Machine Learning I

Wrangling & EDA

Pavi Gupta

tgz5gw

Problem #6

1. $m(a+bx) = \frac{1}{N} \sum_{i=1}^{N} (a + bx_i)$

$= \frac{1}{N} \left(\sum_{i=1}^{N} a + b \sum_{i=1}^{N} x_i \right)$

$= a + b \left(\frac{1}{N} \sum_{i=1}^{N} x_i \right)$

$\therefore \boxed{m(a+bx) = a + b \times m(x)}$

2. $cov(x, x) = \frac{1}{N} \sum_{i=1}^{N} (x_i - m(x))(x_i - m(x))$

$= \frac{1}{N} \sum_{i=1}^{N} (x_i - m(x))^2$

$\boxed{cov(x, x) = s^2}$

3. $cov(x, a+bY) = \frac{1}{N} \sum_{i=1}^{N} (x_i - m(x)) \left[(a + by_i) - (a - bm(Y)) \right]$

$= \frac{1}{N} \sum_{i=1}^{N} (x - m(x)) (a + by_i - a - bm(Y))$

$= \frac{1}{N} \sum_{i=1}^{N} (x_i - m(x)) \left[b(y_i - m(Y)) \right]$

$\boxed{cov(X, a+bY) = b\, cov(X, Y)}$

4. $cov(a+bX, a+bY) = \frac{1}{N} \sum_{i=1}^{N} \left[(a+bx_i) - (a+bm(X)) \right] \left[(a+by_i) + (a+bm(Y)) \right]$

$= \frac{1}{N} \sum_{i=1}^{N} (a + bx_i - a - bm(X)) (\cancel{a} + by_i - \cancel{a} - bm(Y))$

$= \frac{1}{N} \sum_{i=1}^{N} b(x_i - m(X)) \cdot b(y_i - m(Y))$

$= \frac{1}{N} \sum_{i=1}^{N} b^2 (x_i - m(X))(y_i - m(Y))$

$\boxed{cov(a+bx, a+bY) = b^2 cov(X, Y)}$

5. $b > 0$, median of $X = med(x)$

↳ order is preserved

$\boxed{\text{Yes, } a + bX = a + b \cdot med(x)}$

$IQR = Q_3 - Q_1$

$IQR(a+bx) = (a + bQ_3) - (a + bQ_1)$

$= \cancel{a} + bQ_3 - \cancel{a} - bQ_1$

$\boxed{IQR(a+bx) \neq b(IQR(x))} = b(Q_3 - Q_1)$

$= b(IQR(x))$

6. ex.: $x = \{3, 5\}$

$m(x) = 4$

$m(x^2) = \frac{9 + 25}{2} = \frac{34}{2} = 17$

$(m(x))^2 = 4^2 = 16$

$\boxed{(m(x))^2 \neq m(x^2)}$

$m(\sqrt{x}) = \frac{\sqrt{3} + \sqrt{5}}{2} = \frac{3.968}{2} = 1.984$

$\sqrt{m(x)} = \sqrt{4} = 2$

$\boxed{m(\sqrt{x}) \neq \sqrt{m(x)}}$