LEARNING

- Machine Learning is the study of how to build computer systems that adapt and improve with experience. It is a subfield of Artificial Intelligence and intersects with cognitive science, information theory, and probability theory, among others.

Machine learning is particularly attractive in several real life problem because of the following reasons:

- Some tasks cannot be defined well except by example
- Working environment of machines may not be known at design time
- Explicit knowledge encoding may be difficult and not available
- Environments change over time

learning is widely used in a number of application areas including,

- Data mining and knowledge discovery
- Speech/image/video (pattern) recognition
- Adaptive control
- Autonomous vehicles/robots
- Decision support systems
- Bioinformatics
- WWW

Decision tree learning

- It used in data mining and machine learning, uses a decision tree as a predictive model which maps observations about an item to conclusions about the item's target value.
- More descriptive names for such tree models are **classification trees** or **regression trees**.
- In these tree structures, leaves represent classifications and branches represent conjunctions of features that lead to those classifications.

- In decision theory and decision analysis, a decision tree is a graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility.
- It can be used to create a plan to reach a goal.
- Decision trees are constructed in order to help with making decisions.
- A decision tree is a special form of tree structure.
- Another use of trees is as a descriptive means for calculating conditional probabilities.

- In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making.
- In data mining, a decision tree describes data but not decisions; rather the resulting classification tree can be an input for decision making. This page deals with trees in data mining.

Practical example

- David is the manager of a famous golf club. Sadly, he is having some trouble with his customer attendance.
- There are days when everyone wants to play golf and the staff are overworked.
- On other days, for no apparent reason, no one plays golf and staff have too much slack time.
- David's objective is to optimize staff availability by trying to predict when people will play golf.
- To accomplish that he needs to understand the reason people decide to play and if there is any explanation for that.
- He assumes that weather must be an important underlying factor, so he decides to use the weather forecast for the upcoming week.

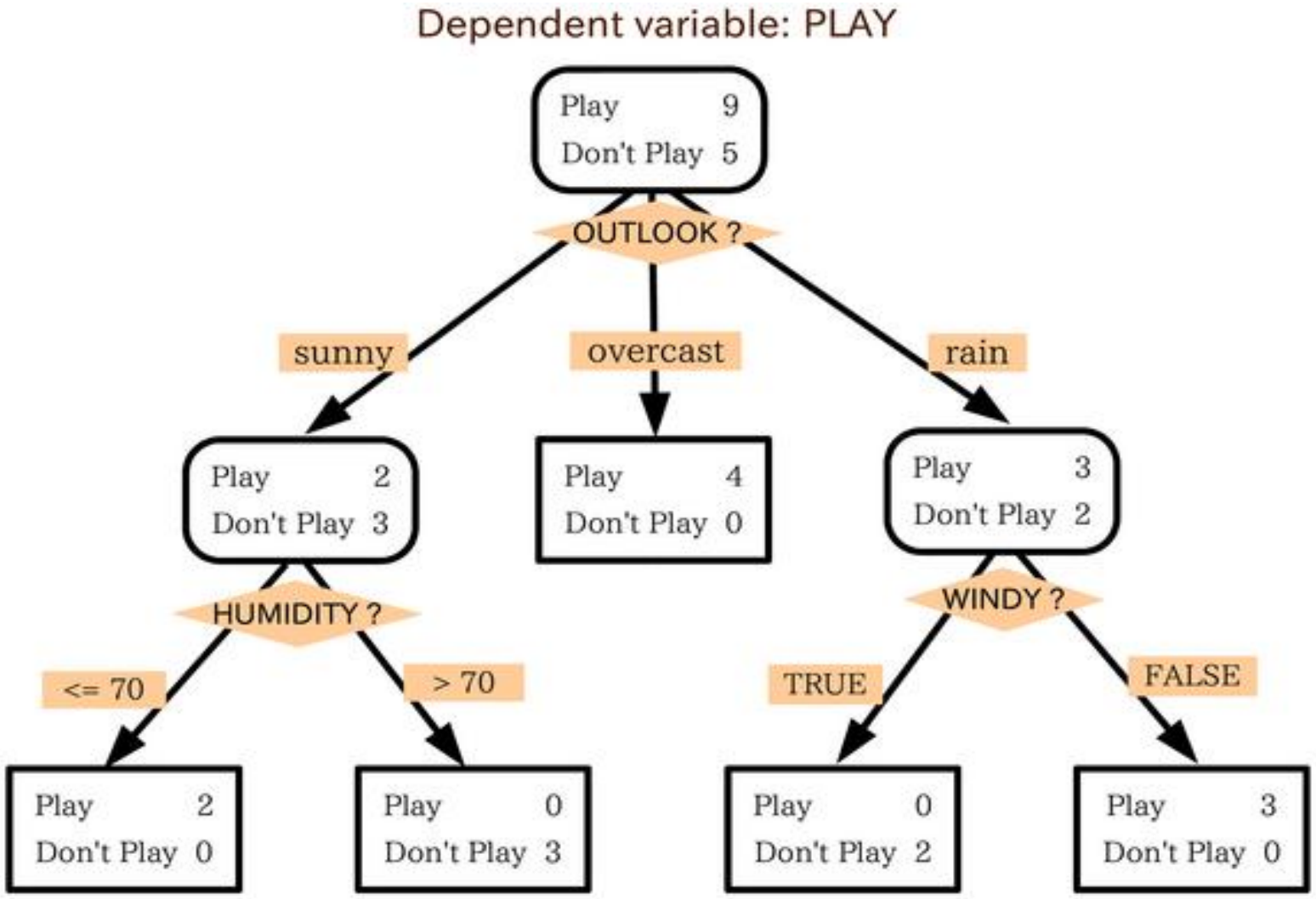
So during two weeks he has been recording:

- The outlook, whether it was sunny, overcast or raining.
- The temperature (in degrees Fahrenheit).
- The relative humidity in percent.
- Whether it was windy or not.
- Whether people attended the golf club on that day.

David compiled this dataset into a table containing 14 rows and 5 columns as shown below.

Play golf dataset

Independent variables				Dep. var
OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY
sunny	85	85	FALSE	Don't Play
sunny	80	90	TRUE	Don't Play
overcast	83	78	FALSE	Play
rain	70	96	FALSE	Play
rain	68	80	FALSE	Play
rain	65	70	TRUE	Don't Play
overcast	64	65	TRUE	Play
sunny	72	95	FALSE	Don't Play
sunny	69	70	FALSE	Play
rain	75	80	FALSE	Play
sunny	75	70	TRUE	Play
overcast	72	90	TRUE	Play
overcast	81	75	FALSE	Play
rain	71	80	TRUE	Don't Play


Dependent variable: PLAY
Play 9
Don't Play 5
OUTLOOK ?
sunny
overcast
rain
Play 2
Don't Play 3
Play 4
Don't Play 0
Play 3
Don't Play 2
HUMIDITY ?
WINDY ?
<= 70
> 70
TRUE
FALSE
Play 2
Don't Play 0
Play 0
Don't Play 3
Play 0
Don't Play 2
Play 3
Don't Play 0

- A decision tree is a model of the data that encodes the distribution of the class label (again the Y) in terms of the predictor attributes.
- It is a directed acyclic graph in form of a tree.
- The top node represents all the data.
- The classification tree algorithm concludes that the best way to explain the dependent variable, play, is by using the variable "outlook".
- Using the categories of the variable outlook, three different groups were found:

- One that plays golf when the weather is sunny,
- One that plays when the weather is cloudy, and
- One that plays when it's raining.

David's first conclusion: if the outlook is overcast people always play golf, and there are some fanatics who play golf even in the rain. Then he divided the sunny group in two. He realized that people don't like to play golf if the humidity is higher than seventy percent.

- Finally, he divided the rain category in two and found that people will also not play golf if it is windy.

- And lastly, here is the short solution of the problem given by the classification tree:
- David dismisses most of the staff on days that are sunny and humid or on rainy days that are windy, because almost no one is going to play golf on those days.
- On days when a lot of people will play golf, he hires extra staff.
- The conclusion is that the decision tree helped David turn a complex data representation into a much easier structure.

Bayesian networks

These are also called *Belief Networks* or *Probabilistic Inference Networks*. Initially developed by Pearl (1988).

The basic idea is:

- Knowledge in the world is *modular* -- most events are conditionally independent of most other events.
- Adopt a model that can use a more local representation to allow interactions between events that *only* affect each other.
- Some events may only be *unidirectional* others may be *bidirectional* -- make a distinction between these in model.
- Events may be causal and thus get chained together in a network.

Implementation

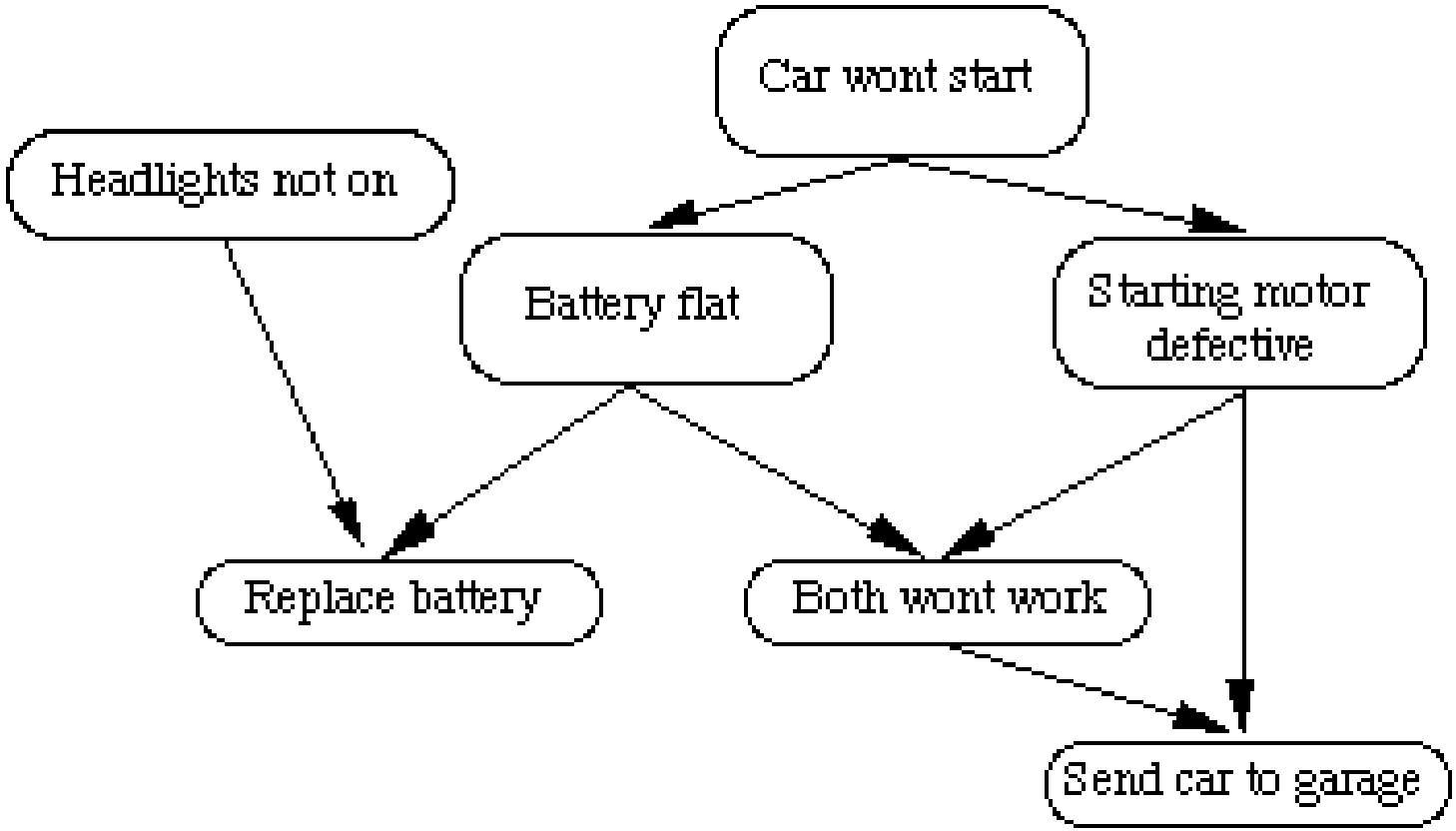
- **A *Bayesian Network* is a *directed acyclic graph*:**
  - A graph where the directions are links which indicate dependencies that exist between nodes.
  - Nodes represent propositions about events or events themselves.
  - Conditional probabilities quantify the strength of dependencies.

- Consider the following example:
- The probability, that my car won't start.
- If my car won't start then it is likely that
  - The battery is flat or
  - The staring motor is broken.

In order to decide whether to fix the car myself or send it to the garage I make the following decision:

- If the headlights do not work then the battery is likely to be flat so i fix it myself.
- If the starting motor is defective then send car to garage.
- If battery and starting motor both gone send car to garage.
- The network to represent this is as follows:

Reasoning in Bayesian nets



Explaination based learning

- *Explanation based learning (EBL)* uses an explicit **domain theory** to construct an **explanation** or **proof** of a training example.
- By then **generalizing** from the proof, new knowledge is acquired that can be applied in non-training situations.
- This differs from **inductive** learning in that the domain theory implies the new knowledge. It is sometimes called **deductive learning** or **analytic learning**.

Example: Learning When an Object is a Cup

- **Target Concept:** cup(C) :- premise(C).
- **Domain Theory:**
- cup(X) :- liftable(X), holds_liquid(X).
- holds_liquid(Z) :- part(Z,W), concave(W),points_up(W).
- liftable(X) :- light(X), part(Y,handle).
- light(X) :- small(X).
- light(X) :- made_of(X,feathers).

Rote learning

- Rote learning is a learning technique which avoids understanding of a subject and instead focuses on memorization.
- The major practice involved in rote learning is learning by repetition.
- The idea is that one will be able to quickly recall the meaning of the material the more one repeats it.

- Rote learning is widely used in the mastery of foundational knowledge.
- Examples include, phonics in reading, the periodic table in chemistry, multiplication tables in mathematics, anatomy in medicine, cases or statutes in law, basic formulas in any science, etc.
- Rote learning, by definition, eschews comprehension, however, and consequently,
- it is an ineffective tool in mastering any complex subject at an advanced level.

- However, rote learning is still useful in passing exams.
- If exam papers are not well designed, it is possible for someone with good memorization techniques to pass the test without any meaningful comprehension of the subject.
- However, learning the context of a particular topic can make the subject more memorable.