1 Covariance matrix

Suppose $\mathbf{X}$ is a d-dimensional random vector (with d random variables), and $\boldsymbol{X_1}$,...,$\boldsymbol{X_n}$ is n independent copies of $\mathbf{X}$.

Write $\boldsymbol{X_i} = (X_i^1, \ldots, X_i^d)^T$, the subscript means the i_{th} copy, the superscript means the number of random variable (i.e. scala). Then we can combine all the $\boldsymbol{X_i}$ together as a new matrix, $\mathbb{X}$ (n by d).

$$\boldsymbol{X} = \begin{pmatrix} X^1 \\ X^2 \\ \ldots \\ X^d \end{pmatrix} \quad \mathbb{X} = \begin{pmatrix} \ldots & \boldsymbol{X_1^T} & \ldots \\ \ldots & \boldsymbol{X_2^T} & \ldots \\ \ldots & \boldsymbol{X_3^T} & \ldots \end{pmatrix} \tag{1}$$

Then we can know the covariance matrix, which means take two different scalas or coordinates (notice the superscript) from a vector and compute their covariance. For convenience, not use bold X again as before.

$$\Sigma = cov(X^i, X^j) \tag{2}$$
$$= \mathbb{E}(XX^T) - \mathbb{E}(X)\mathbb{E}(X)^T \tag{3}$$
$$= \mathbb{E}[(X - \mathbb{E}(X))(X - \mathbb{E}(X))^T] \tag{4}$$

When it comes to empirical data, we use average $\bar{X}$ to replace expectation[1] and use the empirical covariance matrix $\mathbf{S}$ to replace the Σ),

$$\mathbb{E}(X) = \begin{pmatrix} \mathbb{E}(X^1) \\ \vdots \\ \mathbb{E}(X^d) \end{pmatrix} \rightarrow \begin{pmatrix} \frac{\sum}{n} X_i^1 \\ \vdots \\ \frac{\sum}{n} X_i^d \end{pmatrix} \tag{5}$$
$$S = \frac{1}{n} \sum (X_i X_i^T) - \bar{X}\bar{X}^T \tag{6}$$

In order to eliminate the sum character, we multiply a $\mathbb{1}$ to replace the

[1] Here can be a little comfused because in we used subscript before but here we have X_i. This is because in theory, $E(X^1)$ is the expectation of random variable X^1, but empirically we sampled many times and calculate their average

average. $\mathbb{1} = (1, \ldots, 1)^T$

$$\bar{X} = \frac{1}{n}\sum X_i \qquad \mathbb{X} = \begin{bmatrix} \vdots & \vdots & \vdots \\ X_1 & X_2 & X_n \\ \vdots & \vdots & \vdots \end{bmatrix} \tag{7}$$

$$\frac{1}{n}\mathbb{X}^T\mathbb{1} = \frac{1}{n}\sum X_i = \bar{X} \tag{8}$$

And we can see that

$$M_i = \begin{bmatrix} 0 & \vdots & 0 & 0 \\ 0 & X_i & 0 & 0 \\ 0 & \vdots & 0 & 0 \end{bmatrix} \tag{9}$$

$$\mathbb{X}^T\mathbb{X} = \sum_i^n M_i M_i^T = \sum_i^n X_i X_i^T \tag{10}$$

$$\mathbb{X}^T = M_1 + M_2 + \cdots + M_n \tag{11}$$

Then in Eq.6 can be transformed into

$$S = \frac{1}{n}\mathbb{X}^T\mathbb{X} - \frac{1}{n^2}\mathbb{X}^T(\mathbb{1}\mathbb{1}^T)\mathbb{X} \tag{12}$$

$$= \frac{1}{n}\mathbb{X}^{\mathbb{T}}(I_d - \frac{1}{n}\mathbb{1}\mathbb{1}^T)\mathbb{X} \tag{13}$$

$$= \frac{1}{n}\mathbb{X}^T H\mathbb{X} \tag{14}$$

2 What's H?