Decision Trees (for classification)

- Full binary tree
  Every vertex has either 0 children or 2 children (leaves)
- Label of internal nodes:
  Feature compared with threshold
- Label of leaves: +1, -1.

Decision Tree \((S, k)\)

\[
\begin{align*}
\text{if } |S| &\leq k: \\
&\text{return Leaf (label } = \arg \max_{a \in [g + 1, b]} \hat{p}_S(a)\text{)}
\end{align*}
\]

\[
\begin{align*}
\text{else: } \\
&\text{for each } j, \theta: \\
&\quad \Sigma_{i \in S_Y} (y_i - y_i^{(j)})^2 + \Sigma_{i \in S_N} (y_i^{(j)} - y_i)^2 \\
&\quad S_N^{j, \theta} = \Sigma_{i \in S_N} (x_i, y_i) \mid x_i < \theta \\
&\quad S_Y^{j, \theta} = \Sigma_{i \in S_Y} (x_i, y_i) \mid x_i \geq \theta \\
&\quad C(j, \theta) = (1 - \max_{a \in [g + 1, b]} \hat{p}_S(a)) + (1 - \max_{a \in [g + 1, b]} \hat{p}_S(a)) \\
&\quad j^*, \theta^* = \arg \min C(j, \theta) \\
&\text{return Node (label } = j^*, \theta^*, \text{ Decision Tree } (S_N^{j^*, \theta^*}, k), \text{ Decision Tree } (S_Y^{j^*, \theta^*}, k))\end{align*}
\]
Decision Trees are prone to overfitting.

**Bagging**: Boosting Aggregation

- Given training set $S$, construct training sets $S^{(1)}, S^{(2)}, \ldots, S^{(k)}$ by picking (with replacement) $m$ examples from $S$.
- Build decision tree $T^{(1)}, T^{(2)}, \ldots, T^{(k)}$ on $S^{(1)}, S^{(2)}, \ldots S^{(k)}$.

**Aggregation**

- On a new example $(x)$:
  - Classification: $\hat{y} = \text{maj}_i T^{(i)}(x)$
  - Regression: $\hat{y} = \frac{1}{k} \sum_{i=1}^{k} T^{(i)}(x)$

**Random Forests**

- Given training set $S$, construct training sets $S^{(1)}, S^{(2)}, \ldots, S^{(k)}$ by picking (with replacement) $m$ examples from $S$.
- Build a decision tree $T^{(i)}$ on $S^{(i)}$ as follows:
  - At each stage of the decision tree construction, pick a random subset of features $I$, and you use one of the features in $I$ to split.
- On a new example $x$.
  - "Aggregate" the outputs of $T^{(i)}(x)$. 
$k$-Nearest Neighbors:

No hypothesis constructed from the training set.

On a new example $x$, $S = \{ (x^{(0)}, y^{(0)}) \ldots (x^{(n)}, y^{(n)}) \}$.

- Compute permutation of $S$ say $(x^{\pi(0)}, y^{\pi(0)}) \ldots (x^{\pi(n)}, y^{\pi(n)})$ such that

$$d(x, x^{\pi(i)}) \leq d(x, x^{\pi(i+1)})$$

- Output $y = \text{maj} (y^{\pi(1)}, y^{\pi(2)} \ldots y^{\pi(k)})$ (classification)

$$y = \frac{1}{k} \sum_{i=1}^{k} y^{\pi(i)}$$ (regression)

Examples that are close by have outputs that are close.

c-Lipschitz: $c \cdot d(x^{(i)}, x^{(i)}) \geq d(y^{(i)}, y^{(i)})$