Covariance & Correlation

Covariance matrix $\underline{\underline{\Sigma}}$

covariances: interaction between the dimensions

$$\underline{\underline{\Sigma}} = \begin{bmatrix} \sigma_0^2 & \sigma_{01} & \sigma_{02} & \\ \sigma_{01} & \sigma_1^2 & \sigma_{12} & \\ & & \sigma_2^2 & \\ & & & \ddots \end{bmatrix} \in \mathbb{R}^{K \times K}$$

variances: spread in each dimension

data matrix : $\underline{\underline{X}} = \begin{bmatrix} — & \underline{x}^{[0]} & — \\ — & \underline{x}^{[1]} & — \\ & \vdots & \\ — & \underline{x}^{[N-1]} & — \end{bmatrix} \in \mathbb{R}^{N \times K}$

N samples with K feature dimensions each

$$\underline{\mu}^{MLE} = \frac{1}{N} \sum_{i=0}^{N-1} \underline{x}^{[i]} \in \mathbb{R}^K$$

$$\underline{\underline{\Sigma}}^{MLE} = \frac{1}{N} \left(\underline{\underline{X}} - \underline{1}_N \underline{\mu}^T \right)^T \left(\underline{\underline{X}} - \underline{1}_N \underline{\mu}^T \right)$$

(also $\frac{1}{N-1}$ for the unbiased estimator)

⇝ covariances can only be interpreted with the variances

⇒ correlation: covariance of standardized data

Standardization: ① center data

② divide by standard deviation per dimension

$$\hat{\underline{x}}^{[0]} = \frac{\underline{x}^{[0]} - \underline{\mu}}{\underline{\sigma}}$$

– component-wise division

$$\underline{\sigma} = \begin{bmatrix} \sqrt{\sigma_0^2} \\ \sqrt{\sigma_1^2} \\ \vdots \end{bmatrix} \in \mathbb{R}^K$$

Sqrt diagonal of the covariance matrix $\underline{\underline{\Sigma}}$

$$\underline{\sigma} = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} \left(\underline{x}^{[i]} - \underline{\mu} \right)^2}$$

* all operations component-wise

or $\frac{1}{N-1}$ for the unbiased estimator

Standardized data matrix

$$\hat{\underline{\underline{X}}} = \frac{\underline{\underline{X}} - \underline{1}_N \underline{\mu}^T}{\underline{1}_N \underline{\sigma}^T}$$

component-wise

Correlation matrix

$$\underline{\underline{\rho}} = \frac{1}{N} \hat{\underline{\underline{X}}}^T \hat{\underline{\underline{X}}} \in \mathbb{R}^{K \times K}$$

NOT ! $\frac{1}{N-1}$ as this is already done with $\underline{\sigma}$

$$\underline{\underline{\rho}} = \begin{bmatrix} 1 & \rho_{01} & \rho_{02} & & & \\ \rho_{01} & 1 & \rho_{12} & & & \\ \cdots & & 1 & & & \\ & & & 1 & & \\ & & & & 1 & \\ & & & & & \ddots \end{bmatrix} \in \mathbb{R}^{K \times K}$$

$$\rho_{01} \in [-1, 1]$$

strong linear dependence

if $\rho_{01} = 0$, then no linear dependency

Correlation here: Pearson Correlation:

only linear correlation