

11

개념 이해

[선형회귀 분석 개념]

전통적(통계적) 회귀분석(GLM)과 기본적인 개념은 똑같습니다.

- 종속변수는 연속형 독립변수도 연속형이어야 합니다.

00:00

11 회귀분석

11.1 변수 유형별 분석 방법

지도학습

분류 : Predict between several categories

회귀 : Predict Value

비지도학습

군집분석 : Discover structure

차원 축소 : Dimension reduction

11.2 회귀분석 개념

- 하나 혹은 그 이상의 원인이 종속변수에 미치는 영향을 추적하여 식으로 표현하는 통계기법
- 변수들 사이의 인과관계를 밝히고 오차를 최소화하여 관심있는 변수를 예측하거나 추론하기 위해 사용하는 분석 방법

[머신러닝을 이용한 선형회귀 분석]

선형 모델을 머신러닝 기법으로 사용하면 어떠한 것이 다를까요?

- 개념은 같지만, 내부적으로 머신러닝에서는 다중공선성을 제거해줍니다.

02:27

11.1 머신러닝 기법을 이용한 방법

```
from sklearn.linear_model import LinearRegression
```

- LinearRegression은 계수 $w = (w_1, \dots, w_p)$ 를 사용하여 선형 모델을 피팅합니다.
- 선형 모델은 X_{train} 으로 선형 근사에 의해 예측된 선형 회귀식과 y_{train} 사이의 잔여 제곱합을 최소화합니다.

`LinearRegression()`: 선형회귀분석 모델을 생성하는 함수
`fit()`: 선형 회귀 모델에 필요한 두 가지 변수(x, y)를 전달

- 설명변수의 차원을 조정해야 함
- R의 `summary()`와 같은 기능이 없음
- 기술기와 절편을 따로 호출해서 얻을 수 있음
- 잔차를 자동으로 계산해주는 기능이 없음
- `model.coef_`로 회귀계수 추출가능

11.2.2 Ridge, Lasso, ElasticNet

- Ridge

Ridge 회귀는 L2 정규화로 계수 크기에 패널티를 부과하여 일반 최소 제곱의 일부 문제를 해결합니다.
 매개변수: `alpha = 1.0`

