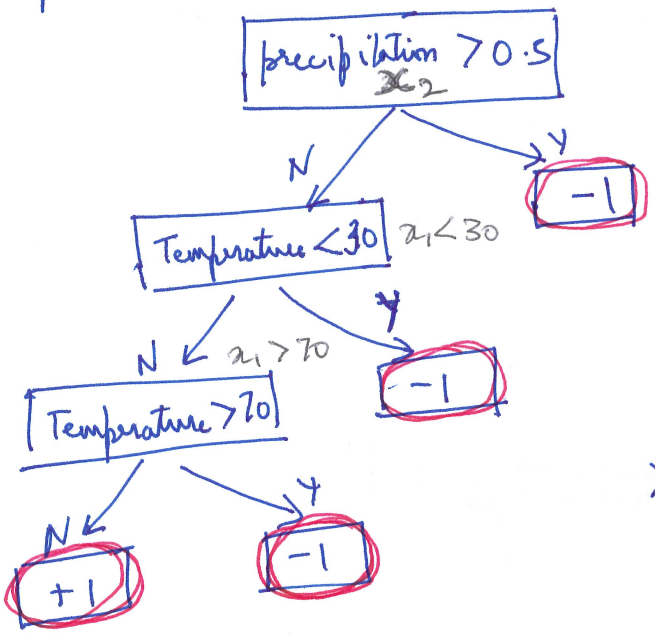
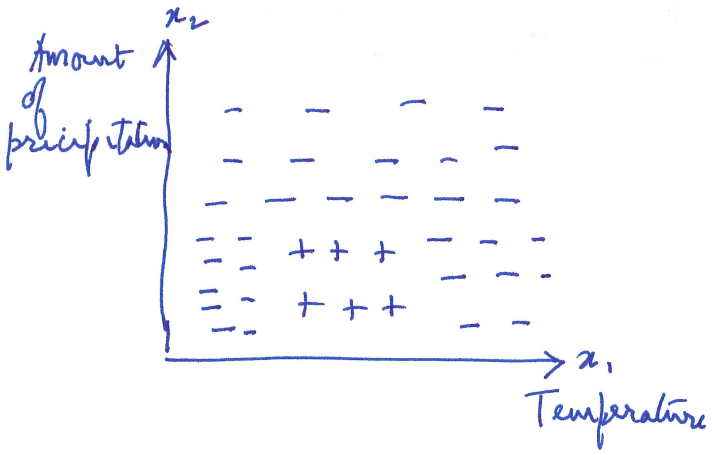
CS 307 Lecture 8

Example

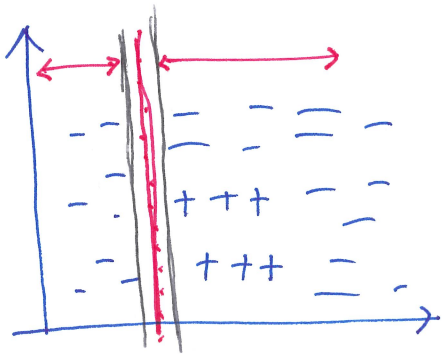




$X = (50, 0.1)$ output $= +1$

Decision Tree: Binary tree
- Internal nodes: $1[x_i \bowtie \theta]$ $(\bowtie \in \{<, >, \leq, \geq\})$
- Leaves: Labeled by the output.

Classification: Output is $\{+1, -1\}$



Decision Tree (S, k)

if construction is terminated $(|S| < k)$

Output a leaf with label $= maj(S)$

else

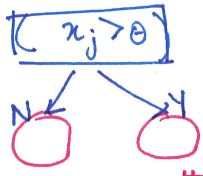
For all j, θ — Finite because you need to consider only finitely many values for θ.

$S_N = \{x \in S \mid x_j < \theta\}$

$S_Y = \{x \in S \mid x_j \geq \theta\}$

$C(j, \theta) = \left(1 - \max_{a=+1,-1} \hat{P}_{S_N}(a)\right) + \left(1 - \max_{a=+1,-1} \hat{P}_{S_Y}(a)\right)$

pick j, θ that minimize $C(j, \theta)$



$\hat{P}_S(a) =$ # examples in $S = a$ / #

Decision Tree (S_N, k)

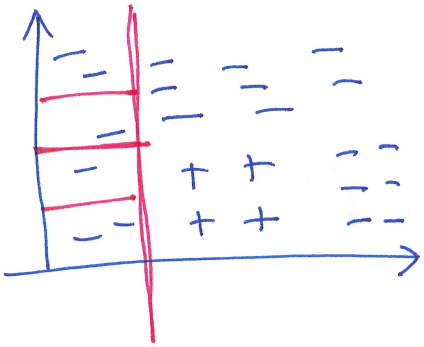
Decision Tree (S_Y, k)

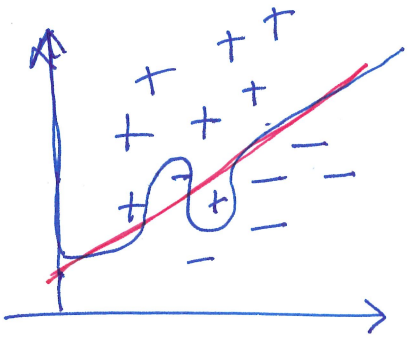
$$\hat{p}_S(a) = \frac{\#\text{ examples in } S \text{ with output } (a)}{\#\text{ examples in } S.}$$

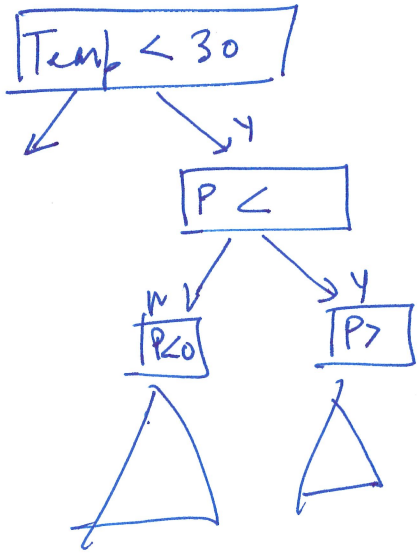
$$C_S = 1 - \max_{a \in \{1,-1\}} \hat{p}_S(a)$$

Cross entropy cost:

$$C_S = \sum_{a=\{-1,1\}} \hat{p}_S(a) \log \hat{p}_S(a)$$







- Construction of the decision
- Prune the decision true

  Bottom-up process where some subtrees are replaced by leaves

Decision Trees are prone to overfitting

Bagging:

- Construct decision trees on multiple training sets
- Actual answer on a new example is "aggregation" of the answers given by each decision tree.

Construct new training sets by sampling with replacement from examples in S.

- pick some example from S $\quad S = \{ (x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}) \ldots \}$