Topic3. Correlation and *R*-Squared

Credit by https://win-vector.com/2011/11/21/correlation-and-r-squared/

$$L(f,y) = \sum_{i=1}^{n}(f_i - y_i)^2 \leq \sum_{i=1}^{n}(af_i + b - y_i)^2 = g(a,b)$$

Let assume *f* is the model that minimizes squared-error loss =>

- there is no scaling of *f* that will improve the fit (*a*=1)
- there is no shift of *f* that will improve the fit (*b*=0)

⇨ Since *g*(1,0) is the optimum, then the derivatives of *g* are zero here

$$\frac{\partial}{\partial_a} g(a,b) = \sum_{i=1}^{n} 2(f_i - y_i)f_i = 2f \cdot f - 2y \cdot y = 0$$

$f \cdot f = y \cdot f$ ... (1)

$$\frac{\partial}{\partial_b} g(a,b) = \sum_{i=1}^{n} 2(f_i - y_i) = 0$$

$\bar{f} = \bar{y}$

Shift the coordinate so $\bar{f} = 0$ & $\bar{y} = 0$

$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - f_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{n}(y_i - f_i)^2}{\sum_{i=1}^{n} y_i^2} = 1 - \frac{y \cdot y - 2y \cdot f + f \cdot f}{y \cdot y}$ ... (2)

Combing (1) & (2)

$R^2 = 1 - \frac{y \cdot y - f \cdot f}{y \cdot y} = \frac{f \cdot f}{y \cdot y}$ ... (3)

Correlation

$\rho = \frac{\frac{1}{n}\sum_{i=1}^{n} f_i y_i}{\sqrt{(\frac{1}{n}\sum_{i=1}^{n} f_i^2)(\frac{1}{n}\sum_{i=1}^{n} y_i^2)}} = \frac{f \cdot y}{\sqrt{(f \cdot f)(y \cdot y)}}$ ... (4)

Combing (1) & (4)

$\rho = \frac{f \cdot f}{\sqrt{(f \cdot f)(y \cdot y)}} = \sqrt{\frac{f \cdot f}{y \cdot y}}$ ... (5)

Combing (3) & (5)

$$R^2 = \rho^2$$