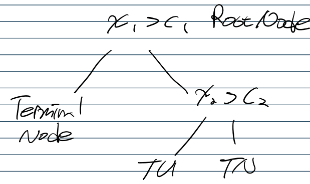
Week 4

Sunday, August 10, 2025 4:18 PM

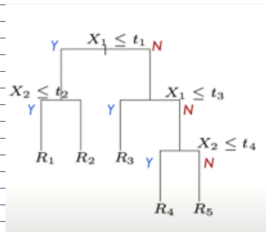
1. Decision Tree

데이터 2개 이상의 부분집합으로 균일해지도록 분할.

- 분류 : 비슷한 범주 관측치 끼리
- 예측 : 비슷한 수의 관측치 끼리



2. Regression Tree



C_m : Regression Tree로 부터 예측한 R_m 부분 예측값 (mean ..)

$\hat{f}(x) = \sum C_m I\{(X_1, X_2) \in R_m\}$

I : Indicator func. $\begin{cases} 0 \to F \\ 1 \to T \end{cases}$

①. 데이터 M개로 분할 $f(x) = \sum C_m I(x \in R_m)$

② 최상분할은 cost function 최소일 때.

$$\min_{C_m} \sum (y_i - f(x_i))^2 \qquad \hat{C}_m = ave(y_i \mid x_i \in R_m)$$

→ 각 분할 y값 평균 예측시 오류가 최소

분할 변수인 j 분할점 지정

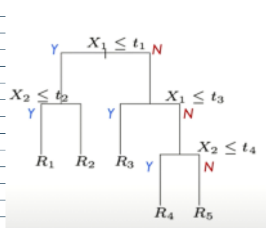
$R_1(j, s) = \{x \mid x_j \le s\}$ $R_2(j, s) = \{x \mid x_j > s\}$

$$\underset{j, s}{\operatorname{argmin}} \left[\min_{c_1} \sum_{x_i \in R_1} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2} (y_i - c_2)^2 \right]$$

$$\sum_{x_i \in R_1} (y - \hat{c}_1)^2 + \sum_{x_i \in R_2} (y - \hat{c}_2)^2$$

~ J, S 변화시키면서 탐색

3. Classification



각 관측치마다 K개 클래스 존재

$\hat{p}_{mk}$: 끝노드 m에서 k클래스에 속한 관측치 비율

$= \frac{1}{N_m} \sum I(y_i = k)$

끝노드 m으로 분류된 관측치는 k(m) 클래스로 분류. $k(m) = \underset{k}{\operatorname{argmax}}\ \hat{p}_{mk}$

$\hat{f}(x) = \sum k(m) I\{(x_1, x_2) \in R_m\}$

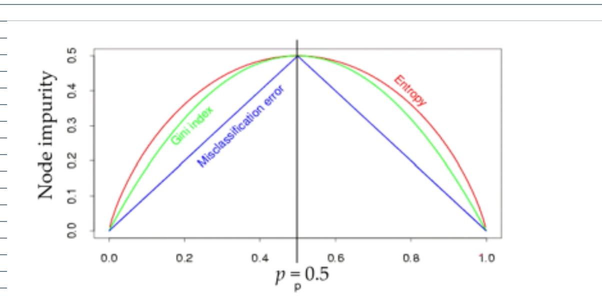
불순도 측정 : 비용함수

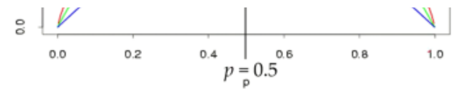
① Misclassification Rate $\left[\frac{1}{N_m} \sum I(y_i \ne k(m)) = 1 - \hat{p}_{m k(m)} \right]$

→ 실제 분류와 얼마나 파악하는가.

② Gini Index $\left[\sum_{k \ne k'} \hat{p}_{mk} \hat{p}_{mk'} = \sum_{k=1} \hat{p}_{mk} (1 - \hat{p}_{mk}) \right]$

③ Cross-Entropy $\left[- \sum_{k=1} \hat{p}_{mk} \log \hat{p}_{mk} \right]$





③ Cross-Entropy $\left[-\sum_{k=1} \hat{p}_{mk} \log \hat{p}_{mk}\right]$

=> 각 노드에서 분할변수 & 분할점 설정은 불순도 감소가 최대가 되도록

Information Gain : IG ∵ 특정변수 사용시 Entropy 감소량.

$IG(S, A) = Entropy(S) - \sum_{v \in Val(A)} \frac{|S_v|}{|S|} Entropy(S_v)$: 변수 A 사용시 IG.

4. 개별 트리 모델 단점

- 계층적 ; 중간에 에러 발생하면 다음 단계로 전파.
- 노이즈에 민감
- 과적합 위험