

协方差与相关性分析

Allen.Huang

Covariance-协方差

期望值分别为 $E[X]$ 与 $E[Y]$ 的两个实随机变量 X 与 Y 之间的**协方差** $Cov(X, Y)$ 定义为：

$$\begin{aligned}Cov(X, Y) &= E[(X - E[X])(Y - E[Y])] \\&= E[XY] - 2E[Y]E[X] + E[X]E[Y] \\&= E[XY] - E[X]E[Y]\end{aligned}$$

性质：

如果两个变量的变化趋势一致，也就是说如果其中一个大于自身的期望值时另外一个也大于自身的期望值，那么两个变量之间的协方差就是正值；

如果两个变量的变化趋势相反，即其中一个变量大于自身的期望值时另外一个却小于自身的期望值，那么两个变量之间的协方差就是负值。

如何计算协方差

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

$$\text{cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m \text{cov}(X_i, Y_j)$$

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

相关性-Correlation

相关性公式：

$$r(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var[X] Var[Y]}}$$

相关性的取值空间

0.8~1.0 : 强相关

0.5~0.8 : 中相关

0.3~0.5 : 弱相关

0~0.3 : 极弱相关

0 : 不相关