Machine learning

September 2021

1.Multivariate Gaussian distribution
Proof Multivariate Gaussian distribution normalization

$$p(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}$$

Gaussian Distribution normalize :

$$\Leftrightarrow \int_{-\infty}^{\infty} p(x \mid \mu, \sigma^2) = 1$$

Have mean:

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

Variance:

$$Var[x] = \int_{-\infty}^{\infty} (x-\mu)^2 p(x \mid \mu, \sigma^2) = \sigma^2$$

Where the $\mu\ is\ a\ D-dimensional\ mean\ \Sigma\ \ isDxD\ covariance,\ and\ \ \mid \Sigma \mid$ $denotes\ the\ of\ \sigma$
$Set:$
$\Delta^2 = \frac{-1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu) = \frac{-1}{2}(x^T - \mu^T)\Sigma^{-1}(x-\mu)$

$= \frac{-1}{2}x^T \Sigma^T x + \frac{1}{2}x^T \Sigma^{-1}\mu + \frac{1}{2}\mu^T \Sigma^{-1} x + \frac{-1}{2}\mu^T \Sigma^{-1}\mu$
$Proof: \Sigma: symmetric, then\ \Sigma^{-1}\ is\ symmetric$
$We\ have: \Sigma\Sigma^{-1} = I$

$$I = I^T$$

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

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

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

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

Then Σ^{-1} is symmetric
So that:

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

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 eigenvector form an orthnormal set.
Proof: $\Sigma = \Sigma_{i=1}^{D} \lambda_i u_i u_i^T => \Sigma^{-1} = \Sigma_{i=1}^{D} \frac{1}{\lambda_i} u_i u_i^T$
$Let : \Sigma = \Sigma_{i=1}^{D} \lambda_i u_i u_i^T = UDU^T$
$UhavedimesionDxD$
$Disdiagonalmatrixwiththeeigenvalue\lambda$ its diagonal
U is orthogonal matrix so $U^{-1} = U^T$
$\Sigma^{-1} = (UDU^T) = (U^T)^{-1} D^{-1} U^{-1} = UD^{-1}U^T = \Sigma^{-1} = \Sigma_{i=1}^{D} \frac{1}{\lambda_i} u_i u_i^T$
$\Delta^2 = (x-\mu)^T \Sigma^{-1} (x-\mu) = \sum_{i=1}^{D} \lambda_i u_i u_i^T = \sum_{i=1}^{D} \frac{1}{\lambda_i} (x-\mu)^T u_i u_I^T (x-\mu)$
$With : y_i = u_i^T (x-\mu)$

$$\Delta^2 = \sum_{i=1}^{D} \frac{y_i^2}{\lambda_i}$$

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

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

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

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

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

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:

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

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

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

Compare with Gaussian distribution:
$\Sigma_{a|b} = A_{aa}^{-1}$
$\mu_{a|b} = \Sigma_{a|b}(A_{aa}\mu_a - A_{ab}(x_b-\mu_b)) = \mu_a - A_{aa}^{-1}A_{ab}(x_b-\mu_b)$
By using Schur complement,
$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} M & -MBD^{-1} \\ -D^{-1}CMD^{-1} & D^{-1}CMBD^{-1} \end{pmatrix}, M = (A - BD^{-1}C)^{-1}$
$\Rightarrow A_{aa} = (\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1}\Sigma_{ab}\Sigma_{bb}^{-1}$
As a result :
$\mu_{a|b} = \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b - \mu_b)$
$\Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$
$\Rightarrow p(x_a \mid x_b) = N(x_{a|b} \mid \mu_{a|b}, \Sigma_{a|b}$

3.*Marginal Gaussian distribution*

The marginal distribution given by
$p(x_a) = \int p(x_a, x_b)dx_b$
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 :
$\int exp\{\frac{-1}{2}(x_b - A_{bb}^{-1})^T A_{bb}(x_b - A_{bb}^{-1}m)\}dx_b$
The remaining 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:

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