

Foundations of Machine Learning

Module 2: Linear Regression and Decision Tree

Part B: Introduction to Decision Tree

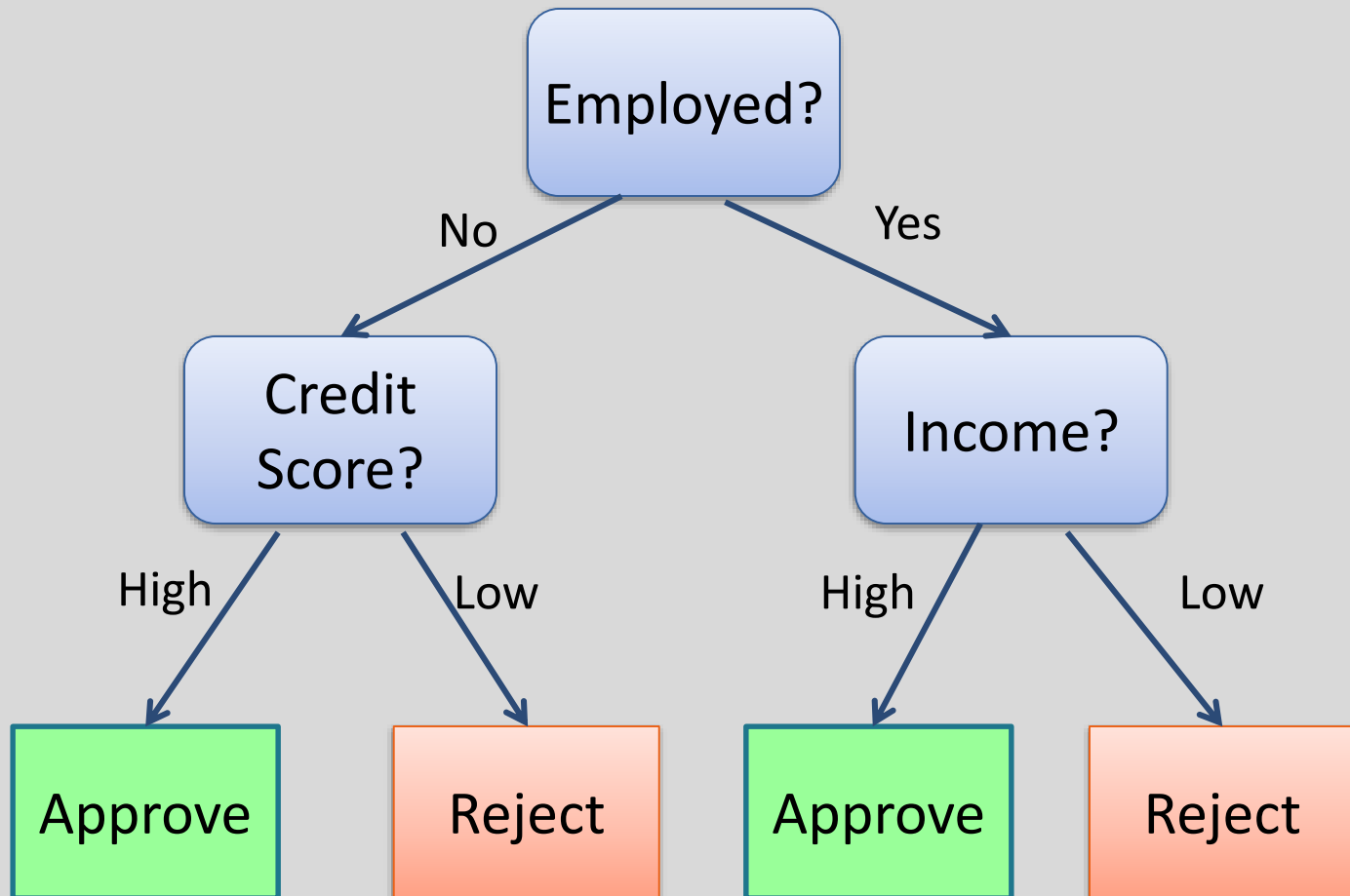
Sudeshna Sarkar
IIT Kharagpur

Definition

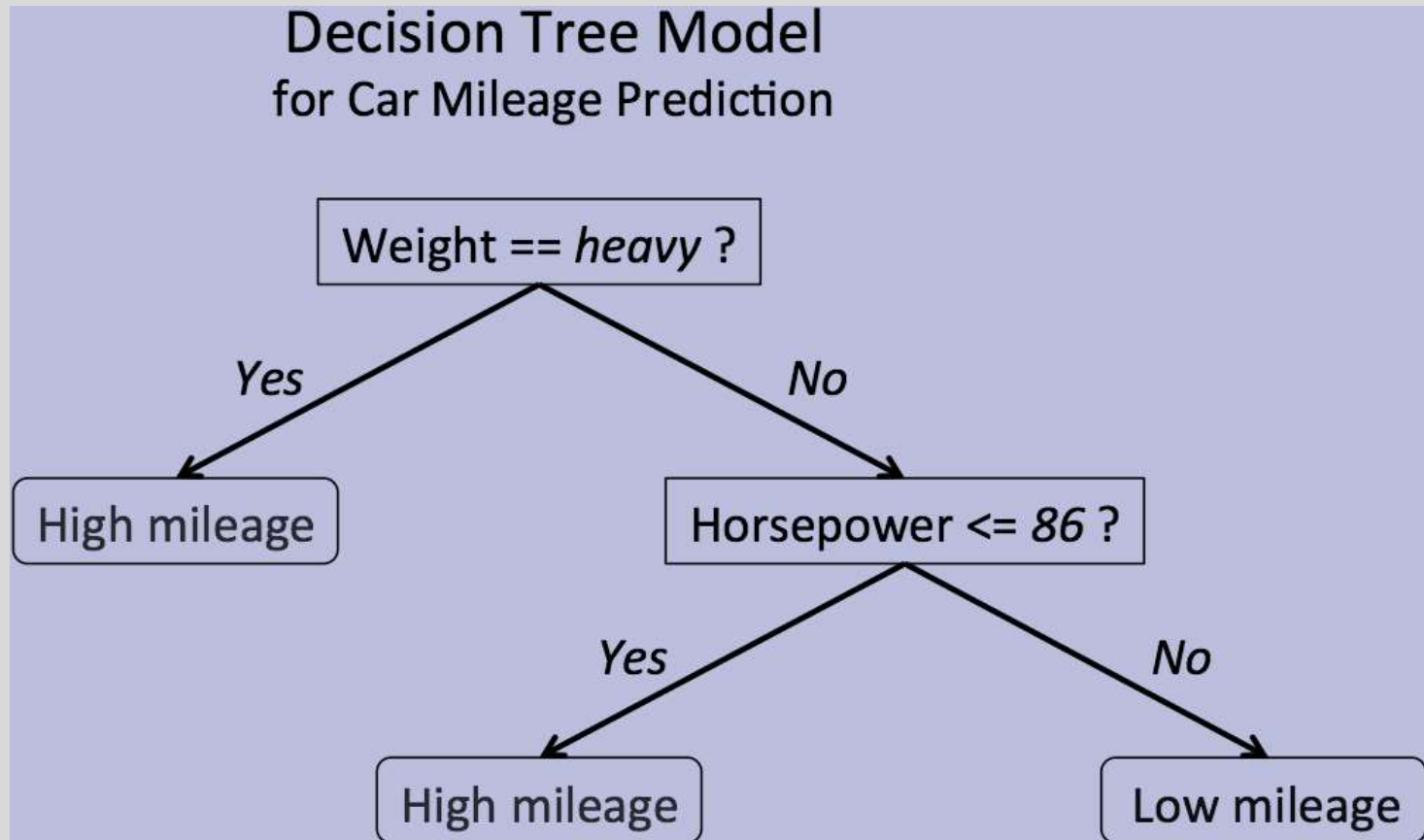
- A decision tree is a classifier in the form of a tree structure with two types of nodes:
 - Decision node: Specifies a choice or test of some attribute, with one branch for each outcome
 - Leaf node: Indicates classification of an example

Decision Tree Example 1

Whether to approve a loan



Decision Tree Example 3



Issues

- Given some training examples, what decision tree should be generated?
- One proposal: prefer the smallest tree that is consistent with the data (Bias)
- Possible method:
 - search the space of decision trees for the smallest decision tree that fits the data

Example Data

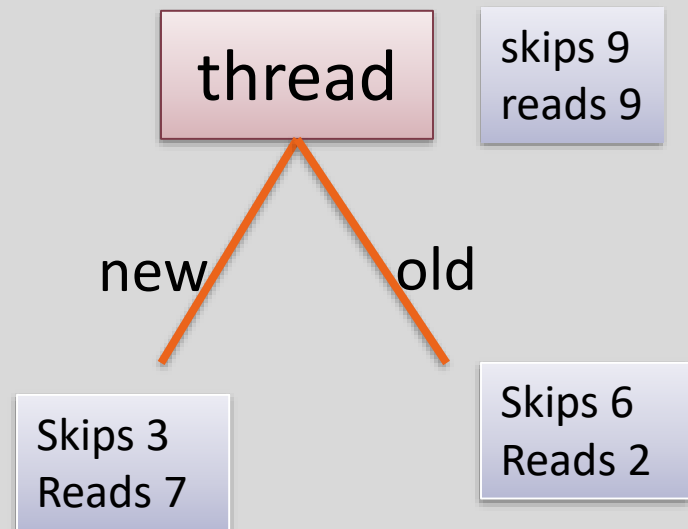
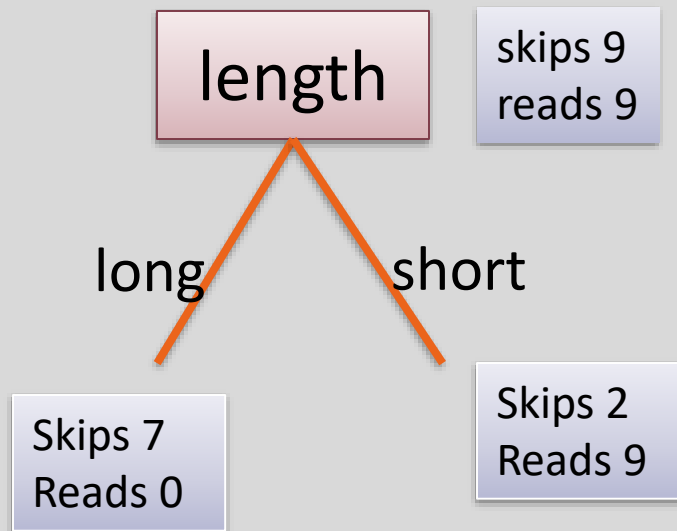
Training Examples:

	Action	Author	Thread	Length	Where
e1	skips	known	new	long	Home
e2	reads	unknown	new	short	Work
e3	skips	unknown	old	long	Work
e4	skips	known	old	long	home
e5	reads	known	new	short	home
e6	skips	known	old	long	work

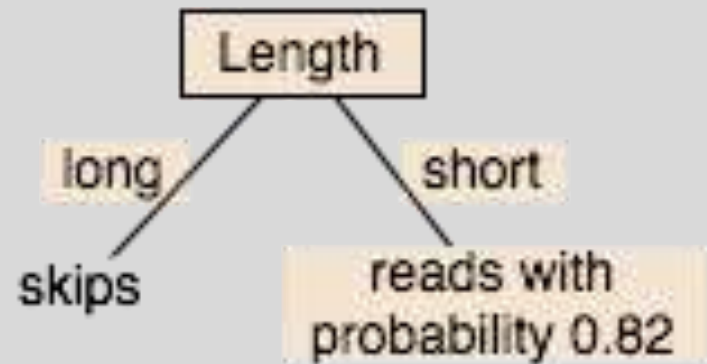
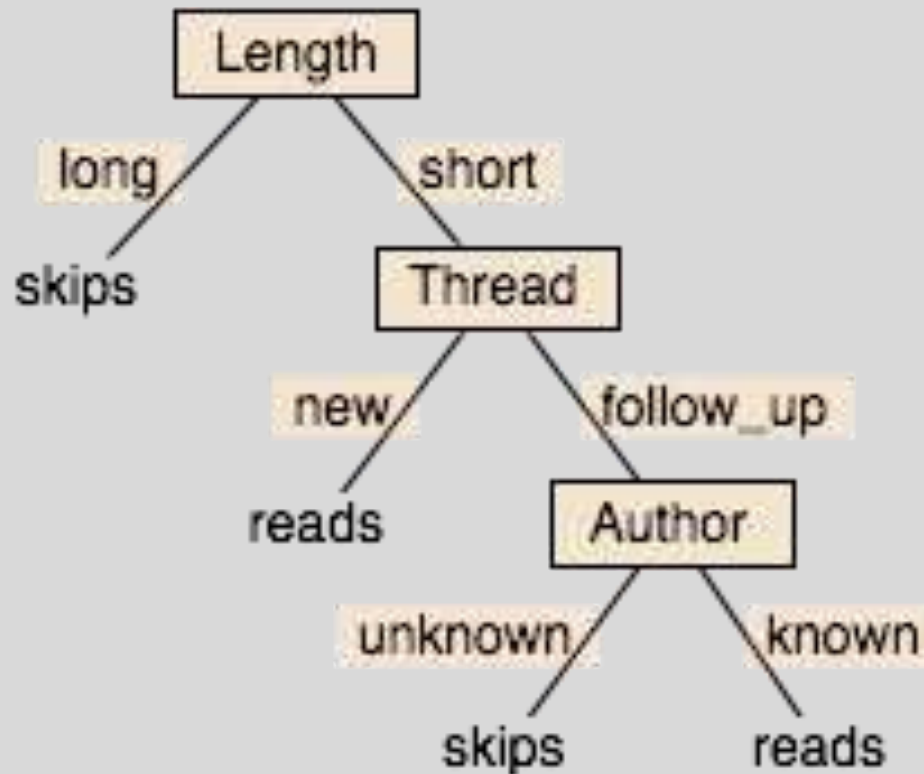
New Examples:

e7	???	known	new	short	work
e8	???	unknown	new	short	work

Possible splits



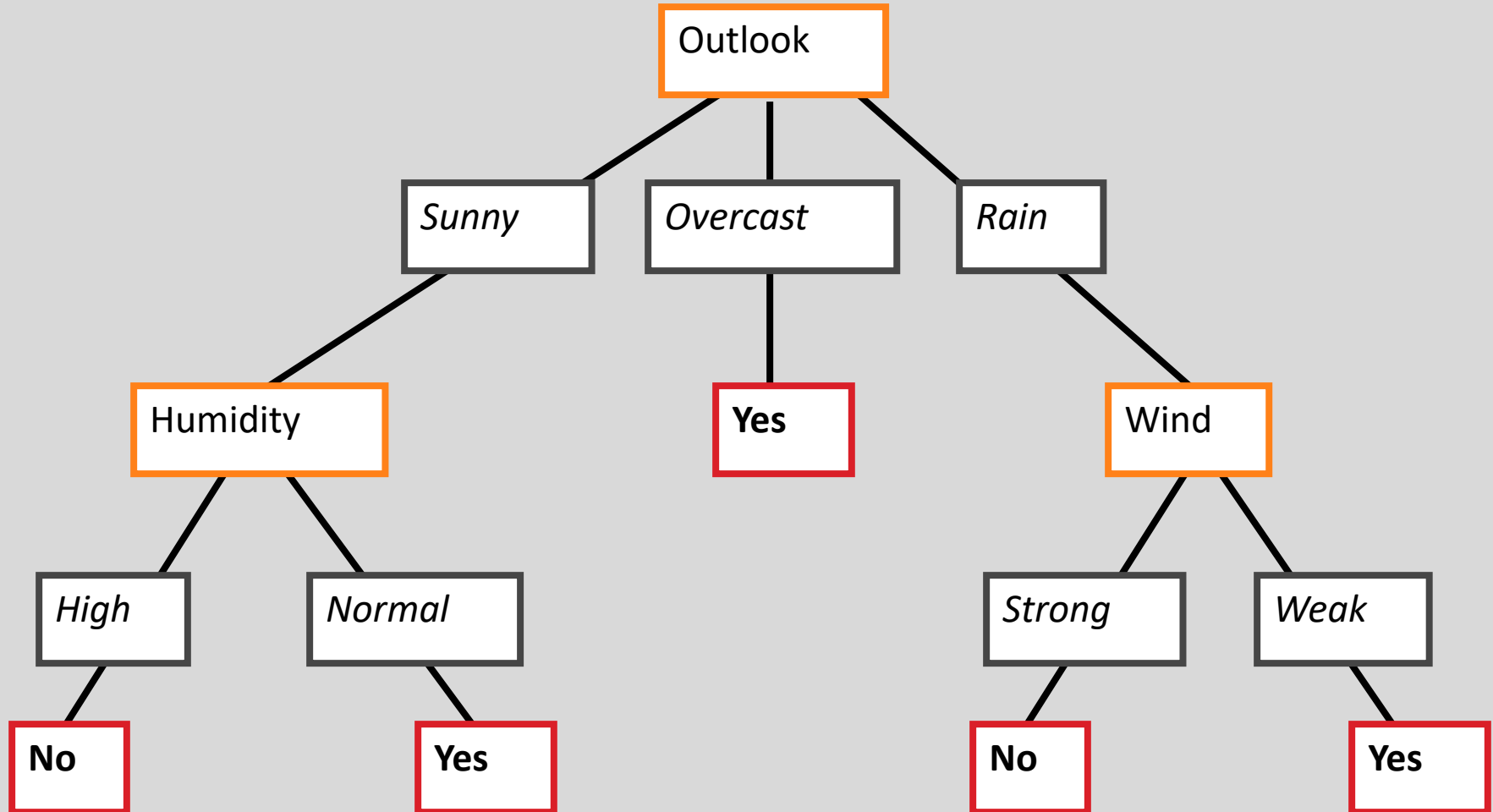
Two Example DTs



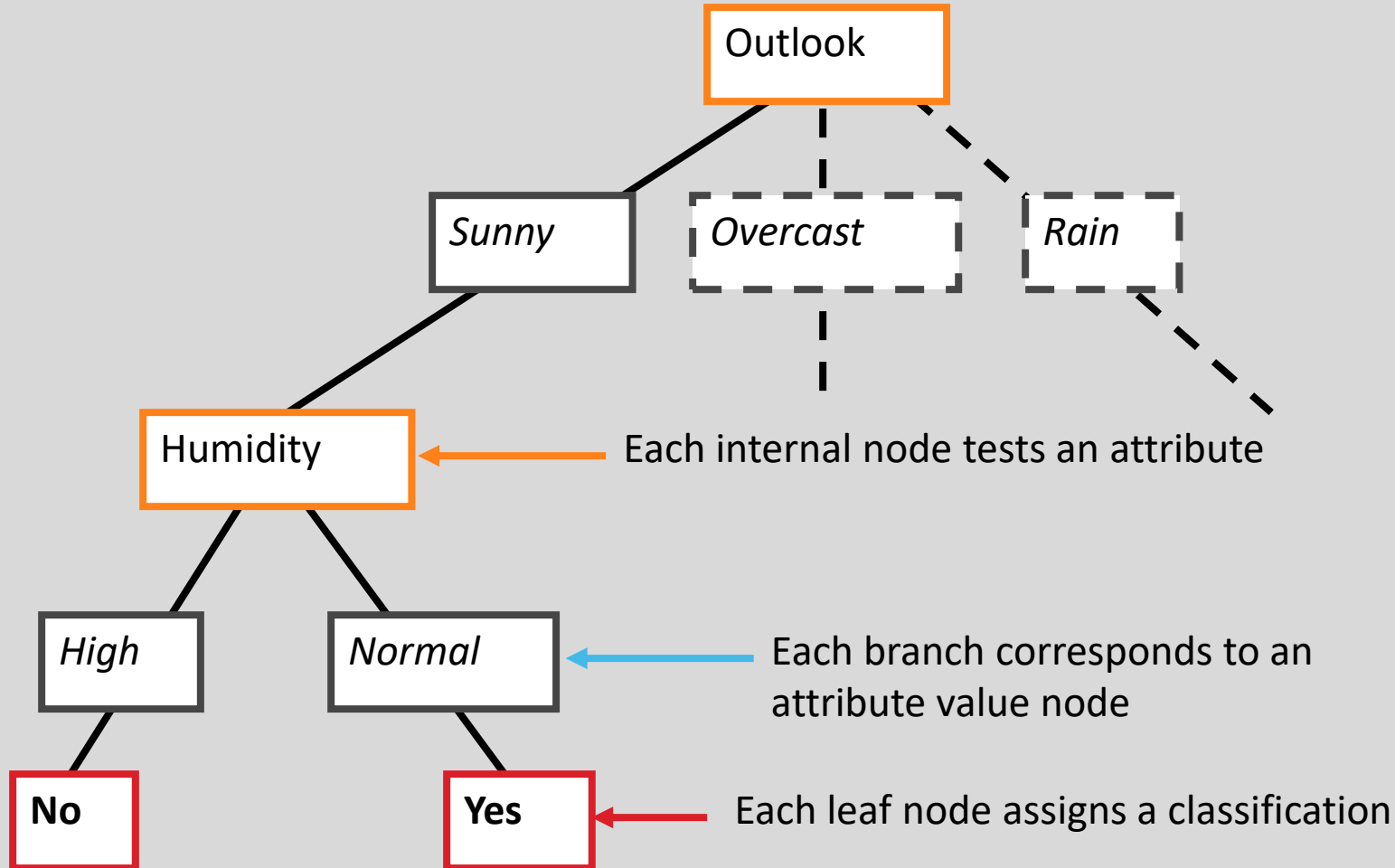
Decision Tree for PlayTennis

- Attributes and their values:
 - Outlook: *Sunny, Overcast, Rain*
 - Humidity: *High, Normal*
 - Wind: *Strong, Weak*
 - Temperature: *Hot, Mild, Cool*
- Target concept - Play Tennis: *Yes, No*

Decision Tree for PlayTennis

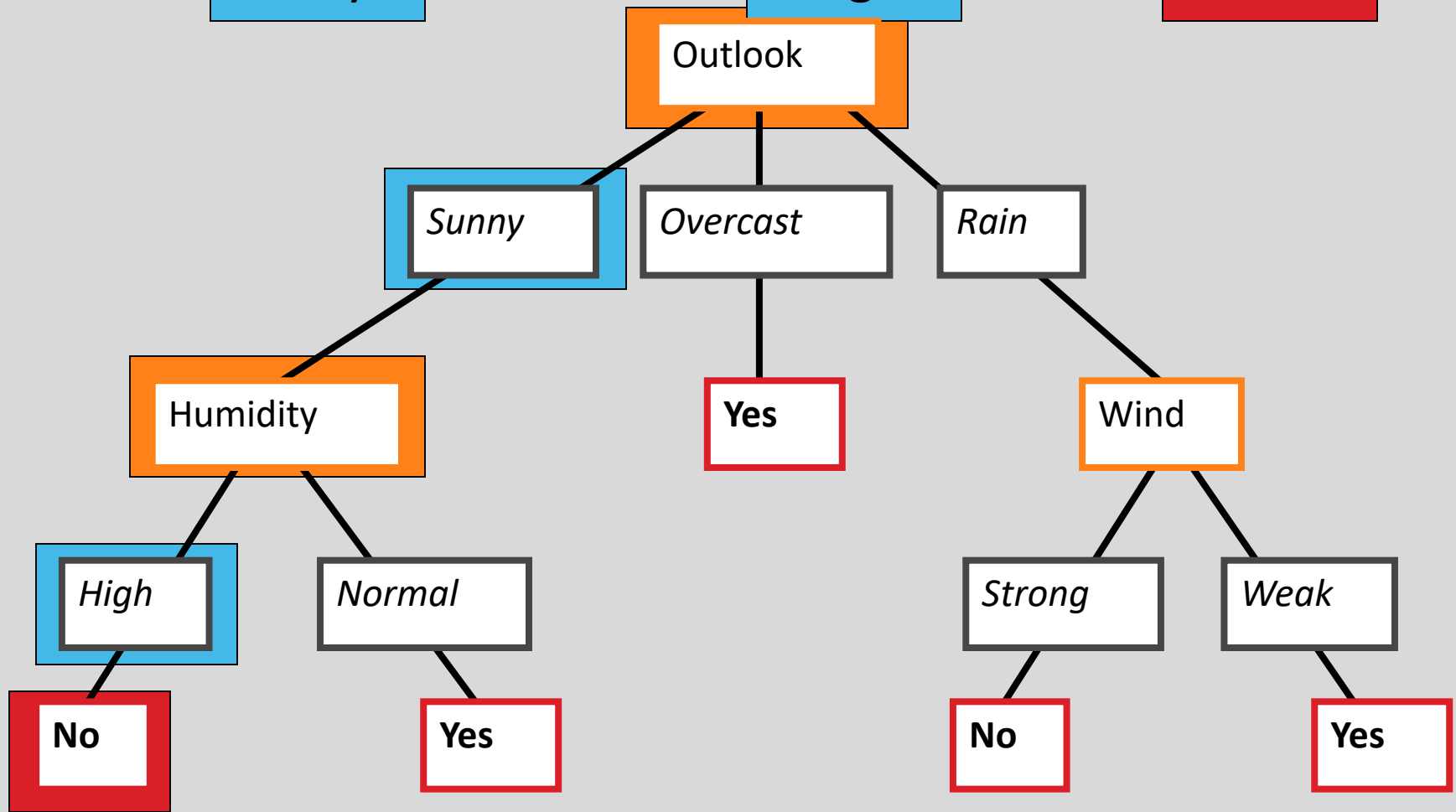


Decision Tree for PlayTennis



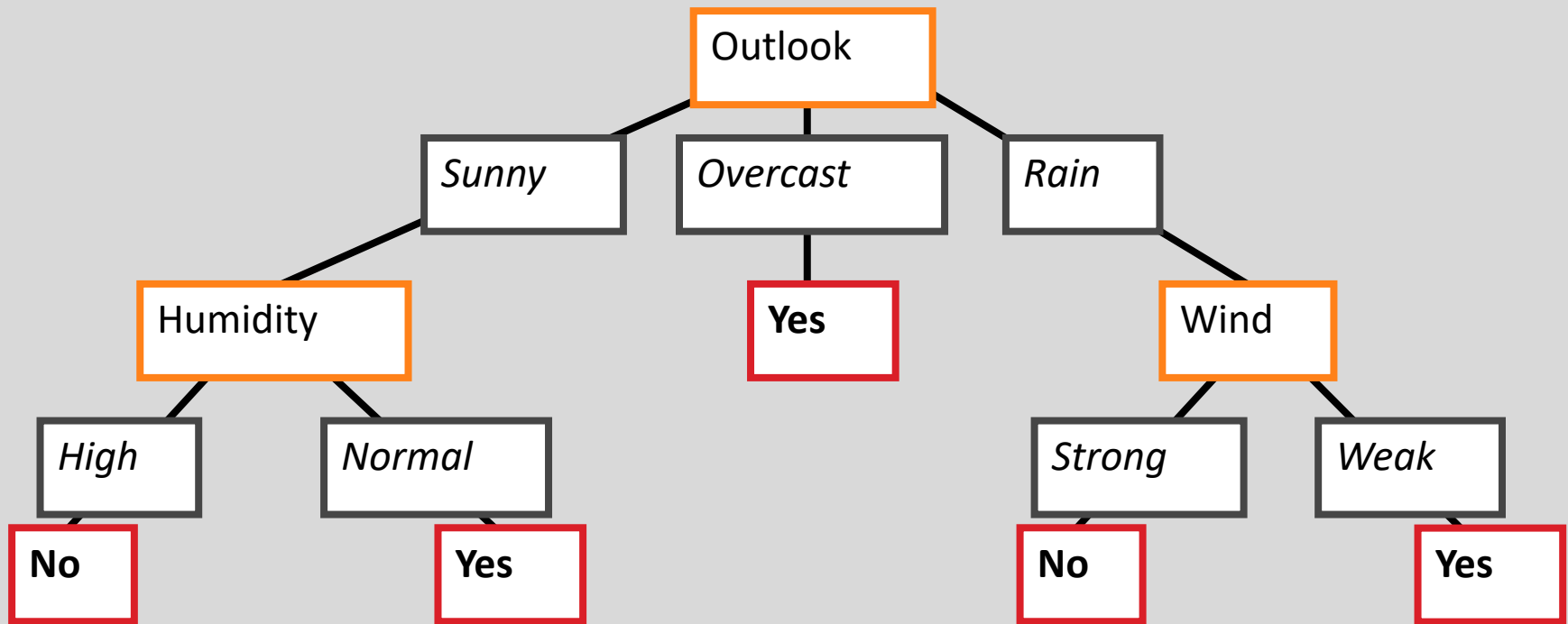
Decision Tree for PlayTennis

Outlook	Temperature	Humidity	Wind	PlayTennis
Sunny	Hot	High	Weak	? No



Decision Tree

decision trees represent disjunctions of conjunctions



(Outlook=Sunny \wedge Humidity=Normal)

✓ (Outlook=Overcast)

✓ (Outlook=Rain \wedge Wind=Weak)

Searching for a good tree

- The space of decision trees is too big for systematic search.
- **Stop** and
 - return the a value for the target feature or
 - a distribution over target feature values
- **Choose** a test (e.g. an input feature) to split on.
 - For each value of the test, build a subtree for those examples with this value for the test.

Top-Down Induction of Decision Trees ID3

1. Which node to proceed with?

1. $A \leftarrow$ the “best” decision attribute for next *node*
2. Assign A as decision attribute for *node*
3. For each value of A create new descendant
4. Sort training examples to leaf node according to the attribute value of the branch
5. If all training examples are perfectly classified (same value of target attribute) stop, else iterate over new leaf nodes.

2. When to stop?

Choices

- When to stop
 - no more input features
 - all examples are classified the same
 - too few examples to make an informative split
- Which test to split on
 - split gives smallest error.
 - With multi-valued features
 - split on all values or
 - split values into half.