Decision Trees - Preliminaries

Prof. Dan A. Simovici

UMB
1 Decision Trees

2 Equivalence Relations

3 Partitions

4 Trace of a Partition on a Set
Decision trees learning is one of the most widely used for approximative learning of discrete-valued functions that is robust relative to noise in data.
Consider a table that shows the decision of playing tennis depending on certain climatic factors. The attributes and their domains are shown below:

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outlook</td>
<td>{sunny, overcast, rain}</td>
</tr>
<tr>
<td>Temperature</td>
<td>{hot, mild, cool}</td>
</tr>
<tr>
<td>Humidity</td>
<td>{normal, high}</td>
</tr>
<tr>
<td>Wind</td>
<td>{weak, strong}</td>
</tr>
</tbody>
</table>

The decision attribute is PlayTennis; this attribute has the domain {yes, no}. 
The data set is shown below:

<table>
<thead>
<tr>
<th>Outlook</th>
<th>Temperature</th>
<th>Humidity</th>
<th>Wind</th>
<th>PlayTennis</th>
</tr>
</thead>
<tbody>
<tr>
<td>sunny</td>
<td>hot</td>
<td>high</td>
<td>weak</td>
<td>no</td>
</tr>
<tr>
<td>sunny</td>
<td>hot</td>
<td>high</td>
<td>strong</td>
<td>no</td>
</tr>
<tr>
<td>overcast</td>
<td>hot</td>
<td>high</td>
<td>weak</td>
<td>yes</td>
</tr>
<tr>
<td>rain</td>
<td>mild</td>
<td>high</td>
<td>weak</td>
<td>yes</td>
</tr>
<tr>
<td>rain</td>
<td>cool</td>
<td>normal</td>
<td>weak</td>
<td>yes</td>
</tr>
<tr>
<td>rain</td>
<td>cool</td>
<td>normal</td>
<td>strong</td>
<td>no</td>
</tr>
<tr>
<td>overcast</td>
<td>cool</td>
<td>normal</td>
<td>strong</td>
<td>yes</td>
</tr>
<tr>
<td>sunny</td>
<td>mild</td>
<td>high</td>
<td>weak</td>
<td>no</td>
</tr>
<tr>
<td>sunny</td>
<td>cool</td>
<td>normal</td>
<td>weak</td>
<td>yes</td>
</tr>
<tr>
<td>rain</td>
<td>mild</td>
<td>normal</td>
<td>weak</td>
<td>yes</td>
</tr>
<tr>
<td>sunny</td>
<td>mild</td>
<td>normal</td>
<td>strong</td>
<td>yes</td>
</tr>
<tr>
<td>overcast</td>
<td>mild</td>
<td>high</td>
<td>strong</td>
<td>yes</td>
</tr>
<tr>
<td>rain</td>
<td>hot</td>
<td>normal</td>
<td>weak</td>
<td>yes</td>
</tr>
<tr>
<td>rain</td>
<td>mild</td>
<td>high</td>
<td>strong</td>
<td>no</td>
</tr>
</tbody>
</table>

The goal of a decision tree is to formulate a rule (as simple as possible) that will allow us to decide when to play tennis as a function of climate factors.
Equivalence relations and partitions are essential instruments in the study of decision trees.

**Definition**

An *equivalence relation* on a set $S$ is a relation $\rho$ that is reflexive, symmetric, and transitive.
This means that

- \((x, x) \in \rho\) for every \(x \in S\);
- \((x, y) \in \rho\) if and only if \((y, x) \in \rho\);
- \((x, y) \in \rho\) and \((y, z) \in \rho\) imply \((x, z) \in \rho\).
Example

Let $U$ and $V$ be two sets, and consider a function $f : U \rightarrow V$. The relation $\ker(f) \subseteq U \times U$, called the kernel of $f$, is given by

$$\ker(f) = \{(u, u') \in U \times U \mid f(u) = f(u')\}.$$ 

In other words, $(u, u') \in \ker(f)$ if $f$ maps both $u$ and $u'$ into the same element of $V$. 
Equivalence Relations

\[ S \rightarrow T \]

\[ f \rightarrow \ldots \rightarrow t \]
Example

Let \( m \in \mathbb{N} \) be a positive natural number. Define the function \( f_m : \mathbb{Z} \to \mathbb{N} \) by \( f_m(n) = r \) if \( r \) is the remainder of the division of \( n \) by \( m \). The range of the function \( f_m \) is the set \( \{0, \ldots, m - 1\} \).

The relation \( \ker(f_m) \) is usually denoted by \( \equiv_m \). We have \((p, q) \in \equiv_m\) if and only if \( p - q \) is divisible by \( m \); if \((p, q) \in \equiv_m\), we also write \( p \equiv q \pmod{m} \).
Definition

Let $\rho$ be an equivalence on a set $U$ and let $u \in U$. The *equivalence class* of $u$ is the set $[u]_\rho$, given by

$$[u]_\rho = \{ y \in U \mid (u, y) \in \rho \}.$$ 

When there is no risk of confusion, we write simply $[u]$ instead of $[u]_\rho$. 
Note that an equivalence class \([u]\) of an element \(u\) is never empty since \(u \in [u]\) because of the reflexivity of \(\rho\).

**Theorem**

*Let \(\rho\) be an equivalence on a set \(U\) and let \(u, v \in U\). The following three statements are equivalent:*

1. \((u, v) \in \rho;\)
2. \([u] = [v];\)
3. \([u] \cap [v] \neq \emptyset.\)
Definition

Let $S$ be a set and let $\rho \in \text{EQ}(S)$. A subset $U$ of $S$ is $\rho$-saturated if it equals a union of equivalence classes of $\rho$.

It is easy to see that $U$ is a $\rho$-saturated set if and only if $x \in U$ and $(x, y) \in \rho$ imply $y \in U$. It is clear that both $\emptyset$ and $S$ are $\rho$-saturated sets.
Definition

Let $S$ be a nonempty set. A *partition* of $S$ is a nonempty collection

$$\pi = \{ B_i \mid i \in I \}$$

of nonempty subsets of $S$, such that $\bigcup \{ B_i \mid i \in I \} = S$, and $B_i \cap B_j = \emptyset$ for every $i, j \in I$ such that $i \neq j$.

Each set $B_i$ of $\pi$ is a *block* of the partition $\pi$.

The set of partitions of a set $S$ is denoted by $\text{PART}(S)$. The partition of $S$ that consists of all singletons of the form $\{s\}$ with $s \in S$ will be denoted by $\alpha_S$; the partition that consists of the set $S$ itself will be denoted by $\omega_S$. 
Example

For the two-element set $S = \{a, b\}$, there are two partitions: the partition $\alpha_S = \{\{a\}, \{b\}\}$ and the partition $\omega_S = \{\{a, b\}\}$.

For the one-element set $T = \{c\}$, there exists only one partition, $\alpha_T = \omega_T = \{\{t\}\}$. 
Example

A complete list of partitions of a set $S = \{a, b, c\}$ consists of

$$\pi_0 = \{\{a\}, \{b\}, \{c\}\}, \quad \pi_1 = \{\{a, b\}, \{c\}\},$$
$$\pi_2 = \{\{a\}, \{b, c\}\}, \quad \pi_3 = \{\{a, c\}, \{b\}\},$$
$$\pi_4 = \{\{a, b, c\}\}.$$  

Clearly, $\pi_0 = \alpha_S$ and $\pi_4 = \omega_S$. 
Definition

Let $S$ be a set and let $\pi, \sigma \in \text{PART}(S)$. The partition $\pi$ is \term{finer} than the partition $\sigma$ if every block $C$ of $\sigma$ is a union of blocks of $\pi$. This is denoted by $\pi \leq \sigma$. 
Theorem

Let $\pi = \{B_i \mid i \in I\}$ and $\sigma = \{C_j \mid j \in J\}$ be two partitions of a set $S$. For $\pi, \sigma \in \text{PART}(S)$, we have $\pi \leq \sigma$ if and only if for every block $B_i \in \pi$ there exists a block $C_j \in \sigma$ such that $B_i \subseteq C_j$. 
Proof

If \( \pi \leq \sigma \), then it is clear for every block \( B_i \in \pi \) there exists a block \( C_j \in \sigma \) such that \( B_i \subseteq C_j \).

Conversely, suppose that for every block \( B_i \in \pi \) there exists a block \( C_j \in \sigma \) such that \( B_i \subseteq C_j \). Since two distinct blocks of \( \sigma \) are disjoint, it follows that for any block \( B_i \) of \( \pi \), the block \( C_j \) of \( \sigma \) that contains \( B_i \) is unique.

Therefore, if a block \( B \) of \( \pi \) intersects a block \( C \) of \( \sigma \), then \( B \subseteq C \).

Let \( Q = \bigcup \{B_i \in \pi \mid B_i \subseteq C_j\} \). Clearly, \( Q \subseteq C_j \). Suppose that there exists \( x \in C_j - Q \). Then, there is a block \( B_\ell \in \pi \) such that \( x \in B_\ell \cap C_j \), which implies that \( B_\ell \subseteq C_j \). This means that \( x \in B_\ell \subseteq C \), which contradicts the assumption we made about \( x \). Consequently, \( C_j = Q \), which concludes the argument.
Note that $\alpha_S \leq \pi \leq \omega_S$ for every $\pi \in \text{PART}(S)$.

Two equivalence classes either coincide or are disjoint. Therefore, starting from an equivalence $\rho \in \text{EQ}(U)$, we can build a partition of the set $U$.

**Definition**

The *quotient set* of the set $U$ with respect to the equivalence $\rho$ is the partition $U/\rho$, where

$$U/\rho = \{ [u]_\rho \mid u \in U \}.$$ 

An alternative notation for the partition $U/\rho$ is $\pi_\rho$. 

Theorem

Let $\pi = \{ B_i \mid i \in I \}$ be a partition of the set $U$. Define the relation $\rho_\pi$ by $(x, y) \in \rho_\pi$ if there is a set $B_i \in \pi$ such that $\{x, y\} \subseteq B_i$. The relation $\rho_\pi$ is an equivalence.
Proof

Let $B_i$ be the block of the partition that contains $u$. Since $\{u\} \subseteq B_i$, we have $(u, u) \in \rho_{\pi}$ for any $u \in U$, which shows that $\rho_{\pi}$ is reflexive.

The relation $\rho_{\pi}$ is clearly symmetric. To prove the transitivity of $\rho_{\pi}$, consider $(u, v), (v, w) \in \rho_{\pi}$. We have the blocks $B_i$ and $B_j$ such that $\{u, v\} \subseteq B_i$ and $\{v, w\} \subseteq B_j$. Since $v \in B_i \cap B_j$, we obtain $B_i = B_j$ by the definition of partitions; hence, $(u, w) \in \rho_{\pi}$. 
Example

Let $f : S \rightarrow T$ be a function. For $t \in T$ define the set $B_t = \{x \in S \mid f(x) = t\}$. Then, the collection of sets $\{B_t \mid t \in T \text{ and } B_t \neq \emptyset\}$ is a partition of $S$ that corresponds to the equivalence $\ker(f)$.
Let $T$ be a set and let $\pi = \{B_1, \ldots, B_k\}$ be a partition of $T$. If $S$ is a subset of $T$, the trace of $\pi$ on the set $S$ is the collection of sets:

$$\pi_S = \{B_i \cap S \mid B_i \in \pi \text{ and } B_i \cap S \neq S\}.$$ 

Note that $\pi_S$ is a partition of $S$. 
Example
We have \( S = \{5, 6, 7, 12, 13, 14\} \) and \( \pi = \{B_1, B_2, B_3, B_4, B_5\} \). The trace of \( \pi \) on \( S \) denoted by \( \pi_S \) consists of

\[
B_2 \cap S = \{12\}, \\
B_3 \cap S = \{13, 14\}, \\
B_5 \cap S = \{5, 6, 7\}.
\]