Machine Learning Homework Week 2

Tô Đức Anh
11196328

September 2021

1 Problems

1. Proof that covariance matrix are symmetric.
2. Proof that multivariate Gaussian distribution normalization.
3. Calculate marginal normal distribution.
4. Calculate conditional normal distribution.

2 Answers

2.1 Question 1

To proof that covariance matrix are symmetric. We first look at the formula:

$$Q = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^\top$$

We can see that, the result will be:

$$\begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & ..... & x_{1,n-1} & x_{1,n} \\ x_{2,1} & x_{2,2} & x_{2,3} & ..... & x_{2,n-1} & x_{2,n} \\ x_{3,1} & x_{3,2} & x_{3,3} & ..... & x_{3,n-1} & x_{3,n} \\ ...... & ...... & ...... & ...... & ...... & ...... \\ x_{n-1,1} & x_{n-12} & x_{n-1,3} & ..... & x_{n-1,n-1} & x_{n-1,n} \\ x_{n,1} & x_{n,2} & x_{n,3} & ..... & x_{n,n-1} & x_{n,n} \end{bmatrix}$$

Since cov(x,y) = cov(y,x) so that matrix Q is symmetric.

2.2 Question 2

To proof that the covariance are normalized we first look at the form of the distribution:

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

with μ is a D-dimensional mean vector, is a D $\times$ D covariance matrix, and $|\Sigma|$ denotes the determinant of Σ.
Let's set:

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

Break it all down we will have:

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

Let's prove that:

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

Recall the characteristic of transpose matrix:

$$\big(ABC\big)^T = C^T * B^T * A^T$$

We can see that:

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

The matrix Σ is symmetric as proven above so that (Σ^{-1}) and $(\Sigma^{-1})^T$ are identical. We also have that x^T is a 1*n matrix, (Σ^{-1}) is a n*n matrix, and μ is a n*1 matrix, so their product is a real number, a transpose of a real number is itself, so that:

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

Therefore:

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

Next, we have the eigenvalues and eigenvectors of Σ:

$$\Sigma u_i = \lambda_i u_i$$

For all i in (1,..,D)
Because Σ is a real, symmetric matrix, its eigenvalues will be real and its eigenvectors form an orthonormal set.
Let prove that:

$$\Sigma^{-1} = \sum_{n=1}^{D} \frac{1}{\lambda} u_i u_i^T$$

Let take

$$\sum_{n=1}^{D} \lambda u_i u_i^T * \sum_{n=1}^{D} \frac{1}{\lambda} u_i u_i^T$$

We can see that:

$$\lambda u_i u_i^T = \lambda_i * \begin{bmatrix} x_{1,1} * x_{1,1} & x1,2 * x2,1 & ..... & x_{1,n-1} * x_{n-1,1} & x_{1,n} * x_{n,1} \\ x_{2,1} * x_{1,2} & x2,2 * x2,2 & ..... & x_{2,n-1} * x_{n-1,2} & x_{2,n} * x_{n,2} \\ ...... & ...... & ...... & ...... & ...... \\ x_{n-1,1} * x_{1,n-1} & x_{2,n-1} * x_{n-1,2} & ..... & x_{n-1,n-1} * x_{n-1,n-1} & x_{n-1,n} * x_{n,n-1} \\ x_{n,1} * x_{1,n} & x_{2,2} * x_{2,2} & ..... & x_{n,n-1} * x_{n-1,n} & x_{n,n} * x_{n,n} \end{bmatrix}$$

Since the Σ covariance matrix is symmetric, so that eigenvalues will > 0 and eigenvectors will be orthogonal, and we have that for vectors v_1 and v_2 :

$$result = \begin{cases} 1 & v_i = v_j \\ 0 & v_i \neq v_j \end{cases}$$

So that

$$\lambda u_i u_i^T = \lambda * \begin{bmatrix} 1 & 0 & 0 & ... & 0 & 0 & 0 \\ 0 & 1 & 0 & ... & 0 & 0 & 0 \\ 0 & 0 & 1 & ... & 0 & 0 & 0 \\ ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & 1 & 0 & 0 \\ 0 & 0 & 0 & ... & 0 & 1 & 0 \\ 0 & 0 & 0 & ... & 0 & 0 & 1 \end{bmatrix}$$

equal to:

$$\begin{bmatrix} \lambda & 0 & 0 & ... & 0 & 0 & 0 \\ 0 & \lambda & 0 & ... & 0 & 0 & 0 \\ 0 & 0 & \lambda & ... & 0 & 0 & 0 \\ ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & \lambda & 0 & 0 \\ 0 & 0 & 0 & ... & 0 & \lambda & 0 \\ 0 & 0 & 0 & ... & 0 & 0 & \lambda \end{bmatrix}$$

Apply the same method to

$$X = \sum_{n=1}^{D} \frac{1}{\lambda} u_i u_i^T$$

We will have the value of matrix X:

$$\begin{bmatrix} \frac{1}{\lambda} & 0 & 0 & ... & 0 & 0 & 0 \\ 0 & \frac{1}{\lambda} & 0 & ... & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\lambda} & ... & 0 & 0 & 0 \\ ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & \frac{1}{\lambda} & 0 & 0 \\ 0 & 0 & 0 & ... & 0 & \frac{1}{\lambda} & 0 \\ 0 & 0 & 0 & ... & 0 & 0 & \frac{1}{\lambda} \end{bmatrix}$$

So

$$\sum_{n=1}^{D} \lambda u_i u_i^T * \sum_{n=1}^{D} \frac{1}{\lambda} u_i u_i^T$$

will equal:

$$\begin{bmatrix} \lambda & 0 & 0 & ... & 0 & 0 & 0 \\ 0 & \lambda & 0 & ... & 0 & 0 & 0 \\ 0 & 0 & \lambda & ... & 0 & 0 & 0 \\ ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & \lambda & 0 & 0 \\ 0 & 0 & 0 & ... & 0 & \lambda & 0 \\ 0 & 0 & 0 & ... & 0 & 0 & \lambda \end{bmatrix} \times \begin{bmatrix} \frac{1}{\lambda} & 0 & 0 & ... & 0 & 0 & 0 \\ 0 & \frac{1}{\lambda} & 0 & ... & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\lambda} & ... & 0 & 0 & 0 \\ ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & \frac{1}{\lambda} & 0 & 0 \\ 0 & 0 & 0 & ... & 0 & \frac{1}{\lambda} & 0 \\ 0 & 0 & 0 & ... & 0 & 0 & \frac{1}{\lambda} \end{bmatrix}$$

The result will be:

$$\lambda u_i u_i^T = \lambda * \begin{bmatrix} 1 & 0 & 0 & ... & 0 & 0 & 0 \\ 0 & 1 & 0 & ... & 0 & 0 & 0 \\ 0 & 0 & 1 & ... & 0 & 0 & 0 \\ ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & 1 & 0 & 0 \\ 0 & 0 & 0 & ... & 0 & 1 & 0 \\ 0 & 0 & 0 & ... & 0 & 0 & 1 \end{bmatrix}$$

or the Identity matrix, so that So that:

$$\Sigma^{-1} is \sum_{n=1}^{D} \frac{1}{\lambda} u_i u_i^T$$

Moving on:

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

with $y_i = u_i^T (x-\mu)$
We have that:

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

=>

$$\int_{-\infty}^{\infty} p(y) dy = \prod_{i=1}^{n} \int_{-\infty}^{\infty} \frac{1}{2\pi\lambda}^{1/2} e^{-\frac{{y_j}^2}{x\lambda_j}} dy_i$$

but we can see that:

$$\frac{1}{2\pi\lambda}^{1/2} e^{-\frac{{y_j}^2}{x\lambda_j}} dy_i$$

is a univariate and have the total probability = 1.
So that:

$$\prod_{i=1}^{n}\int_{-\infty}^{\infty}\frac{1}{2\pi\lambda}^{1/2} e^{-\frac{{y_j}^2}{x\lambda_j}} dy_i = \prod_{i=1}^{n} 1 = 1$$

2.3 Question 3

If you don't mind these two answers will be added later since i ran out of time, sorry

2.4 Question 4