

Machine learning

Hyeonho Lee

2018년 11월 6일

Contents

1 선형회귀	2
1.1 단순선형회귀	2
1.3 회귀모델에서 다른 고려할 사항	2
1.4 마케팅 플랜	2
1.5 선형회귀와 KNN의 비교	2
2 선형모델 선택 및 Regularization	3
subset(부분집합) 선택	3
Shrinkage 방법	3
차원축소 방법	3
고차원의 고려	3
3 선형성을 넘어서 (비선형성)	4
다항식회귀	4
계단함수	4
기저함수	4
회귀 스플라인	4
평활 스플라인	4
국소회귀	4
일반화방법모델	4
4 트리 기반의 방법	5
의사결정트리의 기초	5
배깅, 랜덤 포레스트, 부스팅	5

1 선형회귀

1. 선형회귀는 양적 반응변수를 예측하는 유용한 도구이다.
2. 중요한 질문들...
 - 1) X와 Y사이에 상관관계가 있는가
 - 2) X와 Y사이에 얼마나 강한 상관관계가 있는가
 - 3) 여러 X들 중 Y에 기여하는 X는?
 - 4) Y에 대한 각 X 효과를 얼마나 정확하게 추정할 수 있는가
 - 5) 미래의 Y에 대해 얼마나 정확하게 예측할 수 있는가
 - 6) 상관관계는 선형인가
 - 7) X들 사이에 시너지 효과가 있는가(상호작용 항)

1.1 단순선형회귀

1. 단순선형회귀는 매우 간단한 기법으로, 하나의 설명변수 X에 기초하여 양적 반응변수 Y를 예측한다. 이 기법은 X와 Y 사이에 선형적 상관관계가 있다고 가정한다. 수학적으로 선형적 상관관계는 다음과 같이 나타낸다.

$$Y \approx \beta_0 + \beta_1 X + \varepsilon$$

2. 계수 추정
 - 1) 실제로 β_0 와 β_1 은 알려져 있지 않다. 그러므로 $Y \approx \beta_0 + \beta_1 X + \varepsilon$ 을 사용하여 예측하기 전에 데이터를 이용하여 계수를 추정해야 한다.
 - 2) n의 데이터 포인트의 개수라고 할 때, n개의 데이터 포인트에 가능한 한 가깝게 되도록 하는 절편 $\hat{\beta}_0$ 와 기울기 $\hat{\beta}_1$ 을 찾고자 한다.
 - 3) 가까움(closeness)을 측정하는 방법은 여러 가지가 있으나, 대표적으로는 최소제곱 기준을 최소화하는 것이다.
3. 계수 추정값의 정확도 평가
 - 1)
 - 2)
 - 3)
4. 모델의 정확도 평가
 - 1) 잔차표준오차(RSE)
 - 2) R^2 통계량 ## 1.2 다중선형회귀

1.3 회귀모델에서 다른 고려할 사항

1.4 마케팅 플랜

1.5 선형회귀와 KNN의 비교

2 선형모델 선택 및 Regularization

subset(부분집합) 선택

Shrinkage 방법

차원축소 방법

고차원의 고려

3 선형성을 넘어서 (비선형성)

다항식회귀

계단함수

기저함수

회귀 스플라인

평활 스플라인

국소회귀

일반화가법모델

4 트리 기반의 방법

의사결정트리의 기초

배깅, 랜덤 포레스트, 부스팅