Belief Decision Trees

Thierry Denœux

tdenoeux@utc.fr

University of Compiègne, France

Decision trees

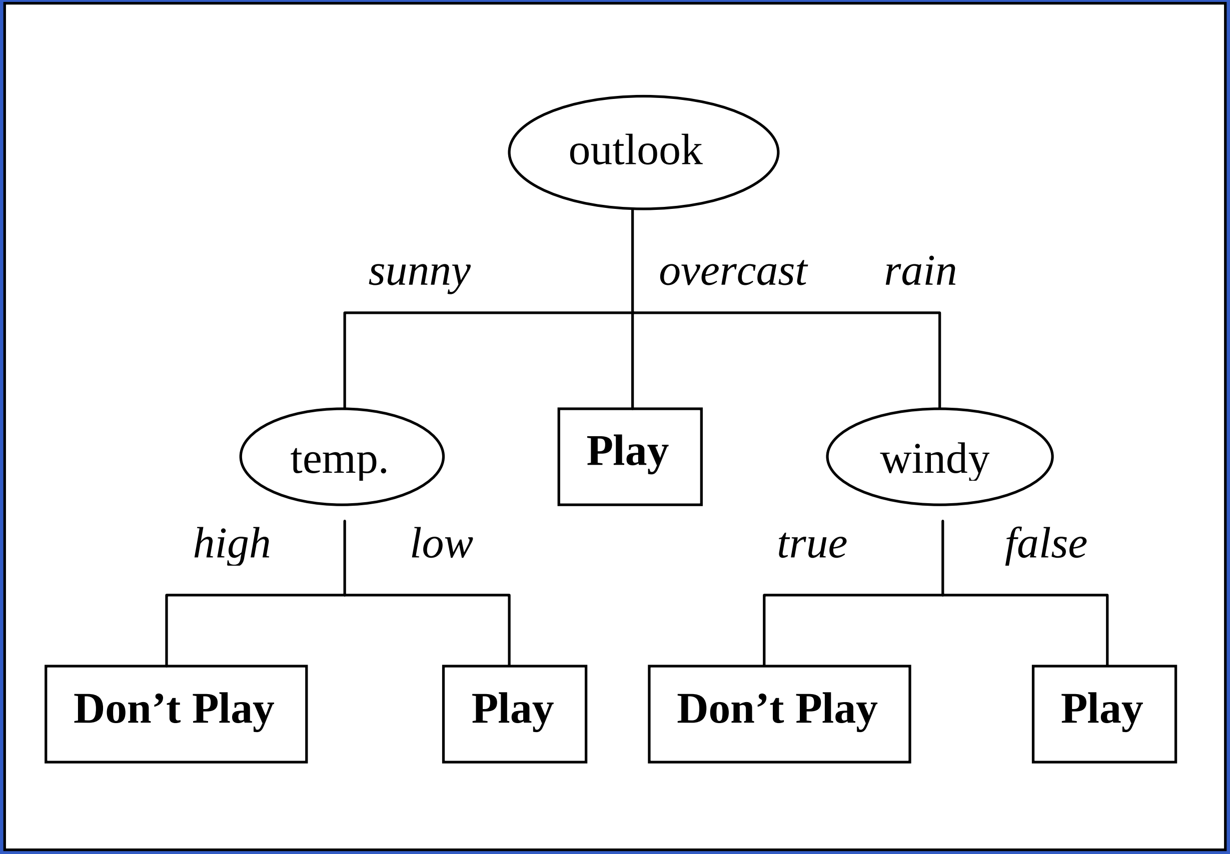
- DT generation techniques = family of classification methods studied in ML and Statistics, widely used in data mining applications.
- Best known programs: CART (Breiman et al., 1984), C4.5 (Quinlan, 1988).
- Advantages:
  - feature selection + learning at the same time
  - accept quantitative and qualitative features
  - invariance to monotonic transf. of inputs
  - lead to easily interpretable decision rules, etc...

Belief decision trees

- Extension of the tree-structured classification methodology in the TBM framework, recently introduced by Elouedi and Smets (2000), Denœux and Skarstein-Bjanger (2000);
- Goal: increase the applicability of DT techniques to weaker forms of data
  - learning data with imprecise or uncertain class labels
  - allow for imprecise or uncertain attribute values in the testing phase
- Several algorithms, e.g. averaging approach.

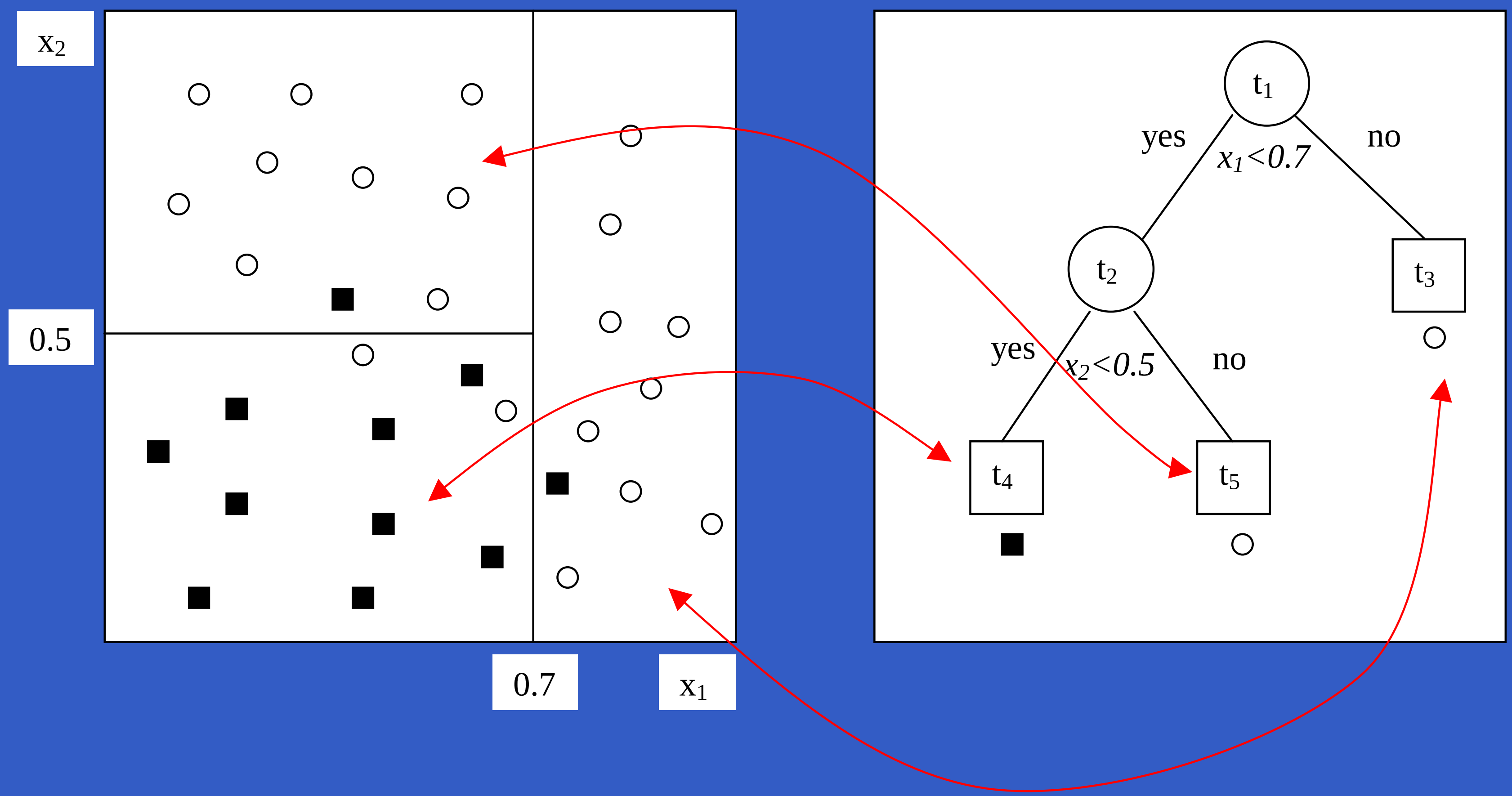
A decision tree

Decision tree = representation of a sequential decision procedure



DT induction (1)

Main idea: recursive partitioning of the learning set.

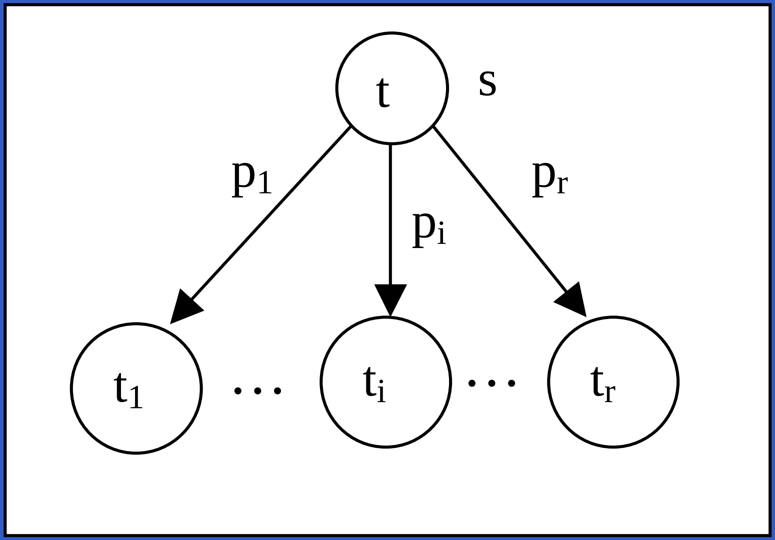


DT induction (2)

- At each step, try to split a node (=subset of patterns) in such a way that the child nodes are, on average, ‘purer’ in one class than their parents.
- Impurity criteria are based on the proportions $\widehat{p}_k(t) = n_k(t)/n(t)$ of each class in node t.
- Classical impurity criterion: entropy of the empirical class distribution $(\widehat{p}_k(t))_{k=1}^K$

$$I(t) = -\sum_{k=1}^{K} \widehat{p}_k(t) \log_2 \widehat{p}_k(t)$$

DT induction (3)



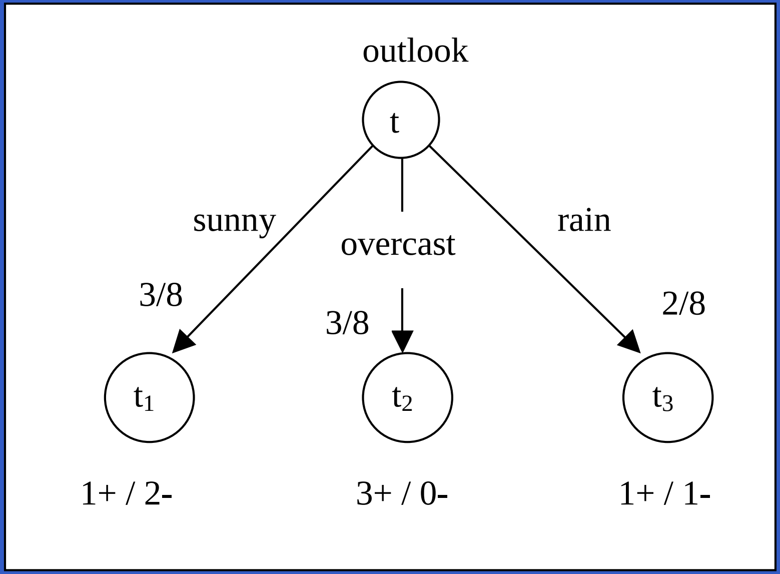
Goodness of split for a candidate split s of node t:

$$\Delta I(s,t) = I(t) - \sum_{i=1}^{r} p_i I(t_i)$$

Example (1)

Outlook	Temperature	Windy?	Class
sunny	low	true	Play
sunny	high	true	Don’t Play
sunny	high	false	Don’t Play
overcast	low	true	Play
overcast	high	false	Play
overcast	low	false	Play
rain	low	true	Don’t Play
rain	low	false	Play

Example (2)

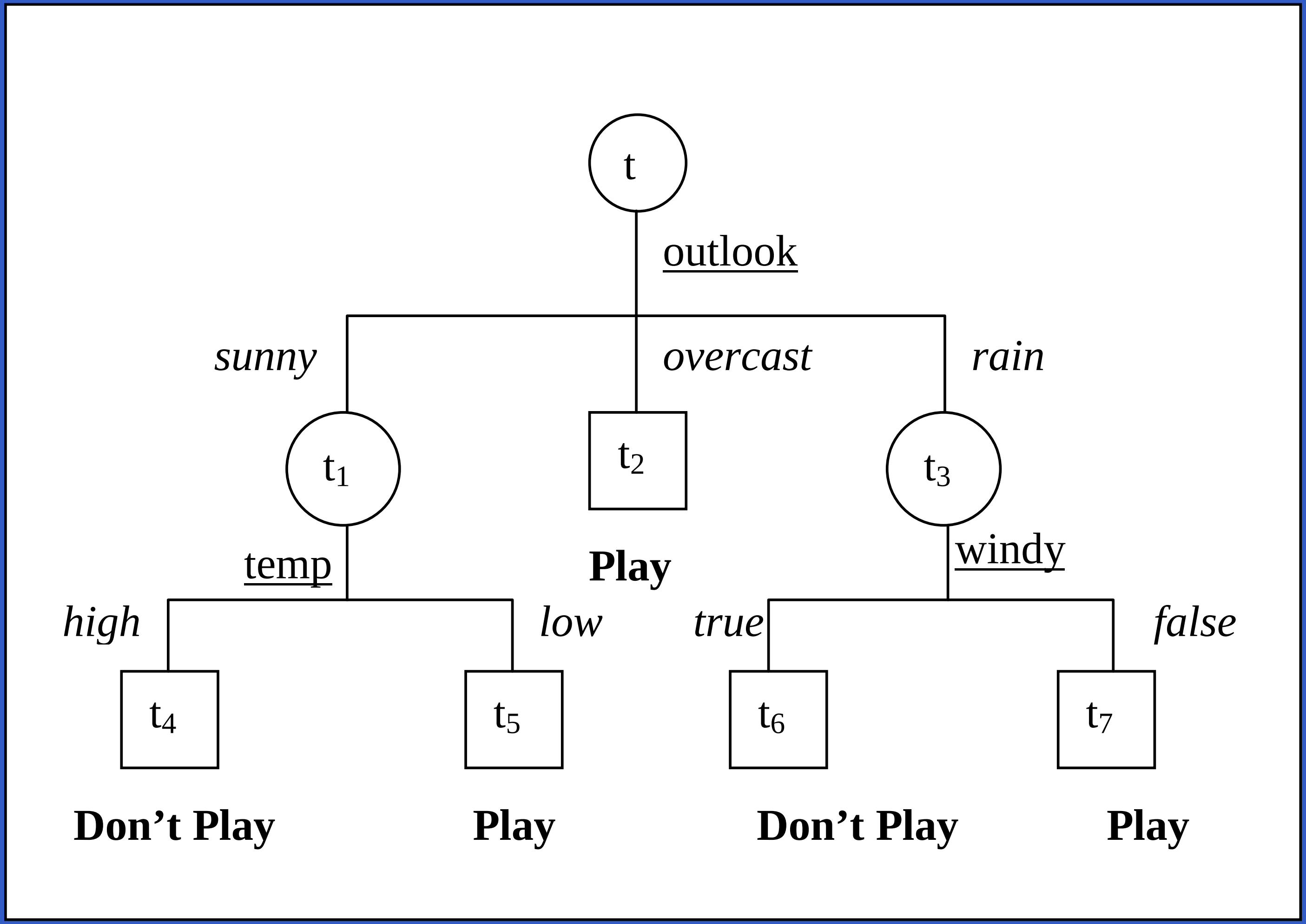


$$I(t) = 0.95 \quad I(t_1) = 0.92 \quad I(t_2) = 0 \quad I(t_3) = 1$$

$$\Delta I(\text{outlook}, t) = 0.95 - (0.92 \times 3/8 + 0 + 1 \times 2/8) = 0.36$$

$$\Delta I(\text{temp}, t) = 0.16 \qquad \Delta I(\text{windy}, t) = 0.05$$

Example (3)



t
outlook
sunny
overcast
rain
t1
t2
t3
temp
Play
windy
high
low
true
false
t4
t5
t6
t7
Don't Play
Play
Don't Play
Play

Extension to uncertain labels - Example

Outlook	Temp.	Windy?	$m(\{P\})$	$m(\{NP\})$	$m(\Omega)$
sunny	low	true	0.7	0	0.3
sunny	high	true	0.3	0.6	0.1
sunny	high	false	0	0.9	0.1
overcast	low	true	0.4	0.3	0.3
overcast	high	false	0.8	0	0.2
overcast	low	false	0.7	0.2	0.1
rain	low	true	0.1	0.9	0
rain	low	false	0.3	0.2	0.5

Extension to uncertain labels (1)

- How can the impurity (entropy) criterion $I(t)$ be generalized ?
- Classical case: $n_k(t)/n(t)$ is the probability that a case selected at random in node t belongs to class ω_k. $I(t)$ is the entropy of the probability distribution on Ω.
- Consider a node t composed of $n(t)$ cases $\{(\mathbf{x}_i, m_i)\}_{i=1}^{n(t)}$. Pick a case at random. What is Your belief that you will get a case from class ω_k ?

Extension to uncertain labels (2)

Let c denote the class of the selected case. $\forall A \subseteq \Omega$,

$$
\begin{aligned}
m^{\Omega}[t](c \in A) &= \sum_{i=1}^{n(t)} P(\text{selected case is } i) m(c_i \in A) \\
&= \frac{1}{n(t)} \sum_{i=1}^{n(t)} m_i(A) = \overline{m}(A)
\end{aligned}
$$

$$
\boxed{m^{\Omega}[t] = \frac{1}{n(t)} \sum_{i=1}^{n(t)} m_i = \overline{m}}
$$

Extension to uncertain labels (3)

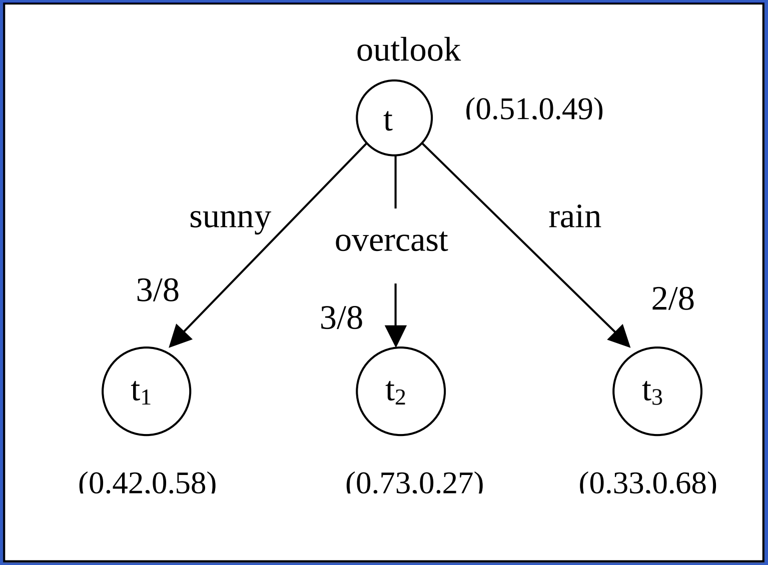
- The impurity of t can be defined as the entropy of the pignistic probability distribution $BetP_{\overline{m}}$ associated to $m^{\Omega}[t]$:

$$I(t) = -\sum_{k=1}^{K} BetP_{\overline{m}}(\omega_k) \log_2 BetP_{\overline{m}}(\omega_k)$$

- Property: $BetP_{\overline{m}}$ is the average of the pignistic probabilities associated to the m_i

$$BetP_{\overline{m}} = \frac{1}{n(t)} \sum_{i=1}^{n(t)} BetP_i = \overline{BetP}$$

Example

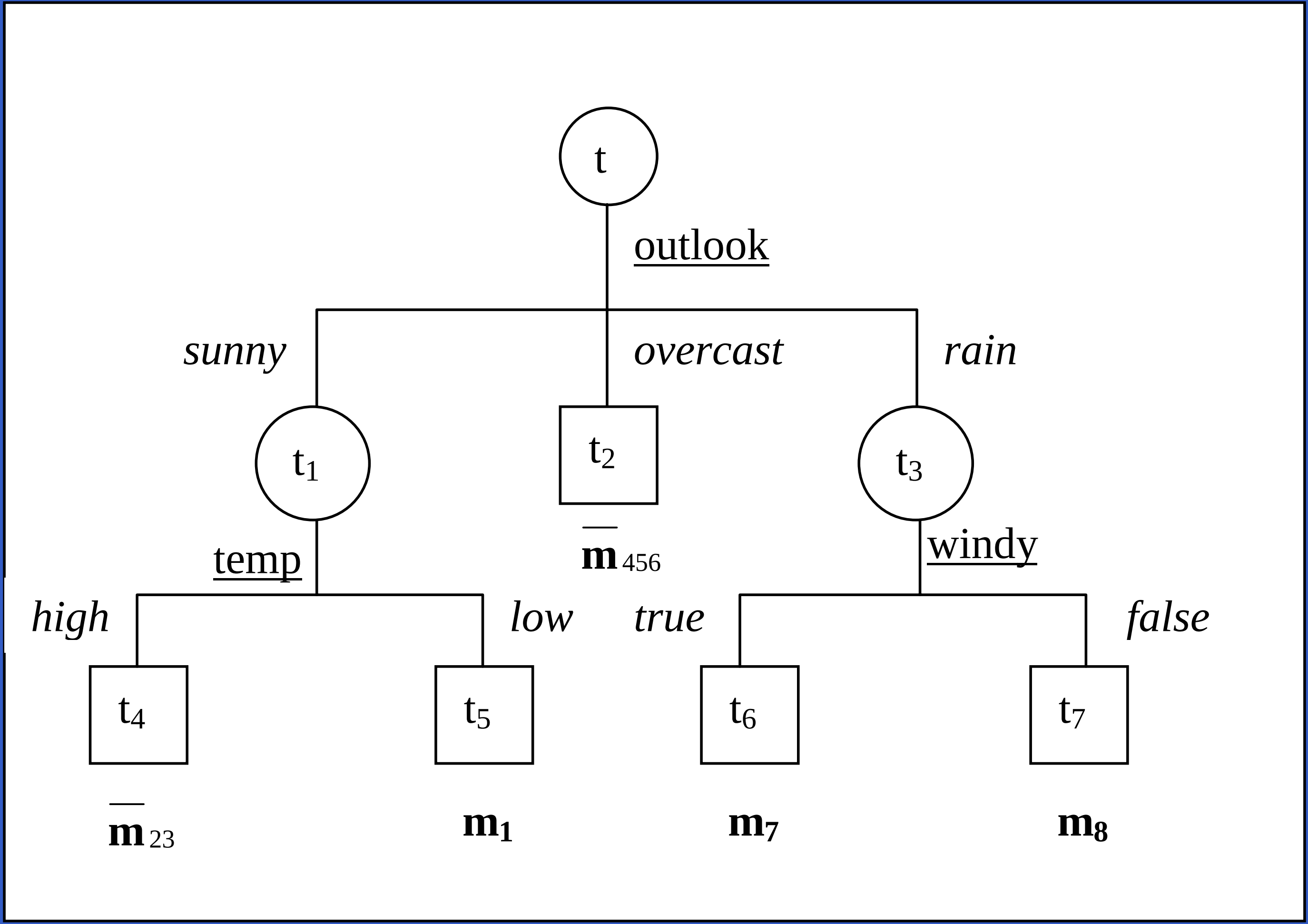


$$I(t) = 0.9995 \quad I(t_1) = 0.9799 \quad I(t_2) = 0.8366 \quad I(t_3) = 0.9097$$

$$\Delta I(\text{outlook}, t) = 0.9995 - (0.98 \times 3/8 + 0.84 \times 3/8 + 0.91 \times 2/8) = 0.090$$

$$\Delta I(\text{temp}, t) = 0.0108 \qquad \Delta I(\text{windy}, t) = 0.072$$

Example: tree

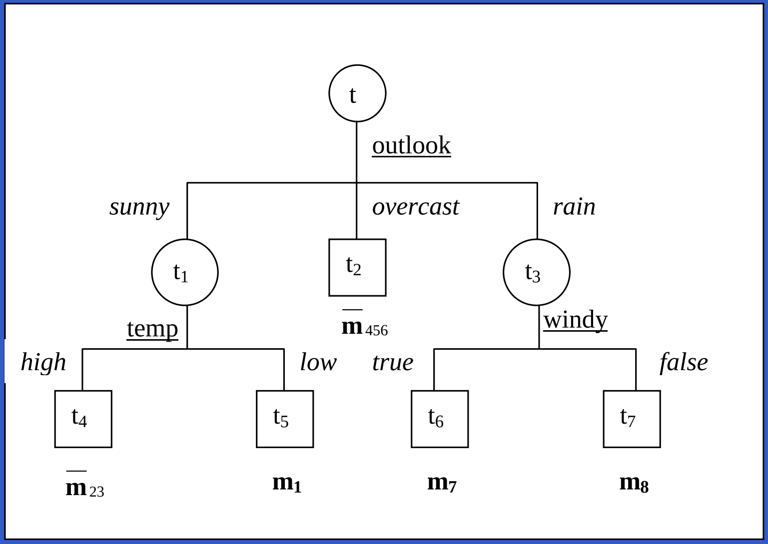


Imprecise/uncertain attribute values

- Disjunctive case: the values of some attributes are only known to belong to subset of values.

Ex: outlook $\neq$ rain,
temp=low
windy =true

$$m = m_1 \textcircled{\cup} \overline{m}_{456}$$



- General case: knowledge about each attribute x_j described by a bba m_{x_j}.

Conclusions

- Extension of decision tree methodology to data with imprecise or uncertain class labels
- Allows to deal with imprecision and uncertainty on attributes in the classification phase.
- Prepruning: discount $\overline{m}$ with reliability factor

$$1 - \alpha = \frac{n(t)}{n(t) + \eta}$$

- Other approaches: conjunctive (Elouedi and Smets, 2000), inferential (Skarstein-Bjanger and Denœux, 2000).