

Part.02 회귀분석

I다중선형회귀분석......

FASTCAMPUS ONLINE

머신러닝과 데이터분석 A-Z

강사. 이경택

1다중 선형 회귀분석

■ 단순 선형 회귀분석 : 변수가 1개인 경우

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

■ 다중 선형 회귀분석: 변수가 여러개인 경우

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

- 다중 선형 회귀계수 추정
 - 회귀계수를 추정하는것은 단순선형회귀분석과 동일하게 SSE를 최소화 하는 방향으로 추정

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

■ 위 식을 각각의 변수에 대해 편미분 하여 회귀계수를 추정

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

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

$$\vdots \qquad \vdots$$

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



I 다중 선형 회귀분석

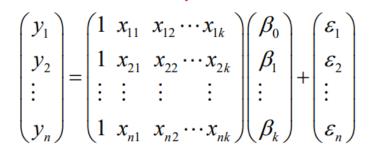
• 다중 선형 회귀계수 추정(행렬)

$$y_{1} = \beta_{0} + \beta_{1}x_{11} + \beta_{2}x_{12} + \dots + \beta_{k}x_{1k} + \varepsilon_{1}$$

$$y_{2} = \beta_{0} + \beta_{1}x_{21} + \beta_{2}x_{22} + \dots + \beta_{k}x_{2k} + \varepsilon_{2}$$

$$\vdots$$

$$y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_k x_{nk} + \varepsilon_n$$





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

$$\sum_{i=1}^{n} \varepsilon_{i}^{2} = \varepsilon' \varepsilon = (y - X\beta)'(y - X\beta)$$
$$= y'y + \beta' X' X\beta - 2\beta' X' y$$

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

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

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

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

$$\hat{Y} = \hat{\beta}_0 + \underline{\hat{\beta}_1} X_1 + \underline{\hat{\beta}_2} X_2 + \underline{\hat{\beta}_3} X_3$$

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

FAST CAMPUS ONLINE 이경택 강사.

Fast campus



Part.02 회귀분석

l다중 선형 회귀분석

(회귀계수/모델에 대한 검정)

FASTCAMPUS ONLINE

머신러닝과 데이터분석 A-Z

강사. 이경택