

Part.02
회귀분석

I 다중 선형 회귀분석

FASTCAMPUS
ONLINE

머신러닝과 데이터분석 A-Z

강사. 이경택

I 다중 선형 회귀분석

- 단순 선형 회귀분석 : 변수가 1개인 경우

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

- 다중 선형 회귀분석 : 변수가 여러개인 경우

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_p X_p$$

- 다중 선형 회귀계수 추정

- 회귀계수를 추정하는것은 단순선형회귀분석과 동일하게 SSE 를 최소화 하는 방향으로 추정

$$SSE = \sum_{i=1}^n e_i^2 = e_1^2 + e_2^2 + \cdots + e_n^2$$

- 위 식을 각각의 변수에 대해 편미분 하여 회귀계수를 추정

$$\frac{\partial L}{\partial \beta_0} = -2 \sum_{i=1}^n (y_i - \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}) = 0$$

$$\frac{\partial L}{\partial \beta_1} = -2 \sum_{i=1}^n (y_i - \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}) x_{1i} = 0$$

$$\vdots \qquad \qquad \qquad \vdots$$

$$\frac{\partial L}{\partial \beta_p} = -2 \sum_{i=1}^n (y_i - \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}) x_{pi} = 0$$

I 다중 선형 회귀분석

■ 다중 선형 회귀계수 추정(행렬)

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_k x_{1k} + \varepsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_k x_{2k} + \varepsilon_2 \\ &\vdots \\ y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_k x_{nk} + \varepsilon_n \end{aligned}$$



$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$



$$y = X\beta + \varepsilon \quad \varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$$

$$\begin{aligned} \sum_{i=1}^n \varepsilon_i^2 &= \varepsilon' \varepsilon = (y - X\beta)'(y - X\beta) \\ &= y'y + \beta' X' X \beta - 2\beta' X'y \end{aligned}$$

$$\frac{\partial L}{\partial \beta} = 2X'X\beta - 2X'y = 0$$

$$\Rightarrow X'X\beta = X'y$$

$$\Rightarrow \beta = (X'X)^{-1}X'y$$

I 다중 선형 회귀분석

■ 다중 선형 회귀계수 해석

- 광고미디어 예에 다중선형회귀를 적용하면 아래와 같음
- 신문 광고의 경우 단순선형회귀에서는 출력 변수인 매출과 연관이 있었지만, 다중선형회귀에서는 p-value가 0.86으로 높아 매출에 유의미한 영향을 미치지 못함
- 신문 광고의 단순선형회귀에서 TV 광고와 라디오 광고의 영향력을 무시했기 때문에 생긴 결과

$$\hat{Y} = \hat{\beta}_0 + \overset{\text{TV}}{\hat{\beta}_1 X_1} + \overset{\text{radio}}{\hat{\beta}_2 X_2} + \overset{\text{newspaper}}{\hat{\beta}_3 X_3}$$

	Coefficient	Std. error	t-statistic	p-value
Intercept	2.939	0.3119	9.42	< 0.0001
TV	<u>0.046</u>	0.0014	32.81	< 0.0001
radio	<u>0.189</u>	0.0086	21.89	< 0.0001
newspaper	<u>-0.001</u>	0.0059	-0.18	0.8599

Part.02
회귀분석

I 다중 선형 회귀분석 (회귀 계수/모델에 대한 검정)

FASTCAMPUS
ONLINE

머신러닝과 데이터분석 A-Z

강사. 이경택