

회귀 트리와 모델 트리 이해

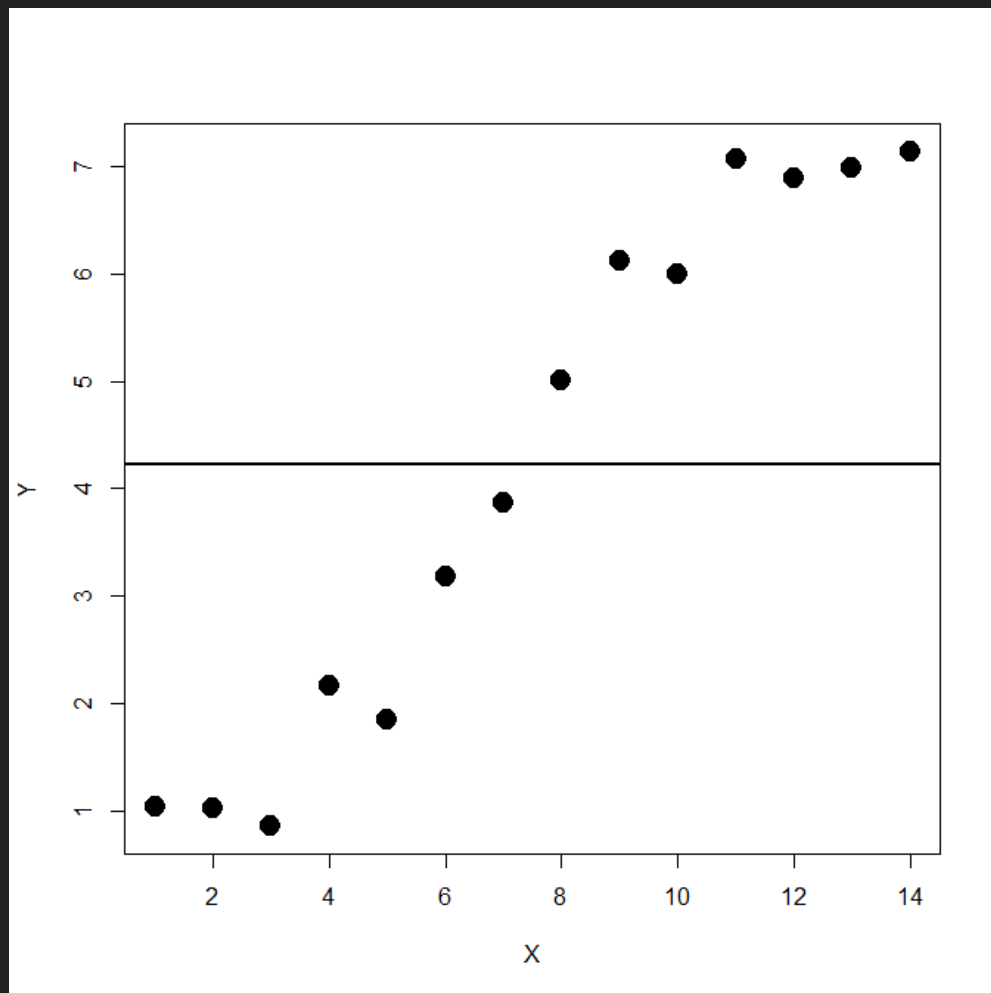
- CART (Classification and Regression Tree)
- 회귀 트리 (Regression Tree)
- 모델 트리 (Model Tree)

회귀 트리

- 1980년대 소개
- 수치 반응변수의 예측
- 선형 회귀 기법을 사용하지 않음

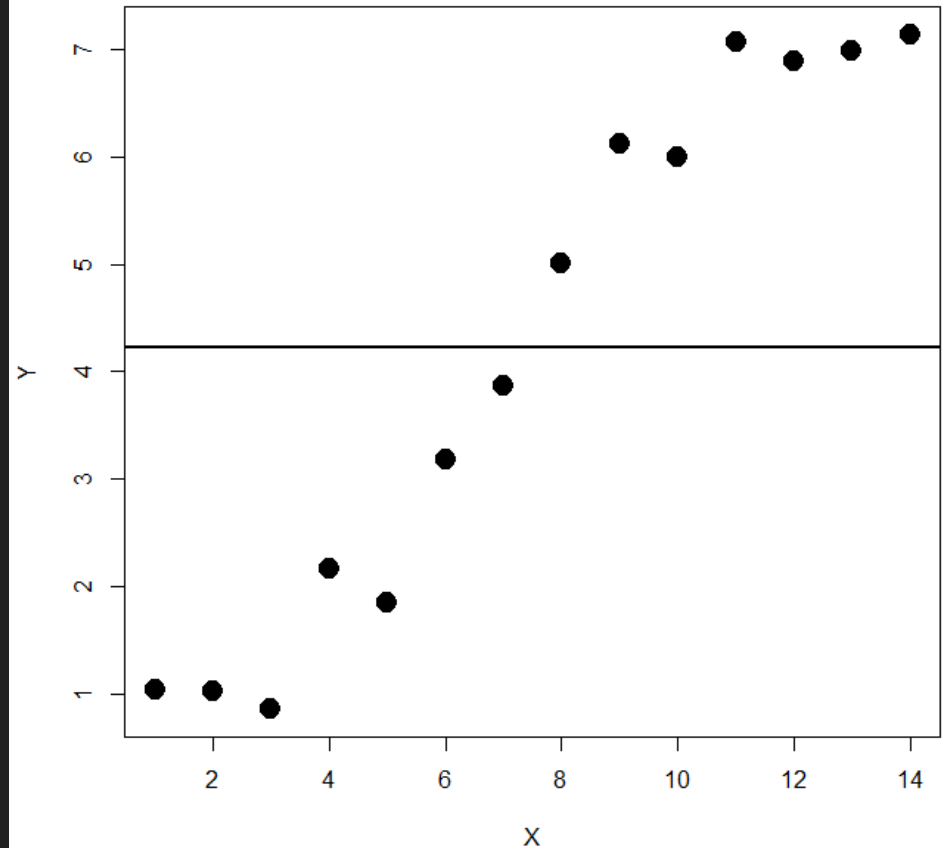
회귀 트리 (cont'd)

- X 를 고려하지 않고 Y 예측?
- 표본평균 이용
- $\hat{E}(Y)=4.22765$



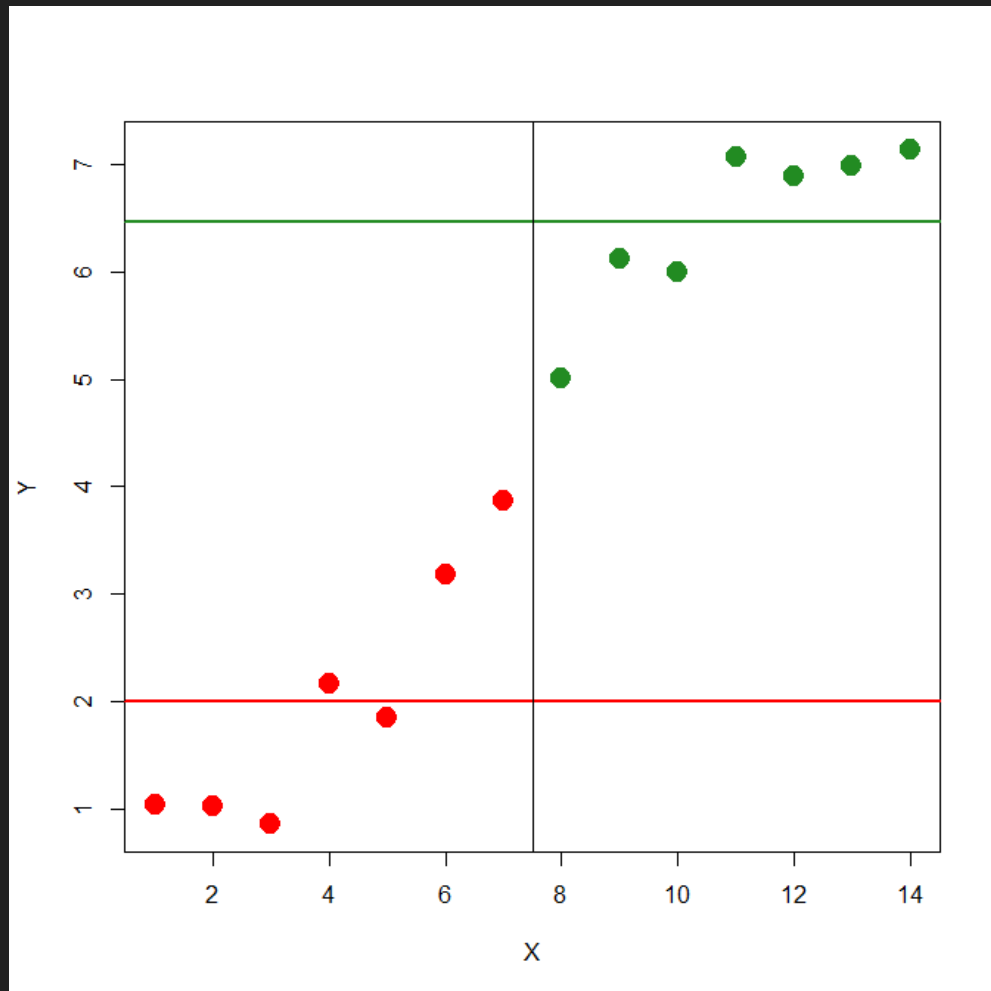
회귀 트리 (cont'd)

- 회귀분석을 꼭 선형 모델로 생각할 필요 없음
- “X가 주어졌을 때의 Y의 조건부 기댓값의 추정”



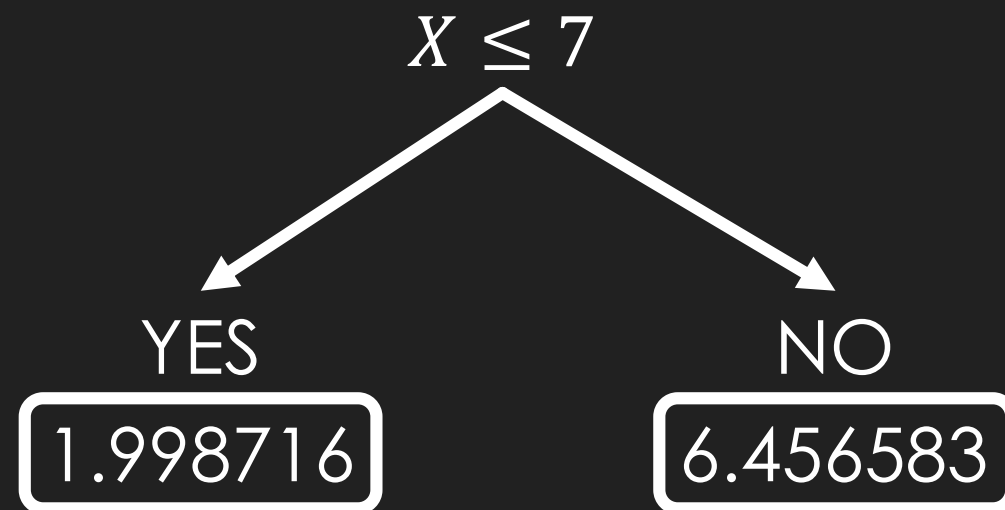
회귀 트리 (cont'd)

- X 를 이용하여 데이터 분할
- 각 분할된 표본에서 Y 예측
- $\hat{E}(Y|X \leq 7) = 1.998716$
- $\hat{E}(Y|X > 7) = 6.456583$



회귀 트리 (cont'd)

○ 트리 다이어그램



회귀 트리 (cont'd)

- 나눔의 기준?
- 표준편차축소
(Standard Deviation Reduction)

$$SDR = sd(T) - \sum_i \frac{|T_i|}{|T|} \times sd(T_i)$$

T : Y 측정값 전체집합
 T_i : T 의 파티션

회귀 트리 (cont'd)

- 가장 Y 를 예측하기에 좋은 X 변수부터 분할한다. (SDR \uparrow)

$$SDR = sd(T) - \sum_i \frac{|T_i|}{|T|} \times sd(T_i)$$

T : Y 측정값 전체집합
 T_i : T 의 파티션

모델 트리

- 분할된 표본에서, 각각을 선형회귀 적합

