ISL
6.1.1

AI명예학회

SKHU

6. Linear Model Selection and Regularization

- 선형모델은 추론에서 뚜렷한 장점이 있으며, 설명력의 관점에서 매우 뛰어나다.
- 비선형적인 모델을 배우기 전에, 선형 모델을 개선하는 방법을 알아보자
- 최소제곱법 대신 다른 최적화 방식을 사용해서 개선

왜 최소제곱법 대신 다른 방법을 사용해서 최적화를 하는 것이 개선에 도움이 될까?

AI명예학회
SKHU

6. Linear Model Selection and Regularization

- **Prediction Acuracy** :

n > p : low bias의 경향, 좋은 test 성능

p > n :무수히 많은 least squares 해 존재. train loss는 0이 되지만, 분산이 커져 test loss는 무척 커지게 된다.
-> 추정된 계수들을 제약하거나 축소함

- **Model Interpretability** :

회귀 분석에 사용되는 변수들이 response를 설명하는데 전혀 도움이 되지 않는 경우 모델을 불필요하게 복잡하게 만들곤 한다.

이런 변수들을 제거함으로서, 더 해석하기 쉬운 모델을 만들 수 있다.

AI명예학회
SKHU

6. Linear Model Selection and Regularization

- 부분집합 선택: 반응변수와 관련이 있다고 생각되는 p개의 예측변수들 중 일부 부분집합을 선별
- Regularization: 계수를 0에 가깝게 축소함으로서 필요없는 계수를 버리고 분산을 축소
- 차원 축소: p의 차원을 M으로 투영해 변수의 개수를 줄임

AI명예학회
SKHU

6.1.1.Best Subset Selection

- 모든 예측 변수 집합의 부분 집합 하나하나에 least squares를 적용
- 어떤 예측 변수들에 대한 부분집합 모델이 최고의 결과를 내는지 분석

Algorithm 6.1 *Best subset selection*

1. Let $\mathcal{M}_0$ denote the *null model*, which contains no predictors. This model simply predicts the sample mean for each observation.

2. For $k = 1, 2, \ldots p$:

   (a) Fit all $\binom{p}{k}$ models that contain exactly k predictors.

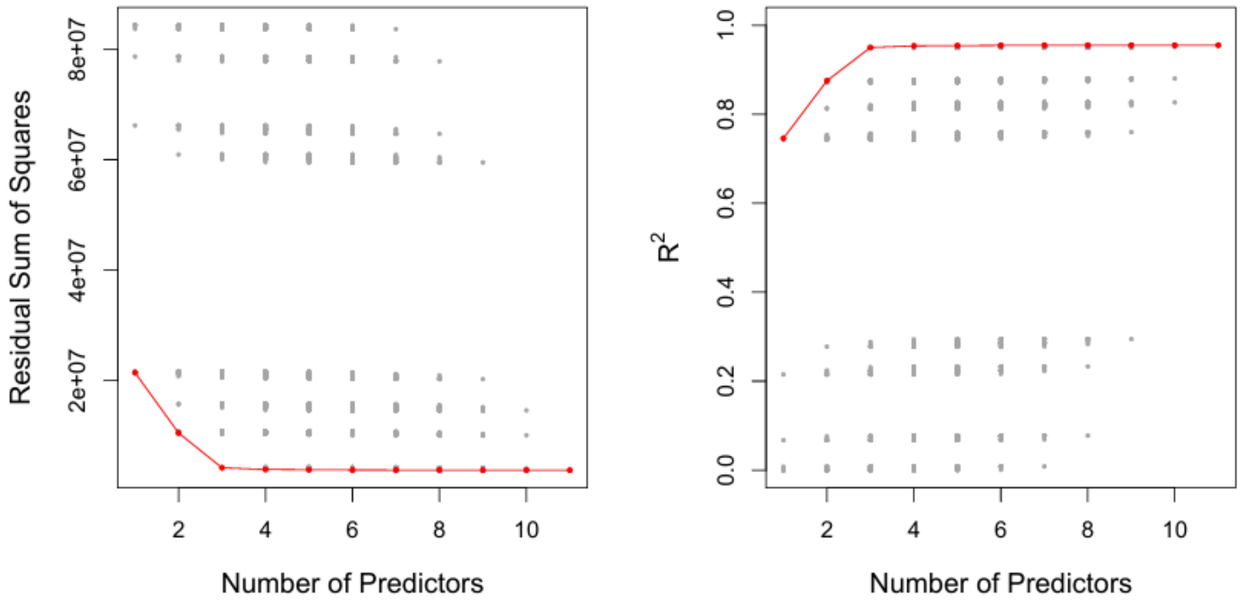
   (b) Pick the best among these $\binom{p}{k}$ models, and call it $\mathcal{M}_k$. Here *best* is defined as having the smallest RSS, or equivalently largest R^2.

3. Select a single best model from among $\mathcal{M}_0, \ldots, \mathcal{M}_p$ using using the prediction error on a validation set, C_p (AIC), BIC, or adjusted R^2. Or use the cross-validation method.

AI명예학회
SKHU

6.1.1.Best Subset Selection

- least squares의 특성 상 포함하는 변수가 늘어날수록 RSS는 줄어들고, R2는 커짐
- 이 문제를 극복하기 위해 여러 검증 방식을 사용



AI명예학회
SKHU

6.1.1.Best Subset Selection

deviance = -2 × log(likelihood)

- logistic regression은 RSS 대신 deviance 사용
- deviance가 작을수록 더 좋은 모델

이 Best Subset Selection은 2^p의 모델에 대해 전부 연산해야 하기 때문에, p가 늘어날수록 지수적으로 연산량이 증가함.
비효율적이라 잘 사용하지 않음.

AI명예학회
SKHU