数据科学基础

Foundations of Data Science

5.1 矩与几何性质

陈振宇
南京大学智能软件工程实验室
www.iselab.cn

形态度量-偏度

设一批数据为$X_1, X_2, \ldots, X_n$，则其偏度s^3为：

$$s^3 = \frac{B_3}{(B_2)^{1.5}} = \frac{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^3}{n}}{\left(\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}\right)^{1.5}}$$

X为随机变量，则其偏度为：

$$s^3 = \frac{E(X - EX)^3}{\sigma^3}$$

形态度量-峰度

设一批数据为$X_1, X_2,\ldots,X_n$，则其峰度s^4为：

$$s^4 = \frac{B_4}{(B_2)^2} = \frac{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^4}{n}}{\left(\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}\right)^2}$$

X为随机变量，则其峰度为：

$$s^4 = \frac{E(X - EX)^4}{\sigma^4}$$

混合矩

设一批数据为$(X_1, Y_1), \ldots, (X_n, Y_n)$, 则其k+l的混合矩为：

$$\frac{\sum_{i=1}^{n} X_i^{\mathrm{k}} \mathrm{Y}_i^{\mathrm{l}}}{n}$$

X, Y为随机变量，则其混合矩为：

$$E\left(X^{\mathrm{k}} Y^{\mathrm{l}}\right)$$

混合中心矩

设一批数据为$(X_1, Y_1), \ldots,(X_{n,}Y_n)$, 则其k+l的混合中心矩为：

$$\frac{\sum_{i=1}^{n}(X_i - \bar{X})^{\mathrm{k}}(Y_i - \bar{Y})^{\mathrm{l}}}{n}$$

X, Y为随机变量，则其混合中心矩为：

$$E[(X - EX)^k(Y - EY)^l]$$

协方差

设一批数据为$(X_1, Y_1), \ldots, (X_n, Y_n)$, 则其协方差为：

$$\mathrm{Cov(X,Y)} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(\mathrm{Y}_i - \bar{\mathrm{Y}})}{n}$$

X, Y为随机变量，则其协方差为：

$$\mathrm{Cov(X,Y)} = E[(X - EX)(Y - EY)]$$

相关系数

设一批数据为$(X_1,Y_1),\ldots,(X_n,Y_n)$，则其（皮尔逊）相关系数为：

$$\rho_{XY}=\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{(X_i-\bar{X})^2(Y_i-\bar{Y})^2}}$$

X, Y为随机变量，则其相关系数为：

$$\rho_{XY}=\frac{E[(X-EX)(Y-EY)]}{\sigma_X\sigma_Y}$$

相关系数

设一批数据为$(X_1,Y_1),\ldots,(X_n,Y_n)$, 则其（皮尔逊）相关系数为：

$$\rho_{XY}=\frac{\sum_{i=1}^{n}(X_i-\bar{X})(\mathrm{Y}_i-\bar{\mathrm{Y}})}{\sqrt{(X_i-\bar{X})^2(\mathrm{Y}_i-\bar{\mathrm{Y}})^2}}$$

X, Y为随机变量，则其相关系数为：

$$\rho_{XY}=\frac{E[(X-EX)(Y-EY)]}{\sigma_X\sigma_Y}$$

更多详情，请咨询课程助教。