Decision Trees

\[ \text{Decision}: \]

1. Which attribute?  
2. Which value?

\[ n = |D| \rightarrow p_c, p_{c_2}, \ldots, p_{c_k} \]

\[ y_i \in \{c_1, c_2, \ldots, c_k\} \]

\[ \tilde{z}_i \in \mathbb{R}^d \]

\[ \text{Entropy} \]

\[ \text{Split} \]

\[ \text{feature space} \]

\[ D \rightarrow \text{hyperplane/split} \]

\[ \text{Ax}_a - \text{aligned} \] (parallel to one of the axes)
\[ P_{c_i} = \frac{|\{x \mid y = c_i\}|}{n} \]

\[ H(D) = -\sum_{i=1}^{k} P_{c_i} \log P_{c_i} \quad : \text{Amount of disorder in terms of labels} \]

Split: \( A_i \leq v_j \)

\[ H(D_{y}, D_{N}) = \frac{|D_{y}|}{|D|} H(D_{y}) + \frac{|D_{N}|}{|D|} H(D_{N}) \quad \text{Entropy of the split} \]

\[ \text{Info Gain} = H(D) - H(D_{y}, D_{N}) \quad (A_i \leq v_j) \]

\[ D \left[ \begin{array}{l}
\text{we have to try all possible } A_i \text{ & } v_j \text{ value}
\text{Compute } IG(A_i, v_j)
\text{Choose the best one of these and there is the decision at this node}
\end{array} \right. \]

\[ \text{we'd have to try all categories : } A_1, A_2 \ldots A_d \]

<table>
<thead>
<tr>
<th>Temp</th>
<th>Pressure</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>70°C</td>
<td>1 atm</td>
<td>Hot</td>
</tr>
<tr>
<td>110°C</td>
<td>2 atm</td>
<td>Cold</td>
</tr>
</tbody>
</table>
Temp ≤ 5

For 5, choose all mid-points between observed value as the split values.

Temp ≤ 20
Temp ≤ 50
Temp ≤ 60.5
Temp ≤ 90

Gini Order:

\[ G(D) = 1 - \sum_{i=1}^{k} (P_c_i)^2 \]

Decision Trees can easily handle categorical attributes

\[ \text{A}_3 \in \{ R, B \} \]

\[ \begin{array}{c|c|c|c}
  \text{A}_1 & \text{A}_2 & \text{A}_3 & \text{Class} \\
  5 & 10 & R & p \\
  6 & 12 & G & p \\
  20 & 1 & K & N \\
  15 & 100 & B & R \\
\end{array} \]
All possible classifiers

\( A_3 \in V \)

\( V \subseteq \{ \text{set of values for } A_3 \} \)

\( A_3 \in \{ R_3 \} \)

\( p_4 \quad 1p \quad \text{IN} \)

\( 2p \)

Unsupervised approach

Clustering

Find "groups" in the data

Clusters

1) Distance vs Similarity

2) Internal / intra
   \( \Rightarrow \) high sim within a cluster

3) External / inter
   \( \Rightarrow \) low similarity between clusters

Flat vs

1) Hierarchical clustering

2) exclusive vs overlapping
   hard vs soft
   (probabilistic)

3) geometry of the group

1 \( \rightarrow \infty \)

\( d \)

\( n \)

\( \text{x} \in \)
3) geometry of the group

\[ D = \{ \bar{x}_c \}_{c=1}^n \]

\[ C_1 = \{ \} \quad C_2 = \{ \} \quad \ldots \quad C_K = \{ \} \]

Partitioning:
\[ c_i \cap c_j = \emptyset \quad \forall i \neq j \]
\[ \bigcup_{C_i} = D \]

Centroids / mean

\[ \text{SSE:} \quad \sum_{i=1}^{K} \sum_{x_j \in C_i} \left( \bar{x}_j - \bar{\mu}_i \right)^2 \]

NP-hard problem to minimize.
NP-hard problem to minimize:

\[ k \geq 2 \]

**K** - means, greedy,

**Initialize means:**

1. **a)** Choose \( M_1, M_2, \ldots, M_k \)
2. **b)** Max-dissimilarity approach:
   - \( M_1 \leftarrow \) a random point \( \tilde{z} \in D \)
   - \( M_2 \leftarrow \) furthest away from \( M_1 \)
   - \( M_i \leftarrow \) furthest away from \( M_1, M_2, \ldots, M_{i-1} \)

**Iterative update to \( \tilde{M}_i \) value**

1. **a)** A set \( \mathcal{X} \) means indicates a partition
2. **b)** Compute distance to \( M_j \) for \( j = 1, \ldots, k \)
3. **c)** Assign \( \tilde{x}_i \) to the closest cluster
   - \( C_i = C_i \cup \{ \tilde{x}_i \} \)
   - \( i^* = \arg \min \{ \| \tilde{x}_i - C_i \|_2^2 \} \)
   - \( j = 1, \ldots, k \)
4. **Recompute the mean for the partition**
   - \( \tilde{M}_i = \text{mean}(C_i) \) for all \( i \)

- **V** - points
- **k** - clusters
- **d** - dimensions
- **O(nkd)** per iteration
Repeat 1) and 2) until convergence

\[ k = 3 \]

Convex clusters!

Non-linear / non-convex clusters:

**Kernel k-means**

\[ \text{all operations have to be dot-products, i.e. using } \phi \]

**x_i \rightarrow \phi(x_i)**
Kernel k-means

1) $\mu^1, \mu^2, ..., \mu^k$ 
   - not allowed

$K \leftarrow$ Compute $K$

Growable:
   Create a random partitioning!

$\{C_1, C_2, ..., C_k\}$

Compute $d(x_i, \mu_j)$ for $j = 1 \ldots k$ and $i = 1 \ldots n$

Reassign $x_i$ to the closest cluster $C_l$

$\{C_1', C_2', ..., C_k'\}$

Repeat until the clusters do not change too much

$m = \text{linear, regular k-means for 1 round}$

$k \leftarrow O(n^2)$

$n, k, n \over O(n^2 k)$ per iteration
for 1 round
then we start to see the
initial clusters

EM: expectation maximization

clustering

→ soft/probabilistic clustering

$c_1, c_2, \ldots, c_K$

$P(c_i | x_j)$

→ easy to get a hard clustering

$k$ clusters $\rightarrow$ multivariate normal shaped

$\tilde{\mu}_i, \Sigma_i; P(c_i)$

$P(c_i | x_j) = \frac{P(x_j | c_i) \cdot P(c_i)}{P(x_j)}$
Normal distribution

Estimate $\hat{\mu}_i, \hat{\Sigma}_i$

Given $\mu_i, \Sigma_i, p(c_i)$

\[
\begin{cases}
\text{Expectation:} & p(X_j|c_i) = w_{ji} \\
\text{Maximization:} & \mu_i, \Sigma_i, p(c_i) \end{cases}
\]

Weighted estimates