Machine learning

Créée par	
Date de création	@8 novembre 2022
Étiquette	

Given a hypothesis space H, a hypothesis h H is said to overfit the training data if there exists some hypothesis h’ H, such that h has smaller error that h’ over the training instances, but h’ has a smaller error that h over the entire distribution of instances.

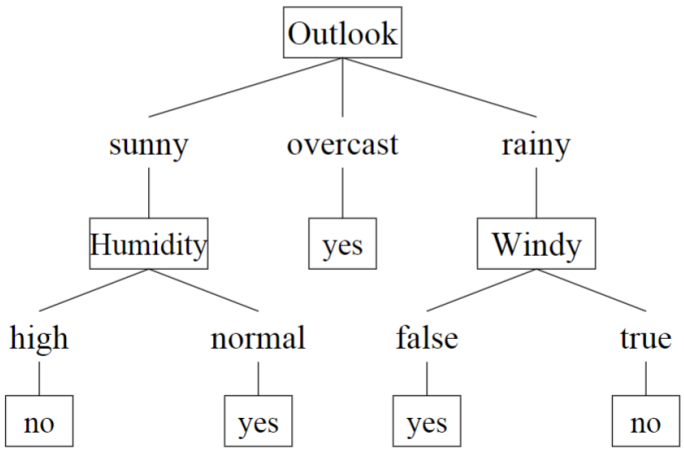
Lecture 2: Decision Trees

Decision Trees:

A decision tree is a tree where:

- Each interior node tests an attribute
- Each branch corresponds to an attribute value
- ach leaf node is labelled with a class (class node)

Example of Decision Tree for playing tennis



Classification with Decision Trees

Classify(x : instance, node : variable containing a node of DT)

- **if** node is classification node then **return** the class of the node
- **else** determine the child of the node that match x. **return** Classify(x ,child)

Entropy

Let **S** be a sample of training examples, and p^+ is the proportion of positive examples in **S** and p^- is the proportion of negative examples in S. Then entropy measures the impurity of **S**:

$$E(S) = -p^+ log_2 p^+ - p^- log_2 p^-$$

Information Gain

Information Gain is the expected reduction in entropy caused by partitioning the instances from S according to a given attribute.

$$E(S) - \sum \frac{|S_v| \times E(S_v)}{|S|}$$

Overfitting

Given a hypothesis space H, a hypothesis h ∈ H is said to overfit the training data if there exists some hypothesis
h' ∈ H, such that h has smaller error that h' over the training instances, but h' has a smaller error that h over the entire distribution of instances.

Implications of Overfitting:

Small number of instances are associated with leaf nodes. In this case it is possible that for coincidental regularities to occur that are unrelated to the actual target concept.

Approaches to Avoiding Overfitting:

- **Pre-pruning**: stop growing the tree earlier, before it reaches the point where it perfectly classifies the training data
- Post-pruning: Allow the tree to overfit the data, and then post-prune the tree.