \neq j \quad \int_{-\infty}^{+\infty} \cdots \exp\left(-\frac{y^T D y}{2}\right) y_i y_j \, dy = 0$$

since $\forall i = j \quad = \int_{-\infty}^{\infty} \cdots \exp\left(-\frac{y_i^2}{2D_i}\right) y_i^2 \, dy = D_i \Leftarrow$ odd function

$$= \sum_i u_i u_i^T D_i = U (D)^{-1} U^T$$

$$= (\Sigma^{-1})^{-1} = \Sigma$$

$$\therefore \quad E[xx^T] = \Sigma + \mu\mu^T$$

Conditional MVN.

$$P(x_a, x_b) = \mathcal{N}_x\left(\begin{bmatrix}\mu_a \\ \mu_b\end{bmatrix}, \begin{bmatrix}\Sigma_{aa}, \Sigma_{ab} \\ \Sigma_{ba}, \Sigma_{bb}\end{bmatrix}\right)$$

$P(x_a \mid x_b)$?

As we saw earlier, the qudratic form completely determines a MVN. Let's look for the qudratic form with respect to x_a

$$\ln P(x_a, x_b) = -\left(x - \begin{bmatrix}\mu_a \\ \mu_b\end{bmatrix}\right)^T \Sigma^{-1} \left(x - \begin{bmatrix}\mu_a \\ \mu_b\end{bmatrix}\right) + \text{Const}$$

$$\log p(x_a, x_b) = -(x-[\mu_b])^T \Sigma^{-1} (x-[\mu_b])/2 + \text{const.}$$

Let $\Theta = \Sigma^{-1}$, $\Theta := \begin{bmatrix} \Theta_{aa}, & \Theta_{ab} \\ \Theta_{ba}, & \Theta_{bb} \end{bmatrix}$

$$
\begin{aligned}
= &-[x_a - \mu_a]^T \Theta_{aa} [x_a - \mu_a]/2 \\
&- [x_a - \mu_a]^T \Theta_{ab} [x_b - \mu_b]/2 \\
&- [x_b - \mu_b]^T \Theta_{ba} [x_a - \mu_a]/2 \\
&- [x_b - \mu_b]^T \Theta_{bb} [x_b - \mu_b]/2
\end{aligned}
$$

The only qudratic form w.r.t x_a : $-x_a^T \Theta_{aa} x_a/2$

We know a MVN, $N_x(\mu, \Sigma)$ has an exponent:

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

$\Longrightarrow$ $P(x_a | x_b)$ has covariance Θ_{aa}^{-1} (NOT Σ_{aa}

Find linear terms includes x_a :

$$
\begin{aligned}
&-(x_a - \mu_a)^T \Theta_{ab} x_b + (x_a - \mu_a)^T \Theta_{ab} \mu_b \\
&+ (x_a - \mu_a)^T \Theta_{aa} \mu_b \quad \text{using } \Theta_{ab} = \Theta_{ba} \\
&= x_a^T [\Theta_{aa} \mu_a - \Theta_{ab} x_b + \Theta_{ab} \mu_b] \\
&= x_a^T \Theta_{aa} [\mu_a - \Theta_{aa}^{-1} \Theta_{ab} x_b + \Theta_{aa}^{-1} \Theta_{ab} \mu_b] \\
&= x_a^T \Theta_{aa} [\underbrace{\mu_a - \Theta_{aa}^{-1} \Theta_{ab} [x_b - \mu_b]}_{\mu_{a|b}}]
\end{aligned}
$$

Marginalization of MVN.

$$P(x_a, x_b) = N_x\left(\begin{bmatrix} \mu_a \\ \mu_b \end{bmatrix}, \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix} \right)$$

find terms w.r.t. x_b

$$
\begin{aligned}
\log P(x_a, x_b) = &-(x_a - \mu_a)^T \Theta_{ab} (x_b - \mu_b)/2 \\
&- (x_b - \mu_b)^T \Theta_{ba} (x_a - \mu_a)/2 \\
&- (x_b - \mu_b)^T \Theta_{bb} (x_b - \mu_b)/2
\end{aligned}
$$

$+ \text{const.}$

$$-x_b^T \Theta_{bb} x_b/2 + x_b^T \Theta_{bb} \mu_b - x_b^T \Theta_{ba} x_a + x_b^T \Theta_{ba} \mu$$

$$= -x_b^T \Theta_{bb} x_b/2 + x_b^T \underbrace{(\Theta_{bb}\mu_b - \Theta_{ba} x_a + \Theta_{ba}\mu_a)}_{m}$$

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

completing the square !!

$$\therefore \int P(x_a, x_b)\, dx_b = \frac{1}{2\pi^{D/2} |\Sigma|^{\frac{1}{2}}} \exp\left(t + \frac{m \Theta_{bb}^{-1} m}{2}\right) \cdot \text{const}$$

$$t = -\frac{x_a \Theta_{aa} x_a}{2} + x_a \Theta_{aa} \mu_a + x_a \Theta_{ab} \mu_b + \text{const.}$$

$$m = (\Theta_{bb}\mu_b - \Theta_{ba} x_a + \Theta_{ba}\mu_a)$$

Hint: look for qudratic and linear terms w.r.t. x_a, other terms does not matter

$$t + m \Theta_{bb}^{-1} m = -\frac{x_a^T \Theta_{aa} x_a}{2} + \frac{x_a^T \Theta_{ab} \Theta_{bb}^{-1} \Theta_{ba} x_a}{2}$$

$$+ x_a^T \Theta_{aa} \mu_a + x_a^T \Theta_{ab} \mu_b - x_a^T \Theta_{ab} \Theta_{bb}^{-1} \Theta_{bb} \mu_b - x_a^T \Theta_{ab} \Theta_{bb}^{-1} \Theta_{ba} \mu_a$$

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

$$= -(x_a - \mu_a)^T \underbrace{(\Theta_{aa} - \Theta_{ab} \Theta_{bb}^{-1} \Theta_{ba})}_{\Sigma_{aa}^{-1}} (x_a - \mu_a) + \text{const}$$

$$\text{so, } \int P(x_a, x_b)\, dx_b \propto \exp\left[-\frac{(x_a - \mu_a)^T \Sigma_{aa}^{-1} (x_a - \mu}{}\right.$$

is a MVN, with mean μ_a, cov Σ_{aa}

MLE of MVN. given $x_1 \cdots x_n$

$$\max_{\mu, \Sigma} L(x_1 \cdots x_n, \mu, \Sigma)$$

$$L = \text{const} - \frac{N}{2}\log|\Sigma| - \sum_{i=1}^{n}(x_i-\mu)\Sigma^{-1}(x_i-\mu)/2$$

$$\frac{\partial L}{\partial \mu} = -\sum_{i=1}^{n} \Sigma^{-1}(x_i-\mu)$$

set $\frac{\partial L}{\partial \mu} = 0 \Rightarrow \quad 0 = -\sum_{i=1}^{n} \Sigma^{-1}(x_i-\mu)$

$$n\mu = \sum_{i=1}^{n} x_i$$

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\frac{\partial L}{\partial \Sigma} = -\frac{n}{2}\Sigma^{-1} - \frac{\partial \, tr[\bar{X}^T \Sigma^{-1} \bar{X}]/2}{\partial \Sigma}, \quad \frac{\partial \log|X|}{\partial X} = X^{-1}$$

$\boxed{\bar{X} = [x_1 \cdots x_n] - \mu} \in R^{d \times n}$

$$\frac{\partial \, tr[X \Sigma^{-1} X]}{\partial} = -(\Sigma^{-1}\bar{X}\bar{X}^T\Sigma^{-1})^T/2$$

$$= -\Sigma^{-1}\bar{X}\bar{X}^T\Sigma^{-1}/2$$

set $\frac{\partial L}{\partial \Sigma} = 0 \Rightarrow -\frac{n}{2}\Sigma^{-1} = -\Sigma^{-1}\bar{X}\bar{X}^T\Sigma^{-1}/$

$$n\Sigma = \bar{X}\bar{X}^T$$

$$\Sigma = \bar{X}\bar{X}^T/n$$