Example

\[ \text{precipitation } > 0.5 \]

\[ \begin{array}{c}
\text{Temperature } < 30 \\
\text{Temperature } \geq 70
\end{array} \]

\[ \begin{array}{c}
\text{if } x_i < 30 \\
\text{if } x_i \geq 70
\end{array} \]

\[ \begin{array}{c}
\text{N} \\
\text{Y}
\end{array} \]

\[ \begin{array}{c}
\text{N} \\
\text{Y}
\end{array} \]

\[ \begin{array}{c}
1 \\
-1
\end{array} \]

\[ \begin{array}{c}
1 \\
-1
\end{array} \]

\[ x = (50, 0.1) \quad \text{output } = +1 \]

Decision Tree: Binary tree
- Internal nodes: \( X_i \times \theta \) (\( \theta \in \{<, >, \leq, \geq\} \))
- Leaves: Labeled by the output.

Classification: Output is \( \{+1, -1\} \)

**Decision Tree \((S, k)\)**

- **if** construction is terminated
  - **Output** a leaf with a label = \( \text{maj}(S) \)
- **else**
  - For all \( j, \theta \)
    - \( S_N = \{ x \in S | x_j < \theta \} \)
    - \( S_Y = \{ x \in S | x_j \geq \theta \} \)
  - \( C(j, \theta) = (1 - \max_{a \in \text{maj}(S)} p_a^N(a)) + (1 - \max_{a \in \text{maj}(S)} p_a^Y(a)) \)
  - Pick \( j, \theta \) that minimize \( C(j, \theta) \)

\[ p_a(x) = \frac{\# \text{ examples in } S = a}{|S|} \]

**Decision Tree \((S_N, k)\)**

**Decision Tree \((S_Y, k)\)**
\[ \hat{p}_S(a) = \frac{\text{# examples in } S \text{ with output } a}{\text{# examples in } S} \]

\[ C_S = 1 - \max_{a \in \{0, 1\}} \hat{p}_S(a) \]

Error entropy cost:

\[ C_S = \sum_{a=0}^{1} \hat{p}_S(a) \log \hat{p}_S(a) \]

- Construction of the decision
- Prune the decision tree
  - Bottom-up process where some sub-trees 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 \( S = \{(x_i, y_i) \} \ldots \)}