Decision Trees - Preliminaries

Prof. Dan A. Simovici

UMB

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:

Attribute	**Domain**
Outlook	$\{$sunny, overcast, rain$\}$
Temperature	$\{$hot, mild, cool$\}$
Humidity	$\{$normal, high$\}$
Wind	$\{$weak, strong$\}$

The decision attribute is PlayTennis; this attribute has the domain $\{$yes, no$\}$.

The data set is shown below:

	Outlook	Temperature	Humidity	Wind	PlayTennis
1	sunny	hot	high	weak	no
2	sunny	hot	high	strong	no
3	overcast	hot	high	weak	yes
4	rain	mild	high	weak	yes
5	rain	cool	normal	weak	yes
6	rain	cool	normal	strong	no
7	overcast	cool	normal	strong	yes
8	sunny	mild	high	weak	no
9	sunny	cool	normal	weak	yes
10	rain	mild	normal	weak	yes
11	sunny	mild	normal	strong	yes
12	overcast	mild	high	strong	yes
13	rain	hot	normal	weak	yes
14	rain	mild	high	strong	no

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 ρ that is reflexive, symmetric, and transitive.

This means that

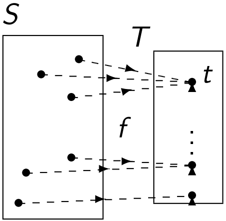
- $(x, x) \in \rho$ for every $x \in S$;
- $(x, y) \in \rho$ if and only if $(y, x) \in \rho$;
- $(x, y) \in \rho$ and $(y, z) \in \rho$ imply $(x, z) \in \rho$.

Example

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

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

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


S
T
t
f

Example

Let $m \in \mathbb{N}$ be a positive natural number. Define the function $f_m : \mathbb{Z} \longrightarrow \mathbb{N}$ 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 $\mathbf{ker}(f_m)$ 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 $p \equiv q (\text{mod } m)$.

Definition

Let ρ 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 ρ.

Theorem

Let ρ 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 $\rho \in \mathrm{EQ}(S)$. A subset U of S is *ρ-saturated* if it equals a union of equivalence classes of ρ.

It is easy to see that U is a ρ-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 ρ-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 π is a *block* of the partition π.
The set of partitions of a set S is denoted by $\mathrm{PART}(S)$. The partition of S that consists of all singletons of the form $\{s\}$ with $s \in S$ will be denoted by α_S; the partition that consists of the set S itself will be denoted by ω_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

$$\begin{array}{lllllll} \pi_0 & = & \{\{a\}, \{b\}, \{c\}\}, & \pi_1 & = & \{\{a, b\}, \{c\}\}, \\ \pi_2 & = & \{\{a\}, \{b, c\}\}, & \pi_3 & = & \{\{a, c\}, \{b\}\}, \\ \pi_4 & = & \{\{a, b, c\}\}. \end{array}$$

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

Definition

Let S be a set and let $\pi, \sigma \in \mathsf{PART}(S)$. The partition π is *finer* than the partition σ if every block C of σ is a union of blocks of π. This is denoted by $\pi \leqslant \sigma$.

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 $\pi, \sigma \in PART(S)$, we have $\pi \leqslant \sigma$ 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 $\pi \leqslant \sigma$, 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 σ are disjoint, it follows that for any block B_i of π, the block C_j of σ that contains B_i is unique. Therefore, if a block B of π intersects a block C of σ, 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 $\alpha_S \leqslant \pi \leqslant \omega_S$ 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 ρ is the partition U/ρ, where

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

An alternative notation for the partition U/ρ is π_ρ.

Theorem

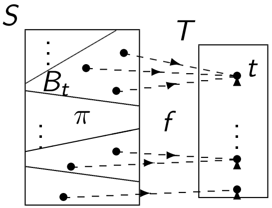
Let $\pi = \{B_i \mid i \in I\}$ *be a partition of the set* U*. Define the relation* ρ_π *by* $(x, y) \in \rho_\pi$ *if there is a set* $B_i \in \pi$ *such that* $\{x, y\} \subseteq B_i$*. The relation* ρ_π *is an equivalence.*

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 ρ_π is reflexive.
The relation ρ_π is clearly symmetric. To prove the transitivity of ρ_π, 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 $f : S \longrightarrow T$ 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 $\ker(f)$.

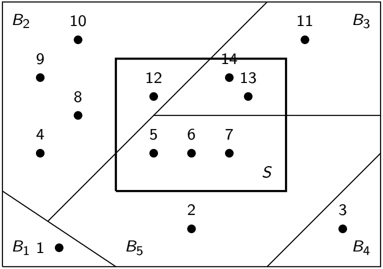


Let T be a set and let $\pi = \{B_1, \ldots, B_k\}$ be a partition of T. If S is a subset of T, the *trace* of π on the set S is the collection of sets:

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

Note that π_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 π on S denoted by π_S consists of

$$\begin{aligned} B_2 \cap S &= \{12\}, \\ B_3 \cap S &= \{13, 14\}, \\ B_5 \cap S &= \{5, 6, 7\}. \end{aligned}$$

On slide 17 we introduced the partial order $\leqslant$ on the set $\mathrm{PART}(S)$.

Definition

Let $\pi, \sigma \in \mathrm{PART}(S)$ be two partitions of the finite set S. We say that σ *covers* π (and write $\pi \lhd \sigma$) if $\pi = \{B_1, \ldots, B_m\}$ and the blocks of σ are the same as the blocks of π, except for a block C of σ that is the union of two blocks B', B'' of π.

Note that if $\pi \lhd \sigma$ and π has m blocks, then σ has $m - 1$ blocks.

Example

Let $S = \{1, 2, 3, 4, 5, 6\}$ and let $\pi = \{\{1\}, \{2, 4\}, \{3, 5, 6\}\}$. Then, the partitions

$$\begin{aligned} \sigma_1 &= \{\{1, 2, 4\}, \{3, 5, 6\}\}, \\ \sigma_2 &= \{\{1, 3, 5, 6\}, \{2, 4\}\}, \\ \sigma_3 &= \{\{1\}, \{2, 3, 3, 5, 6\}\} \end{aligned}$$

all cover π.

Note that if $\pi \leqslant \sigma$ for $\pi, \sigma \in \text{PART}(S)$ there exists a sequence of partitions $\tau_1, \ldots, \tau_k$ such that

$$\pi = \tau_1 \lessdot \tau_2 \lessdot \cdots \lessdot \tau_k = \sigma.$$

In other words, if $\pi \leqslant \sigma$ it is possible to "interpolate" the partitions $\tau_1, \ldots, \tau_k$ such that each partition τ_ℓ is covered by $\tau_{\ell+1}$ for $1 \leqslant \ell \leqslant k - 1$.

Example

Let $\pi = \{\{1,2\},\{3\},\{4,5,6\},\{7,8\},\{9\}\}$ and $\sigma = \{\{1,2,3,7,8\},\{4,5,6,9\}\}$. Clearly, $\pi \leqslant \sigma$.
The chain of partitions $\tau_1 \lhd \tau_2 \lhd \tau_3 \lhd \tau_4$ can be interpolated between π and σ, where

$$\begin{aligned}
\tau_1 &= \{\{1,2\},\{3\},\{4,5,6\},\{7,8\},\{9\}\} = \pi, \\
\tau_2 &= \{\{1,2,3\},\{4,5,6\},\{7,8\},\{9\}\}, \\
\tau_3 &= \{\{1,2,3\},\{4,5,6,9\},\{7,8\}\} \\
\tau_4 &= \{\{1,2,3,7,8\},\{4,5,6,9\}\} = \sigma.
\end{aligned}$$