Machine Learning week 2

Kieu Son Tung

5. Oktober 2021

1 Prove that

1.1 First we have

$$\Sigma u_i = \lambda_i u_i$$

If we put all the eigenvectors into columns of the matrix U and all eigenvalues as the entries of a diagonal matrix A we can write the covariance matrix Σ as

$$\Sigma U = UA$$

We multiply both sides with U^{-1} and $U^{-1} = U^T$.So Σ will be represented as

$$\Sigma = UAU^T$$

If Σ is symmetry, then all eigenvectors are orthogonal $\implies UU^T = U^TU = I$. So we have

$$\Sigma = \sum_{i=1}^{D} \lambda_i u_i u_i^T$$

1.2 Prove

$$\Sigma = \sum_{i=1}^{D} \lambda_i u_i u_i^T => \Sigma^{-1} = \sum_{i=1}^{D} \frac{1}{\lambda_i} u_i u_i^T$$

Assume that eigenvalues and eigenvectors of Σ:

$$\begin{aligned} \Sigma u_1 &= \lambda_1 u_1 \\ \Sigma u_2 &= \lambda_2 u_2 \end{aligned} \tag{1}$$

Multiply both sided with u_2^T and u_1^T respectively, we have

$$\begin{aligned} u_2^T \Sigma u_1 &= \lambda_1 u_2^T u_1 \\ u_1^T \Sigma u_2 &= \lambda_2 u_1^T u_2 \end{aligned} \tag{2}$$

Subtract two equations we have

$$0 = (\lambda_1 - \lambda_2)\, u_1^T u_2$$

Because $u_2^T \Sigma u_1$ and $u_1^T \Sigma u_2$ are numbers so they are unchanged when we transpose them. We have

$$\left(u_1^T \Sigma u_2\right)^T = u_2^T \Sigma u_1$$

$\Sigma^T = \Sigma$ since it is symmetric. λ_1 and λ_2 are considered to be distinct so $\lambda_1 - \lambda_2 \neq 0$. So we have

$$u_1^T u_2 = 0$$

$\Longrightarrow$ u_1 and u_2 are orthogonal. So we have

$$\begin{aligned} \Sigma &= \sum_{i=1}^{D} \lambda_i u_i u_i^T \\ &= \sum_{i=1}^{D} \lambda_i I \\ \Sigma^{-1} &= \sum_{i=1}^{D} \frac{1}{\lambda_i} u_i u_i^T \\ &= \sum_{i=1}^{D} \frac{1}{\lambda_i} I \\ \Sigma\Sigma^{-1} &= \sum_{i=1}^{D} \lambda_i \cdot \sum_{i=1}^{D} \frac{1}{\lambda_i} \\ &= 1 \end{aligned} \tag{3}$$

2 Prove that multivariate Gaussian distribution is normalized

We have

$$\begin{aligned}\Delta^2 &= (x-\mu)^T \Sigma^{-1} (x-\mu) \\ &= \sum_{i=1}^{D} \frac{1}{\lambda_i} (x-\mu)^T (x-\mu) \\ &= \sum_{i=1}^{D} \frac{y_i^2}{\lambda_i}\end{aligned} \tag{4}$$

with $y_i = u_i^T (x-\mu)$ We also have $\left|\Sigma\right|^{\frac{1}{2}} = \prod_{i=1}^{D} \lambda_i^{\frac{1}{2}}$.
For a D-dimensional vector x, the multivariate Gaussian distribution takes the form

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

μ is a D-dimensional mean vector, Σ is a D x D covariance matrix, and $|\Sigma|$ denotes the determinant of Σ.
We replace $y_i = u_i^T (x-\mu)$ into the equation, we have

$$\begin{aligned}p(y) &= \frac{1}{(2\pi)^{\frac{D}{2}} (\prod_{i=1}^{D} \lambda_i)^{\frac{1}{2}}} exp(-\frac{1}{2}\sum_{i=1}^{D}\frac{y_i^2}{\lambda_i}) \\ &= \frac{1}{(2\pi)^{\frac{D}{2}} (\prod_{i=1}^{D} \lambda_i)^{\frac{1}{2}}} \prod_{i=1}^{D} exp(-\frac{1}{2}\frac{y_i^2}{\lambda_i}) \\ &= \prod_{j=1}^{D} \frac{1}{(2\pi\lambda_i)^{\frac{1}{2}}} exp(-\frac{y_j^2}{2\lambda_j}) \\ \implies \int_{-\infty}^{\infty} p(y)dy &= \prod_{j=1}^{D} \int_{-\infty}^{\infty} \frac{1}{(2\pi\lambda_i)^{\frac{1}{2}}} exp(-\frac{y_j^2}{2\lambda_j})dy_j \\ &= 1\end{aligned} \tag{5}$$

3 Calculate mean and variance of conditional distribution

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

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

So we have corresponding mean vector μ

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

And the covariance matrix Σ

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

We have

$$\begin{aligned} \frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) &= \frac{1}{2}(x-\mu)^T A(x-\mu) \\ &= \begin{pmatrix} x_a-\mu_a \\ x_b-\mu_b \end{pmatrix}^T \begin{pmatrix} A_{aa} & A_{ab} \\ A_{ba} & A_{bb} \end{pmatrix} \begin{pmatrix} x_a-\mu_a \\ x_b-\mu_b \end{pmatrix} \end{aligned} \tag{6}$$

And when $x = nx1, y = mx1, A = nxm$, we have

$$x^T A y = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}x_i y_j$$

So from (6)

$$\begin{aligned} (6) &= -\frac{1}{2}(x_a-\mu_a)^T A_{aa}(x_a-\mu_a) - \frac{1}{2}(x_a-\mu_a)^T A_{ab}(x_b-\mu_b) - \frac{1}{2}(x_b-\mu_b)^T A_{ba}(x_a-\mu_a) - \frac{1}{2}(x_b \\ &= -\frac{1}{2}x_a^T A_{aa}x_a + x_a^T(A_{aa}\mu_a - A_{ab}(x_b-\mu_b)) + const \end{aligned} \tag{7}$$

Compare with Gaussian distribution

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

So we have

$$\Sigma = A_{aa}^{-1}$$

We have

$$\begin{aligned}\Sigma^{-1}\mu &= A_{aa}\mu_a - A_{ab}(x_b - \mu_b)\\ \Sigma^{-1}\mu\Sigma &= (A_{aa}\mu_a - A_{ab}(x_b - \mu_b))A_{aa}^{-1}\\ \mu &= \mu_a - A_{aa}^{-1}A_{ab}(x_b - \mu_b)\end{aligned} \tag{8}$$

By using Schur complement

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

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

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

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