

선형 회귀에서 비선형성을 포착하는 방법.

과차광 도입. $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2$

X_1^2 와 같은 새 변수 X_2 와 같이 포함하는.

변수의 갯수가 늘어날수록 모델의

유연성이 증가함.

범주형 변수 X 를 처리하는 방법.

성별이나 지역과 같은 범주형 변수는

수치가 아니기 때문에 수치화 해야 함.

0과 1과 같은 형태로 인코딩할 수 있음.

ex) $X_1 = \text{cat or dog. (cat}=1, \text{dog}=0)$

$$y_i = \beta_0 + \beta_1 X_{i1} + \epsilon_i \quad \begin{cases} \text{cat} : \beta_0 + \beta_1 + \epsilon_i \\ \text{dog} : \beta_0 + \epsilon_i \end{cases}$$

이전 레벨이 아닌 경우엔 다른 변수가 필요함.

ex) $X_1 = \text{cat or dog or horse}$

$$X_{i1} = \begin{cases} 1 & \text{if } i\text{th animal is cat} \\ 0 & \text{" not cat} \end{cases}$$

$$X_{i2} = \begin{cases} 1 & \text{" dog} \\ 0 & \text{" not dog} \end{cases}$$

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i$$

$$\begin{cases} \text{cat} : \beta_0 + \beta_1 + \epsilon_i \\ \text{dog} : \beta_0 + \beta_2 + \epsilon_i \\ \text{horse} : \beta_0 + \epsilon_i \end{cases}$$

k 개의 레벨을 가진 범주형 변수는 $k-1$ 개의

0/1 변수를 필요로 함.