머신러닝 프로그래밍 과제

20235127김민주

첫 번째 강의와 Simple Linear Regression(선형 회귀) 두 번째 강의(Hypothesis Testing and Confidence Intervals)와 세 번째 강의인(Multiple Linear Regression)에서 머신러닝을 대통령 선거에서 이용하였다. 그 머신러닝은 매우 정확하게 예측했다고 했다. 통계를 사용하고, 신중하게 샘플링 된 데이터를 사용하고 다양한 장소들과 신중한 분석을 통해선거 예측을 매우 정확하게 수행하였다고 한다. 강의자들이 살펴본 영상에서의 첫 번째 데이터 세트인 전립선 암에 대해 얘기한다. 97명의 남성으로 구성된 비교적 작은 데이터 세트를 실제로 스탠포드 대학에서 전립선 암을 앓고 있는 97명의 남성들을 80년대 후반, 의사 스테이미 박사가 연구하였습니다. 암에 관련해서 측정하는데 혈액에서 측정한 측정값, 사용할 측정값 암의 크기와 암의 심각성을 나타낸다. 그리고 이게 산점도 행렬이다. 로그의 무게를 재며 로그를 분석할 수 있습니다.

위 식에서 는 다른 모든 예측 변수를 고정한 상태에서 의 한 단위 증가가 Y에 미치는 평균 효과로 해석된다. 광고 예제에서 모델은 다음과 같습니다.

Sales = 가 나옵니다.

* 이상적인 시나리오는 예측 변수들이 상관관계가 없는 균형 잡힌 설계일 때 입니다.

- 각 계수는 개별적으로 추정하고 테스트할 수 있습니다.  
- "의 단위 변화는 Y의 변화와 관련이 있지만, 다른 모든 변수는 고정되어 있다"와 같은 해석이 가능합니다.

* 예측 변수 간의 상관관계는 문제를 일으킵니다.

- 모든 계수의 분산은 때때로 극적으로 증가하는 경향이 있습니다  
- 해석은 위험해 집니다 - 가 변하면 다른 모든 것이 변합니다.  
관측 데이터에 대한 인과관계 주장은 피해야 합니다.

* 추정치 이 주어지면 다음 공식을 사용하여 예측할 수 있습니다.
* 우리는 제곱 잔차의 합을 최소화하는 값으로 를 추정합니다.

RSS =

이 작업은 표준 통계 소프트웨어를 사용하여 수행됩니다. RSS를 최소화하는 값인 은 다중 최소 제곱 회귀 계수 추정치 임을 알 수 있었다.

기본적으로 모든 모델이 잘못되었지만 일부는 유용합니다

"복잡한 와이어템이 방해받을 때 어떤 일이 일어날지 알아낼 수 있는 유일한 방법은 단순히 수동적으로 관찰하는 것이 아니라 시스템을 방해하는 것입니다.