

Part.02
회귀분석

| 회귀계수의 의미

FASTCAMPUS
ONLINE

머신러닝과 데이터분석 A-Z

강사. 이경택

I 회귀계수의 의미

회귀 계수의 해석

- $\hat{\beta}_1$ 의 해석 - X1이 1단위 증가할 때마다 y가 $\hat{\beta}_1$ 만큼 증가한다.
- 예시) radio광고 예산과 매출 간의 관계
- Radio광고 예산이 1증가 할 때 마다 매출은 0.2단위 만큼 증가한다. 그때의 유의성은 매우 높다.
- Radio광고 예산이 35 단위일 때 예상 매출액은 $9.312 + 0.203 \times 35 = 16.42$ 단위이다

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

$$\hat{Y} = 9.312 + 0.203X$$

회귀 계수 회귀 계수의 표준 오차 회귀 계수의 의성을 판단하는 통계치

Simple regression of sales on radio

	Coefficient	Std. error	t-statistic	p-value
Intercept	9.312	0.563	16.54	< 0.0001
radio	0.203	0.020	9.92	< 0.0001

I 회귀계수의 의미

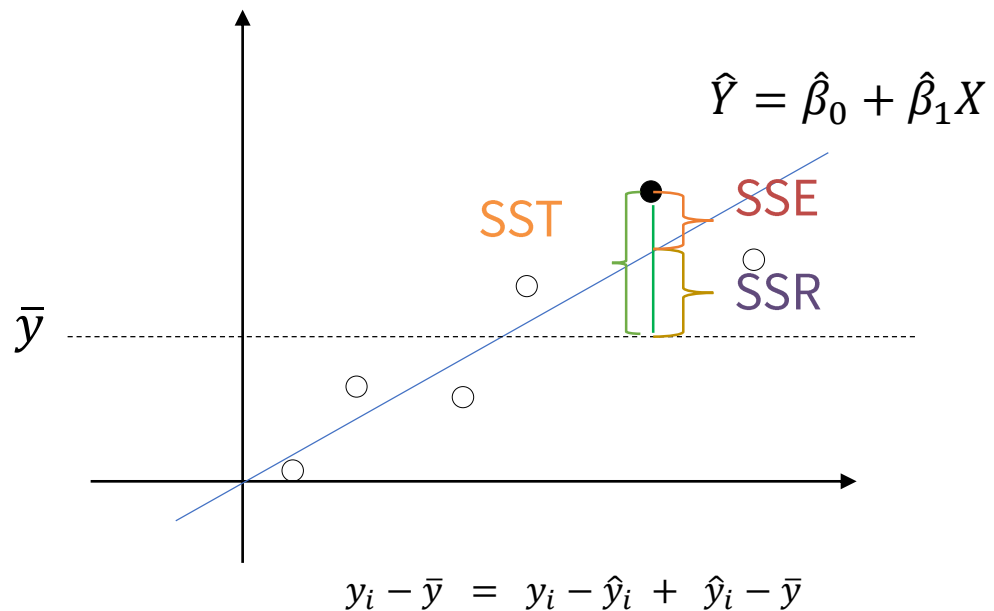
- 선형 회귀의 정확도 평가
 - 선형회귀는 잔차의 제곱합(SSE : Error sum of squares)를 최소화 하는 방법으로 회귀 계수를 추정
 - 즉, SSE가 작으면 작을수록 좋은 모델이라고 볼 수 있음
 - MSE(Mean Squared Error)는 SSE를 표준화한 개념

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$MSE = \frac{1}{n-2} SSE$$

I 회귀계수의 의미

■ 선형 회귀의 정확도 평가



$$\underbrace{\sum_{i=1}^n (y_i - \bar{y})^2}_{\text{SST (Total sum of squares)}} = \underbrace{\sum_{i=1}^n (y_i - \hat{y}_i)^2}_{\text{SSE (Error sum of squares)}} + \underbrace{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}_{\text{SSR (Regression sum of squares)}}$$

$$\begin{aligned}\Sigma(y_i - \bar{y})^2 &= \Sigma(y_i - \hat{y}_i)^2 + \Sigma(\hat{y}_i - \bar{y})^2 + 2(\Sigma e_i)(\Sigma \hat{y}_i - \bar{y}) \\ &= \Sigma(y_i - \hat{y}_i)^2 + \Sigma(\hat{y}_i - \bar{y})^2\end{aligned}$$

※ 회귀 계수 추정

$$\frac{\partial L}{\partial \beta_0} = -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\frac{\partial L}{\partial \beta_1} = -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) x_i = 0$$

Source of Variation	Sum of Squares	Degree of freedom	Mean Square
Regression	SSR	1	SSR
Error	SSE	N-2	MSE
Total	SST	N-1	

I 회귀계수의 의미

■ 선형 회귀의 정확도 평가

$$\underbrace{\sum_{i=1}^n (y_i - \bar{y})^2}_{SST} = \underbrace{\sum_{i=1}^n (y_i - \hat{y}_i)^2}_{SSE} + \underbrace{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}_{SSR}$$

- Y의 총 변동은 회귀직선으로 설명 불가능 한 변동과 회귀직선으로 설명 가능한 변동으로 이루어져 있음
- R^2 는 RSE의 단점을 보완한 평가지표로 0~1의 범위를 가짐
- R^2 은 설명력으로 입력 변수인 X로 설명할 수 있는 Y의 변동을 의미
- R^2 이 1에 가까울 수록 선형회귀 모형의 설명력이 높다는 것을 뜻함

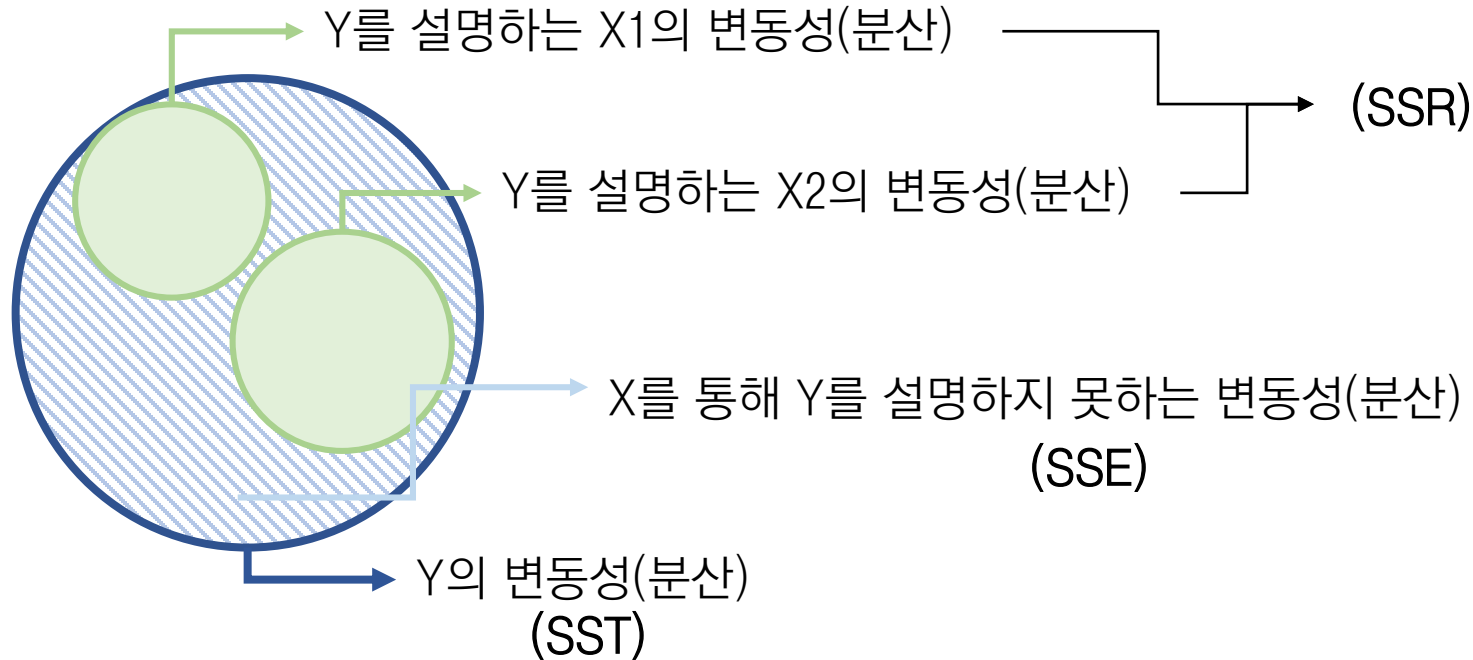
$$R^2 = \frac{SST - SSR}{SST} = 1 - \frac{SSE}{SST} = \frac{SSR}{SST} \quad \text{where } SST = \sum_{i=1}^n (y_i - \bar{y})^2$$

입력 변수로 설명할 수 없는 변동 비율

I 회귀계수의 의미

■ 선형 회귀의 정확도 평가

- 회귀 분석은 결국 Y의 변동성을 얼마나 독립변수가 잘 설명하느냐가 중요
- 변수가 여러 개일 때 각각 Y를 설명하는 변동성이 크면 좋은 변수 -> p-value자연스레 낮아짐



$$R^2 = \frac{SSR}{SST}$$

Part.02
회귀분석

| 회귀계수에 대한 검정

FASTCAMPUS
ONLINE

머신러닝과 데이터분석 A-Z

강사. 이경택