

회귀분석 (regression analysis)은 다중 벡터 독립변수 x 와 이에 대응하는 스칼라 종속변수 y 의 관계를 정량적으로 찾아내는 작업이다. \rightarrow 경의 유도

회귀분석은 결정적 모형 (deterministic model) 확률적 모형 (probabilistic model)이 있다.

결정적 모형은 독립변수 x 에 대해 대응하는 종속변수 y 와 가장 비슷한 값 \hat{y} 를 찾는 함수 $f(x)$ 를 찾는 과정이다.

$$\hat{y} = f(x) \approx y$$

선형 독립변수 x 와 이에 대응하는 종속변수 y 의 관계가 다음과 같은 선형 함수 $f(x)$ 이면, 이를 선형회귀 분석이라 한다.

$$\hat{y} = w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n = w_0 + w^T x$$

\downarrow 회귀식의 오차항 계수 (parameter) 계수 (coefficient)

- 선형회귀

이 선형회귀에서 회귀분석에 독립변수인 1 (bias)를 추가하여 데이터 벡터의 x 와의 내적은

가능하게 표현할 수 있다.

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & & & \\ \vdots & & & \\ x_{n1} & & & x_{nn} \end{bmatrix} \rightarrow X_a = \begin{bmatrix} 1 & x_{11} & \dots & x_{1n} \\ \vdots & x_{21} & & \\ \vdots & & & \\ \vdots & x_{n1} & \dots & x_{nn} \end{bmatrix}$$

상대량 행렬에 대해 모든 원소가 1인 행이 추가됨.

||

$$f(x) = w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n = \begin{bmatrix} 1 & x_1 & x_2 & x_3 & \dots & x_n \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{bmatrix} = x_a^T w_a = w_a^T x_n$$

- OLS (Ordinary Least Square)

가장 적합한 결정적 회귀분석, 잔차 제곱합 (RSS: Residual sum squares)을 최소화하는 가장 벡터

해결 방법을 구하는 방법

- 경의는 선형 -