Bioinformatics Technology "InfoLife"

Introduction to Machine Learning

Bioinformatics Laboratory "InfoLife"

University of Foggia - C.IN.I. Consortium

Prof. Crescenzio Gallo

crescenzio.gallo@unifg.it

What is Machine Learning?

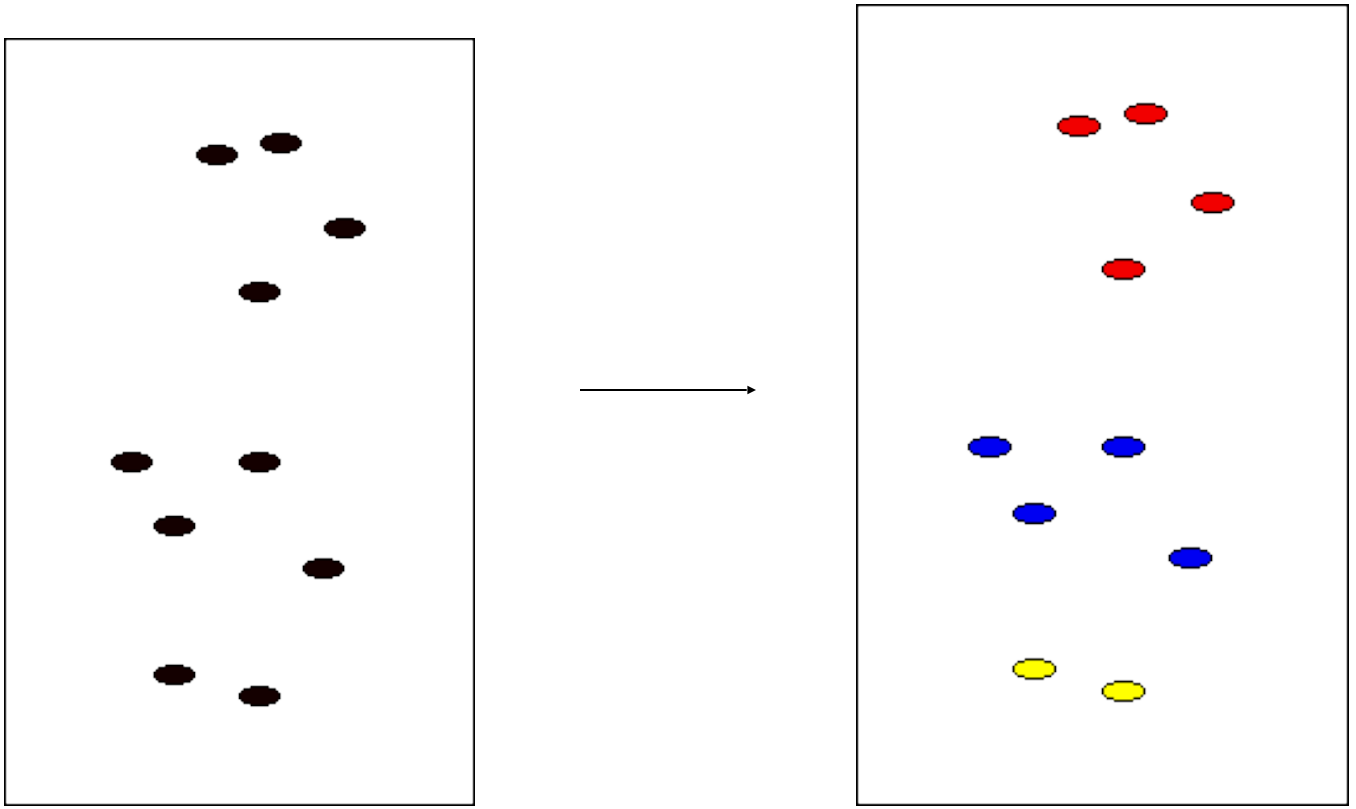
- Is part of Artificial Intelligence
- Relates to particular activities, without the goal of building intelligent automata
- Relates to the design of systems that improve (or at least change) as they gain knowledge or experience

Typical ML activities

- Clustering
- Classification
- Categorization
- Filter/Selection
- Pattern recognition
- Game simulation
- Autonomous behavior

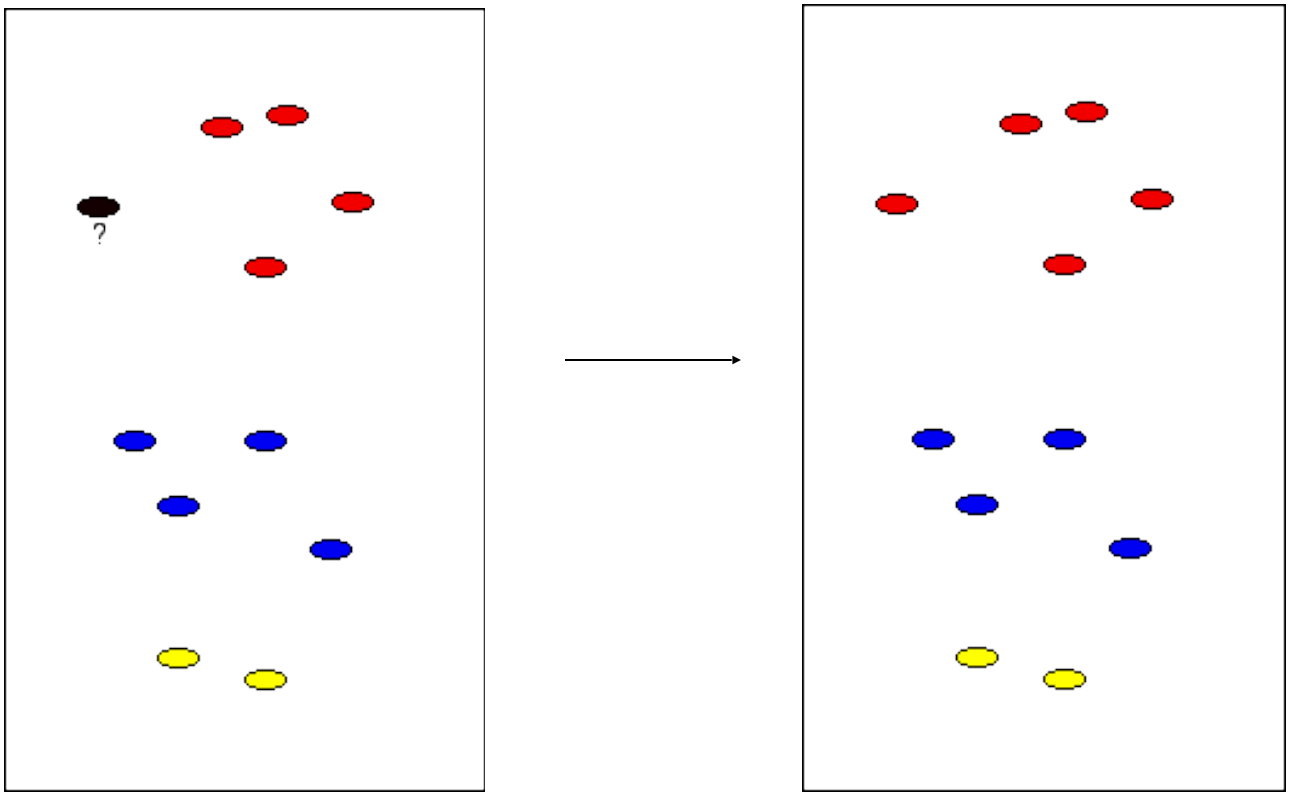
Typical ML activities

- Clustering



Typical ML activities

- Classification/Categorization



Typical ML activities

- Recognition (of images)

Vincent Van Gogh

Joe Di Maggio

Mohammed Ali

Steve Jobs

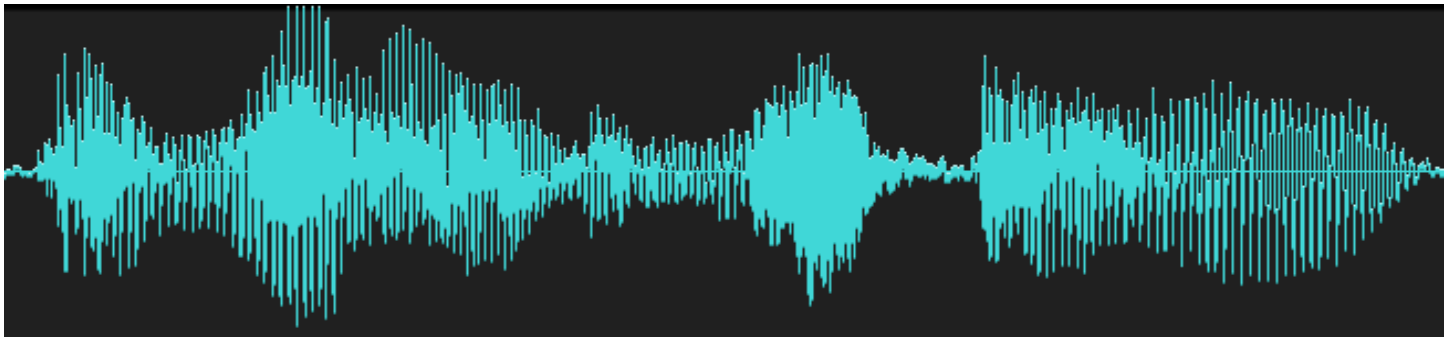
Bill Gates

Winston Churchill

Alfred Hitchcock

Typical ML activities

- Recognition (of sounds)



Mr Tambourine man

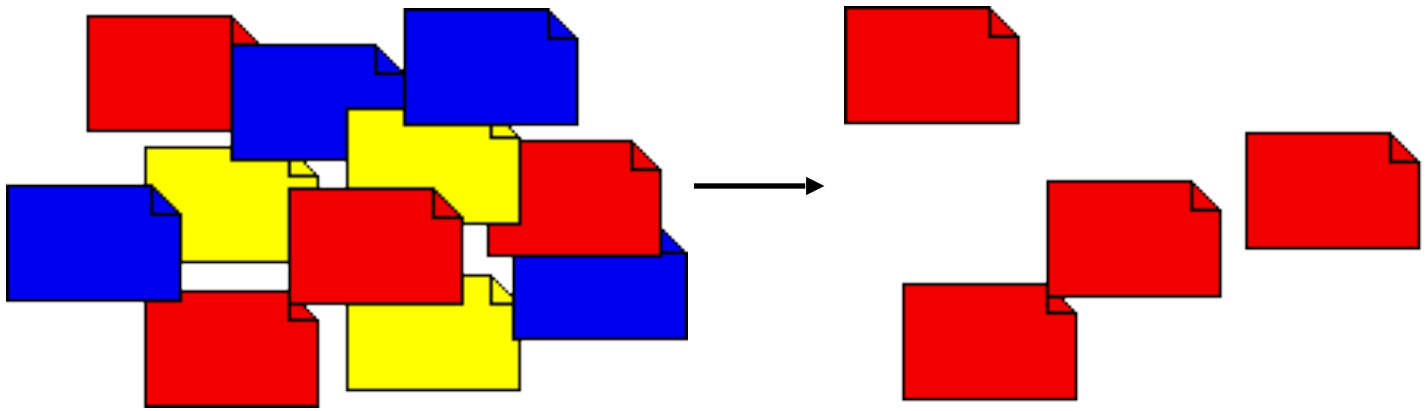
La Canzone del padre

Teach your children

Il suonatore Jones

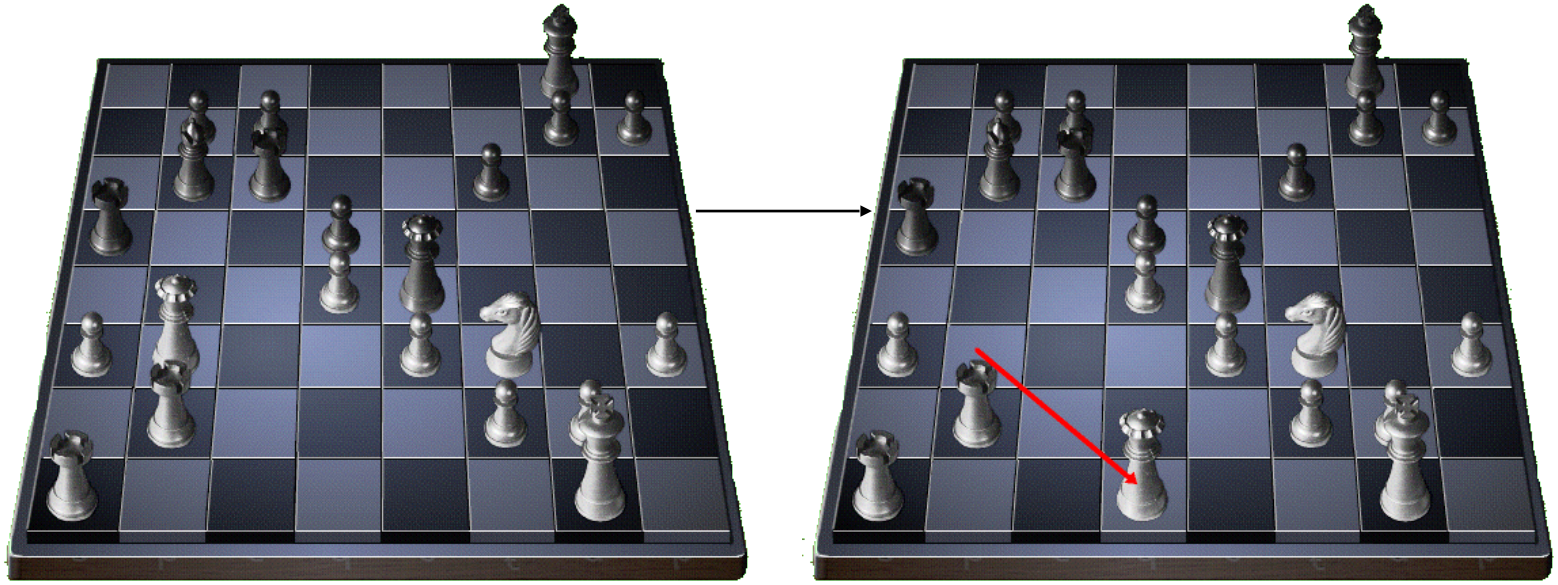
Typical ML activities

- Filter/Selection



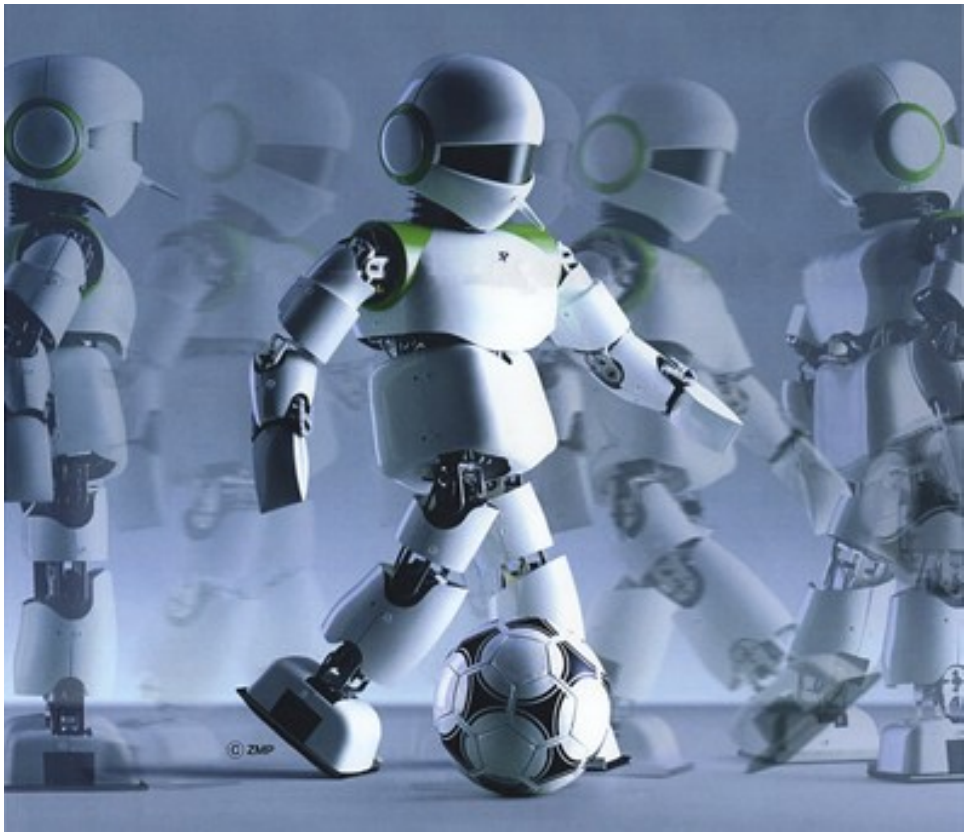
Typical ML activities

- Game simulation



Typical ML activities

- (Learning and) Autonomous behavior



ML Keywords

- Data Mining
- Knowledge Management (KM)
- Information Retrieval
- Expert Systems
- Text Mining (identification and control of topics in texts and social networks)

Who is in charge of ML?

- Research and industry
- In close synergy with each other
- Not many reusable ML/KM modules, outside of a few commercial systems
- KM is seen as a key component of large enterprise strategy
- ML is an extremely active research area with low "cost of entry"

When is ML useful?

- When there are large amounts of data
- When there aren't enough people to solve a problem, or those available can't operate more quickly
- When you can afford to be wrong in some cases (efficiency vs. accuracy)
- When you need to find *patterns*
- When you have nothing left to lose...

Theory and ML Terminology

- The ML is related to an objective function (*target*) linked to a set of examples (cases, instances)
- The objective function is often called the hypothesis
- Example: with a Neural Network, a *trained network* is a hypothesis
- The set of all possible objective functions is called the *space of hypotheses*
- The training process can be viewed as a search in the space of hypotheses

Theory and ML Terminology

- ML techniques
  - they **exclude** some hypotheses
  - they **prefer** some hypotheses regarding others
- The exclusion and preference rules of an ML technique constitute its capacity (*bias*) for induction
- If a technique is not "biased" (i.e. with *good* rules), cannot learn and therefore **generalize**
- Example: children learning to do multiplication (understanding instead of learning by heart: how much damage is done in elementary school…)

Theory and ML Terminology

- Ideally, an ML technique
  - must not exclude the "correct" hypothesis, i.e. the hypothesis space must include the objective hypothesis;
  - prefers the objective hypothesis over the others.
- Measuring the degree to which these criteria are met is important and sometimes complicated

Hypothesis Evaluation

- Often the knowledge of the "goodness" of a hypothesis:
  1. Is needed to know how it will behave in the real world;
  2. Can be used to improve the learning technique or to refine its parameters;
  3. It can be used by a learning algorithm to automatically improve the hypothesis.
- Usually the evaluation is done on test data
  - Test data must be separated from the data used in the learning phase;
  - The test data used is usually known as validation or hold-out data;
  - Training, validation and test data should always be different and not mixed together.

Hypothesis Evaluation

- Some statistical measures used are: error rate, accuracy, precision (specificity), recall (sensitivity, i.e. completeness), F_1 (combines precision and recall)
- They are calculated by means of contingency tables

		Real	
		Yes	No
Predicted	Yes	a	b
	No	c	d

a = True Positive (TP)

b = False Positive (FP)

c = False Negative (FN)

d = True Negative (TP)

Hypothesis Evaluation

Pred-icted \ *Real*	Yes	No
Yes	a	b
No	c	d

- Error rate = $(b+c)/(a+b+c+d)$
- Accuracy = $(a+d)/(a+b+c+d)$
- Precision = $p = a/(a+b)$
- Recall = $r = a/(a+c)$
- $F_1 = 2 \cdot p \cdot r/(p+r)$

- Precision = Fraction of "true positives" among cases found positive
- Recall = Fraction of "true positives" found compared to all true positives
- F_1 combines precision and recall for a more comprehensive evaluation

Hypothesis Evaluation

		Real Yes	*Real* No
Predicted	Yes	2663	467
Predicted	No	1081	2675

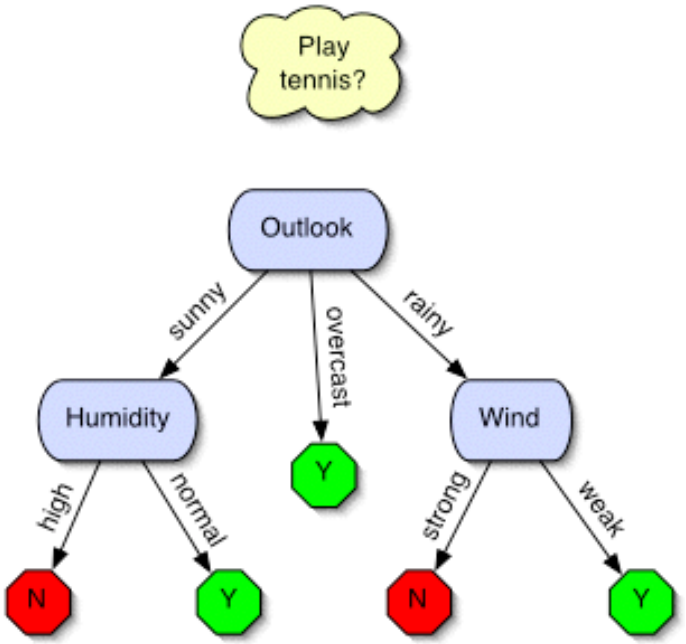
- Example (of classification)
- Note that *precision* is higher than *recall* – this indicates a "cautious" classifier

Precision = 0.851, Recall = 0.711, $F_1 = 0.775$

These values depend on the type of classification (and number of classes), and should not be compared with other classifications.

It is often useful to compare classes separately, then average (*macro-averaging*) them.

Decision Trees



- Conceptually simple
- Rapid evaluation
- Examinable decision structures
- Can *learn* from training data
- Can be difficult to construct
- Possible "overfit" of training data
- Simpler, i.e., *smaller* decision trees are usually preferred

Decision Trees

- Example of training data (the well-known dataset "**Play tennis?**"):

Outlook	Temp	Humid	Wind	Play?
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
		. . .		

Decision Trees

- How do we build the tree from the training data?
- We want to get trees that are as small as possible.
- Which attribute (Outlook, Wind, etc.) is the best classifier?
- We need a measure of how much an attribute contributes to the final classification result.
- We use *Information Gain* (IG) for this purpose, which is based on the *entropy* of the training data.
- The attribute with the highest IG is the "most useful" classifier because it provides the greatest reduction in entropy.

Decision Trees

$$Entropy = \sum_i p_i \log_2 \left(\frac{1}{p_i} \right)$$

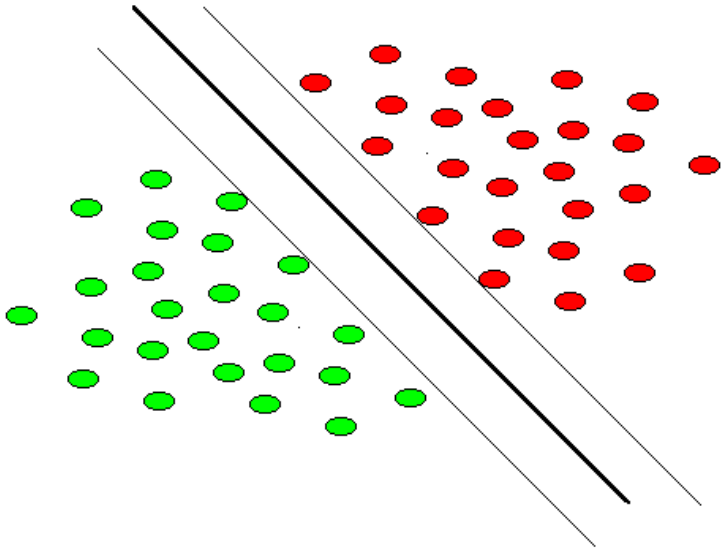
- Derived from Claude Shannon's Information Theory.
- Measures the uncertainty of a decision between alternative options.
- It is represented by the expected value (in a probabilistic sense) of the number of bits required to specify the value of an attribute.
- i represents the value of an attribute I; p_i represents the probability (frequency) of that value.

Decision Trees

$$Gain(S,I) = Entropy(S) - \sum_{i \in I} \frac{|S_i|}{|S|} Entropy(S_i)$$

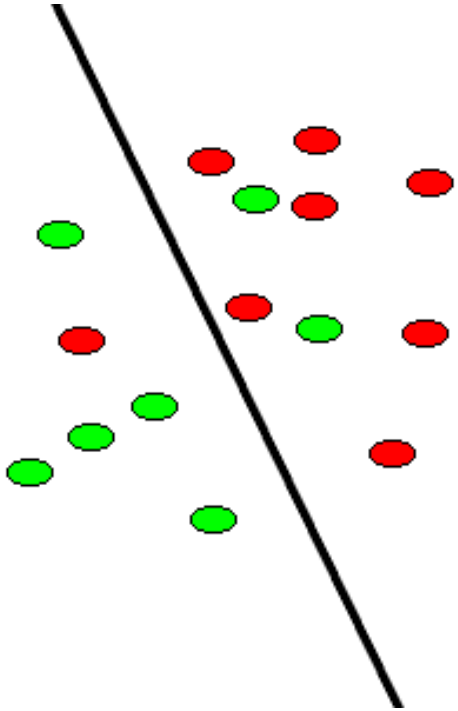
- The S_i are the subsets of S in which the attribute I has value i
- IG is the original entropy minus the entropy given the value i
- At each decision node (*splitting node*) is found $\text{argmax}_I (Gain(S, I))$
- To maximize IG is enough to minimize the second term on the right, being $Entropy(S)$ constant
- This is known as the "ID3 algorithm" (J. R. Quinlan, 1986)

Support Vector Machine (SVM)



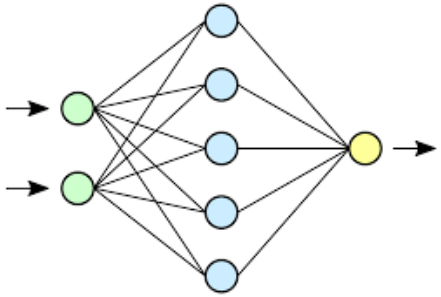
- Another ML technique
- Measure features (attributes) quantitatively by inducing a vector space
- Find the *optimal decision surface*

Support Vector Machine (SVM)



- Data may not be separable
- The same algorithms usually work to find the "best" separating surface
- Different shapes of surfaces can be used
- Usually scales well with number of features, not so well with number of cases

Artificial Neural Network (ANN)



- A neural network is an artificial representation of the human brain that attempts to simulate the learning process.
- Like the human brain, an **artificial neural network** consists of neurons and the connections between them.
- Neurons carry incoming information on their outgoing connections to other neurons. In neural network terminology these connections are called **weights**.
- The "electrical" information is simulated with specific values stored in the weights.
- The change in the connection structure can be simulated simply by changing the values of the weights.

Thanks for your attention!