CHAPTER FIVE

Machine Learning Basics

WHAT IS LEARNING

- Learning is memorizing and remembering
  - Telephone number
  - Reading textbook
  - Understanding Language (memorizing grammar & practicing)
  - Recognize face, signature & fraudulent credit card transactions
- Learning is improving *motor* skills
  - Riding a bike
  - Exercising and practicing the idea
- Learning is understanding the strategy & rule of the game
  - Playing chess and football
- Learning is abstraction and exploration
  - Develop scientific theory
  - Undertaking research

CONT…

- **Learning** is any process by which a system improves performance from experience
- Learning takes place as a result of the interaction between
  - the agent and
  - the world,
  - And from observation by the agent of its own decision-making processes.
- What is the task?
  - Classification
  - Problem solving /planning /control

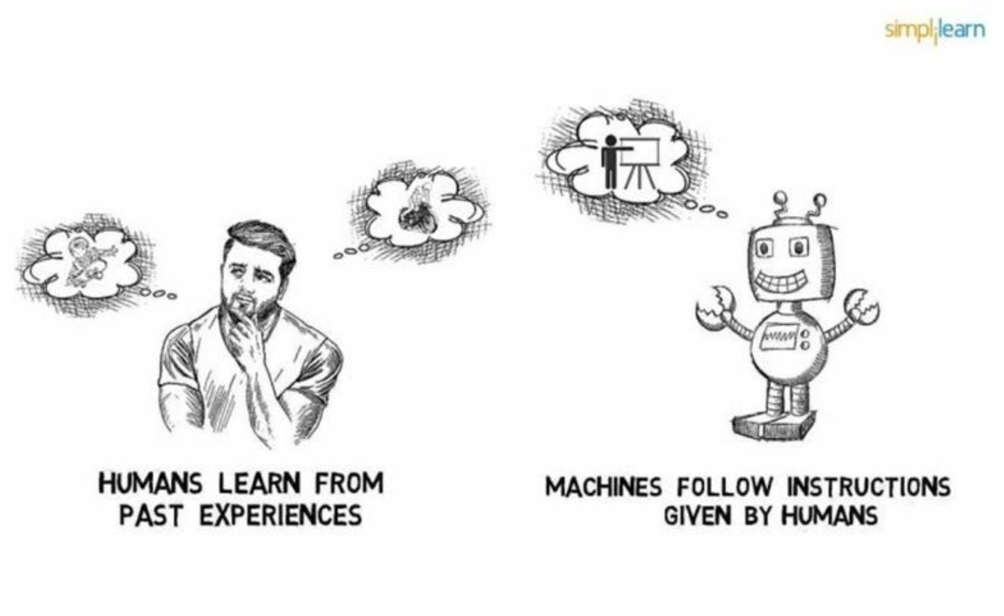
CONT...

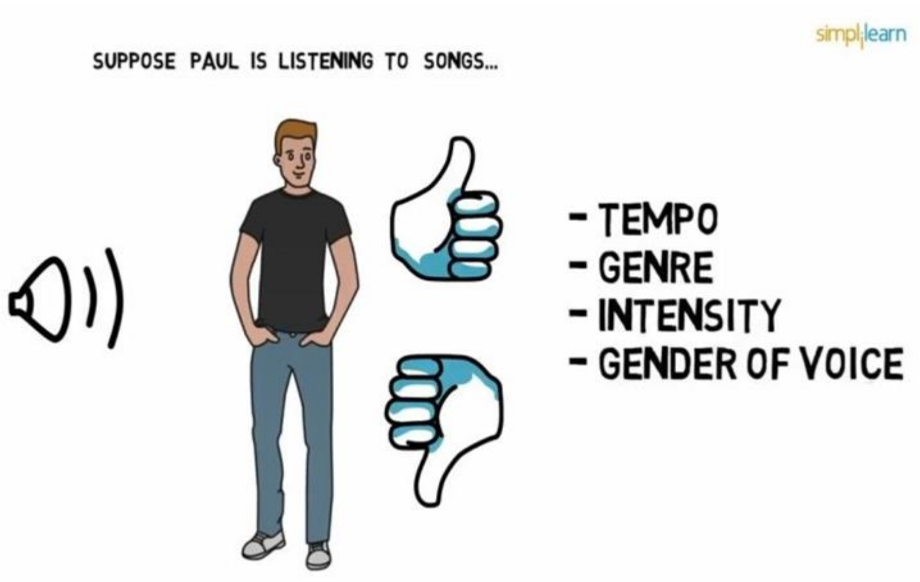
- Learning is one of the keys to human intelligence. Do you agree?
- The idea behind learning is that percepts should not only be used for acting now, but also for improving the agent’s ability to act in the future.
  - Learning is essential for unknown environments, i.e. when the agent lacks knowledge. It enables to organize new knowledge into general, effective representations
  - Learning modifies the agent's decision-making mechanisms to improve performance
- Learning is nothing but
  - Feature extraction
  - Classification
- Learning modifies the agent's decision-making mechanisms to improve performance

MACHINE LEARNING

- Machine Learning is the domain of Artificial Intelligence which is concerned with building adaptive computer systems that are able to improve their competence and/or efficiency through learning from input data or from their own problem-solving experience.
- Competence: A system is improving its competence if it learns to solve a broader class of problems, and to make fewer mistakes in problem solving.
- Efficiency: A system is improving its efficiency, if it learns to solve the problems from its area of competence faster or by using fewer resources.

CONT..





- A) human and machine thinking B) machines classify music based on genre

CONT.

- Feature Extractions

Male =  Hair length y =0

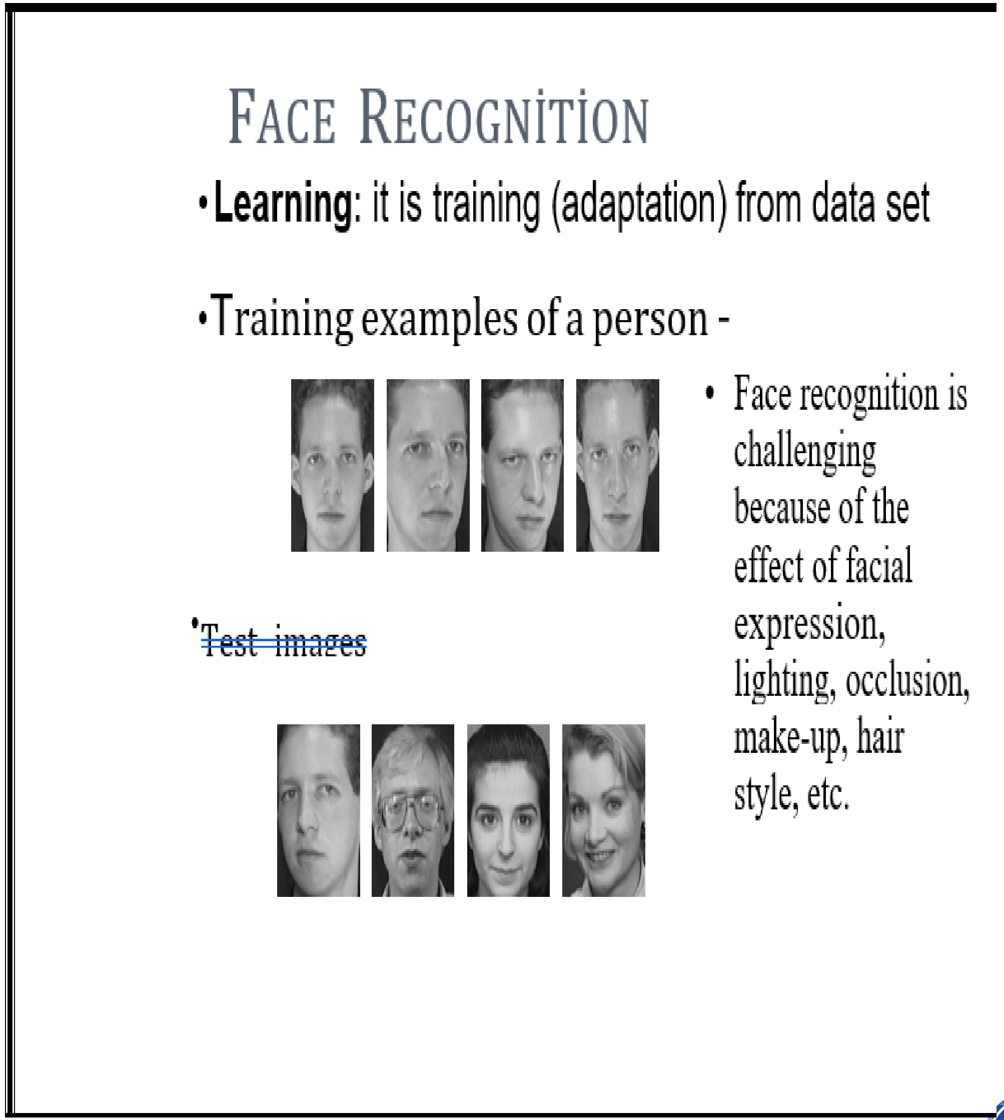
Female =  Hair length y = 30 cm

Task: to extract features which are good for classification.

Good features:

- Objects from the same class have similar feature values.
- Objects from different classes have different values.

“Good” features “Bad” features

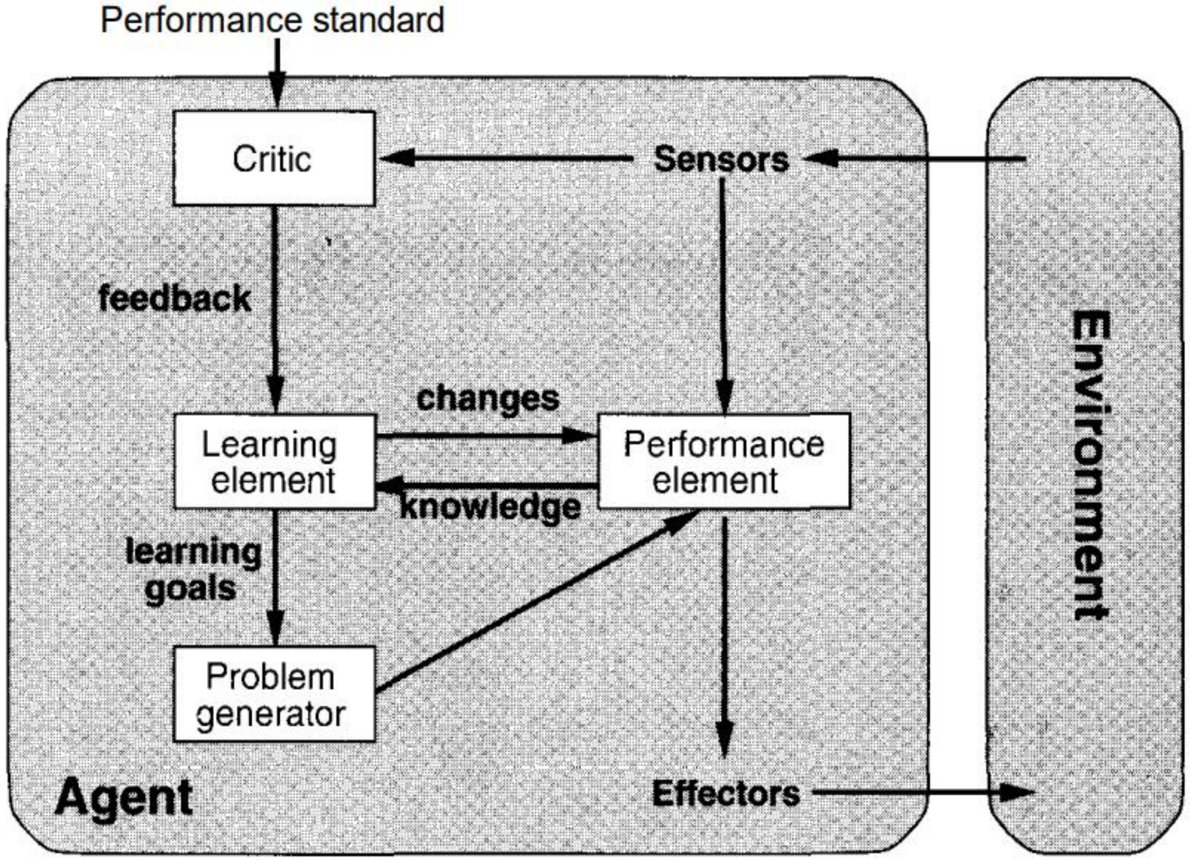


THE BASIC LEARNING MODEL

- A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P,
  - if its performance at tasks T, as measured by P, improves with experience E.
- Learning agents consist of four main components:
  - **Learning element** – the part of the agent responsible for improving its performance
  - **Performance element** – the part that chooses the actions to take
  - **Critic** – provides feedback for the learning element how the agent is doing with respect to a performance standard
  - **Problem generator** – suggests actions that could lead to new, informative experiences (suboptimal from the point of view of the performance element, but designed to improve that element)

CONT…

- The general model of learning agent



- The **critic** is designed to tell the learning element how well the agent is doing.

- **Problem generator** is responsible for suggesting actions that will lead to new and informative experiences

CONT...

- The design of the learning element is affected by four major issues:
  - Which **components** of the performance element are to be improved.
  - What **representation** is used for those components.
  - What **feedback** is available.
  - What **prior information** is available.

TYPES OF LEARNING

- **Supervised learning:** occurs where a set of input/output pairs are explicitly presented to the agent by a teacher
  - The teacher provides a category label for each pattern in a training set, then the learning algorithm finds a rule that does a good job of predicting the output associated with a new input.
- **Unsupervised learning:** Learning when there is no information about what the correct outputs are.
  - In unsupervised learning or clustering there is no explicit teacher, the system forms clusters or natural groupings of the input patterns.
- **Reinforcement learning:** an agent interacting with the world makes observations, takes actions, & is rewarded or punished;
  - it should learn to choose actions in order to obtain a lot of reward.
  - The agent is given an evaluation of its action, but not told the correct action. Reward strengthens likelihood of its action. Typically, the environment is assumed to be stochastic.

A, SUPERVISED LEARNING

- learning algorithms are the most commonly used for predictive analytics.
- requires human interaction to label data read for accurate supervised learning.
- In supervised learning, the model is taught by example using input and output data sets processed by human experts, usually data scientists.
- The model learns the relationships between input and output data and then uses that information to formulate predictions based on new datasets.
- Supervised machine learning methods commonly solve regression and classification problems:

❖ Classification problems: consist of a discrete unknown variable. Typically, the issue involves estimating which specific sample belongs to a set of pre-defined classes.

❖ Regression problems: involve estimating the mathematical relationship(s) between a continuous variable and one or more other variables.

B, UNSUPERVISED LEARNING

- algorithms do not require human experts but autonomously discover patterns in data.
- Unsupervised learning mainly deals with unlabeled data.
- The model must work on its own to find patterns and information.

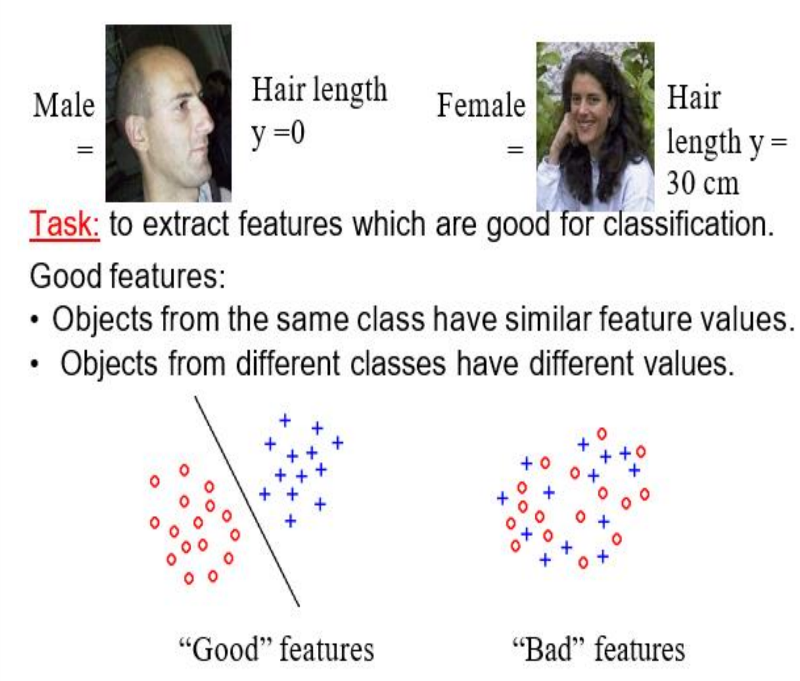
Examples of problems solved with unsupervised methods are clustering and association:

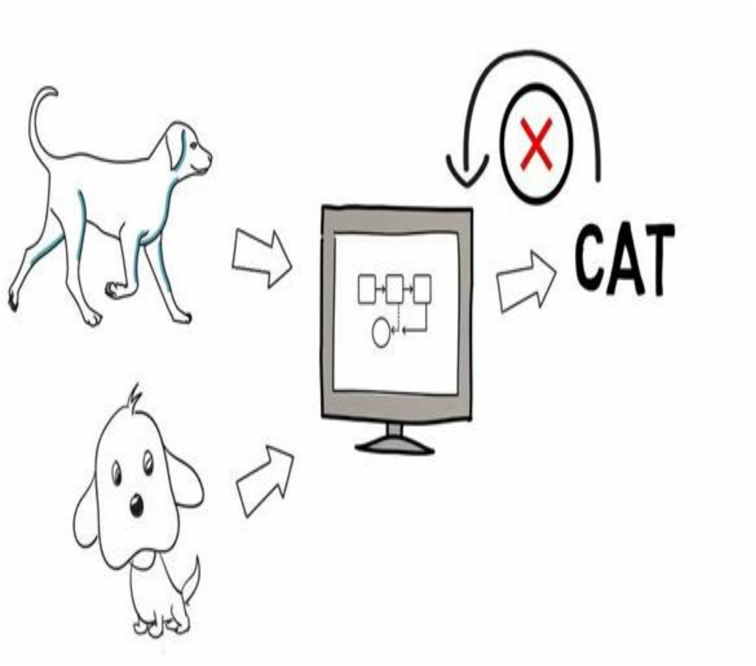
- **Clustering methods** - Clustering is the grouping of data that have similar characteristics.
- **Association methods** - Association consists of discovering groups of items frequently observed together.

C, REINFORCEMENT LEARNING

- **Reinforcement learning** teaches the machine through trial and error using feedback from its actions and experiences, also known as learning from mistakes.
- The system learns by trial and error to come up with solutions.
- It involves assigning positive values to desired outcomes and negative values to undesired effects.
- The result is optimal solutions; the system learns to avoid adverse outcomes and seek the positive.
- Practical applications of reinforcement learning include
  - building ratification intelligence for playing video games and
  - robotics and industrial automation

CONT.



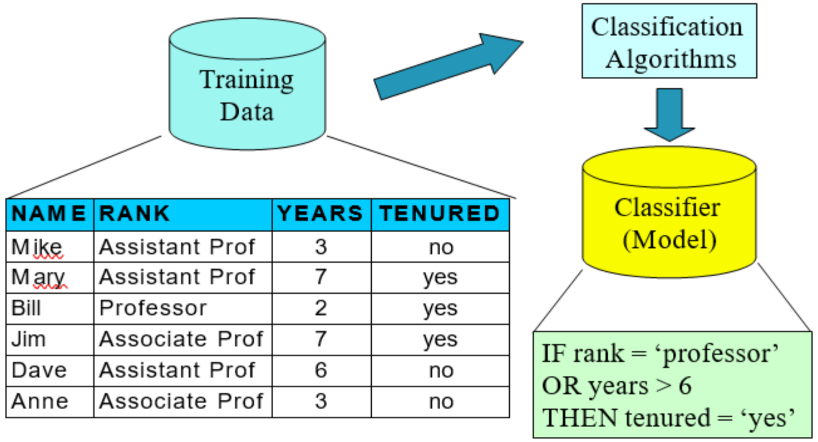


- A) Supervised learning B) Unsupervised learning C) Reinforcement learning

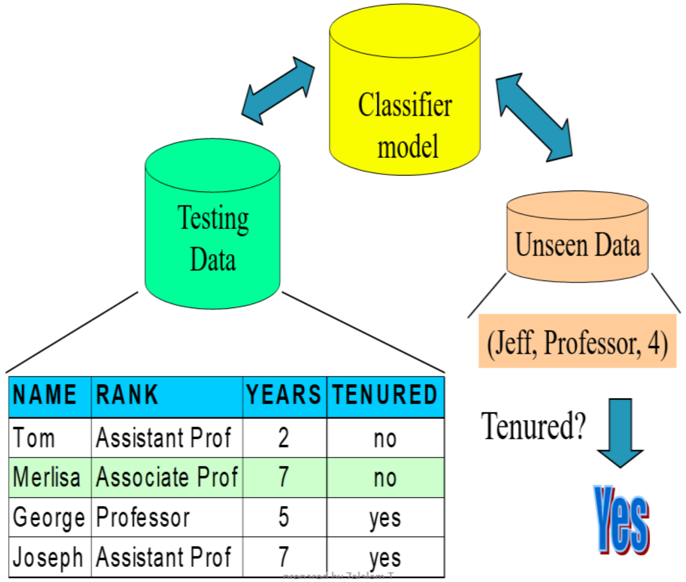
CONT

LEARNING—A TWO-STEP PROCESS

- **Model construction**:
  - A training set is used to create the model.
  - The model is represented as classification rules, decision trees, or mathematical formulae
- **Model usage**:
  - the test set is used to see how well it works for classifying future or unknown objects



NAME	RANK	YEARS	TENURED
Mike	Assistant Prof	3	no
Mary	Assistant Prof	7	yes
Bill	Professor	2	yes
Jim	Associate Prof	7	yes
Dave	Assistant Prof	6	no
Anne	Associate Prof	3	no



NAME	RANK	YEARS	TENURED
Tom	Assistant Prof	2	no
Merlisa	Associate Prof	7	no
George	Professor	5	yes
Joseph	Assistant Prof	7	yes

CONT…

Performance Evaluation metrics

	PREDICTED CLASS		
		Class=Yes	Class=No
ACTUAL CLASS	Class=Yes	a (TP)	b (TN)
	Class=No	c (FP)	d (FN)

- Most widely-used metric:
  - To measure the performance of the model in general:

$$\text{Accuracy} = \frac{a+d}{a+b+c+d} = \frac{TP}{TP+FP} * 100$$

  - To measure the performance of the model on each class: Recall, precision and F-Measure

LEARNING METHODS

- There are various learning methods. Popular learning techniques include the following.
  - **Decision tree** : divide decision space into piecewise constant regions.
  - **Neural networks**: partition by non-linear boundaries
  - **Bayesian network**: a probabilistic model
  - **Regression**: (linear or any other polynomial)
  - **Support vector machine**
  - **Expectation maximization algorithm**

LEARNING WITH DECISION TREES

- Decision tree induction is one of the simplest and yet most successful forms of learning algorithm.
- A decision tree takes input as an object or situation described by a set of properties, and outputs a **yes/no "decision"**
- Decision trees therefore represent Boolean functions.
- Functions with a larger range of outputs can also be represented, but for simplicity we will usually stick to the Boolean case

CONT...

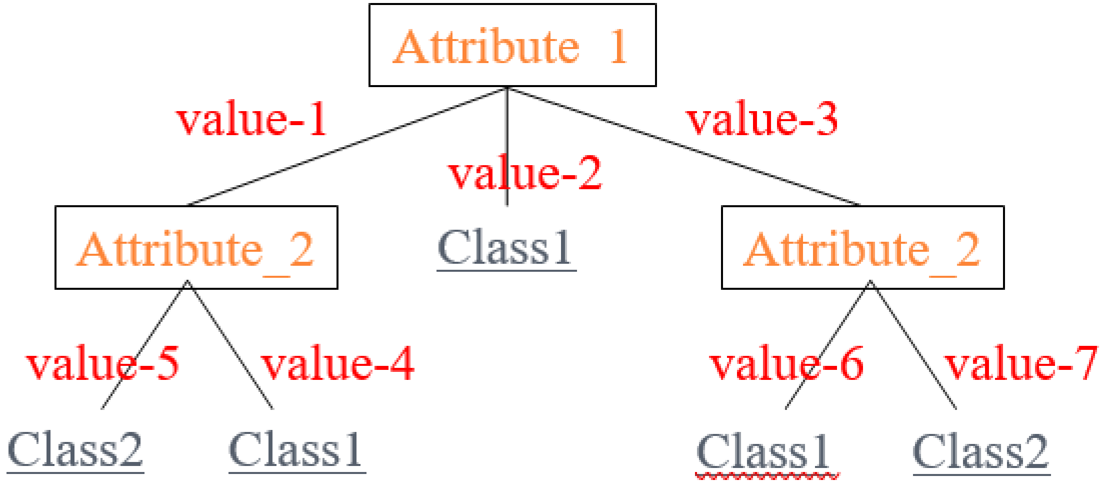
- Each internal node in the tree corresponds to a test of the value of one of the properties, and the branches from the node are labelled with the possible values of the test.
- Each leaf node in the tree specifies the Boolean value to be returned if that leaf is reached.
- Decision trees are implicitly limited to talking about a single object.
- We cannot use decision trees to represent tests that refer to two or more different objects

- Widely used learning method. It has been applied to:
  - classify medical patients based on the disease,
  - equipment malfunction by cause,
  - loan applicant by likelihood of payment.

- Easy to interpret: can be re-represented as if-then-else rules

CONT…

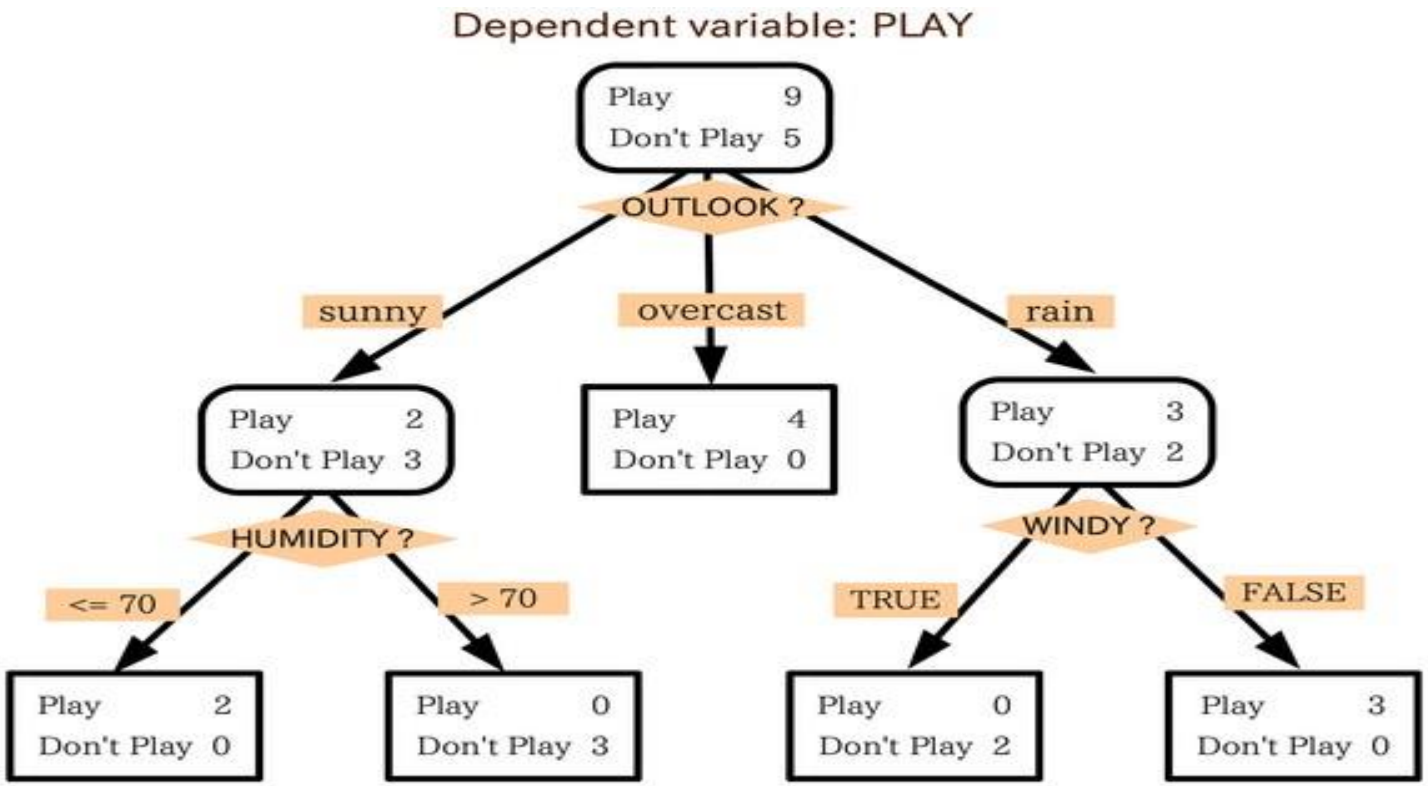
- Tree where internal nodes are simple decision rules on one or more attributes and leaf nodes are predicted class labels; i.e. a Boolean classifier for the input instance.
  - Given an instance of an object or situation, which is specified by a set of properties, the tree returns a "yes" or "no" decision about that instance.



CONT...

- Given an instance of an object or situation, which is specified by a set of properties, the tree returns a "yes" or "no" decision about that instance.

A decision tree for deciding whether to wait for a table features



CONT…

SO WHAT IS THE BEST DECISION TREE?

- Up to now, we used our common sense to select what attributes were most likely to be relevant to classify whether people will get play tennis or not.
- It is better to construct an optimal decision tree by selecting best attributes using Information Gain
  - **Information gain**: Select the attribute with the highest gain, which means select attributes that creates small average disorder.
    - First, **compute the disorder using Entropy**; the expected information needed to classify objects into classes.
    - Second, **measure the Information Gain**; to calculate by how much the disorder of a set would reduce by knowing the value of a particular attribute.

CONT..

ENTROPY: The **Entropy** measures the disorder of a set S containing a total of *n* examples of which *n+* are positive and *n-* are negative and it is given by:

$$D(n+ \,, n- \,) = -\frac{n+}{n}\log 2\,\frac{n+}{n} - \frac{n-}{n}\log 2\,\frac{n-}{n}$$

Some useful properties of the Entropy:

- D(n,m) = D(m,n)
- D(0,m) = 0
  - *D(S)=0* means that all the examples in *S* have the same class
- D(m,m) = 1

- *D(S)=1* means that half the examples in *S* are of one class and half are the opposite class

CONT..

Information Gain

- The Information Gain measures the expected reduction in entropy due to splitting on an attribute A

$$GAIN_{split} = Entropy(S) - \left(\sum_{i=1}^{k} \frac{n_i}{n} Entropy(i) \right)$$

- Information Gain: Measures Reduction in Entropy achieved because of the split. Choose the split that achieves most reduction (maximizes GAIN).
- Let us construct an optimal decision tree with the help of Entropy and Information Gain.
  - Entropy and information gain helps us to select the best attributes in the process of decision tree construction.

CONT..

- **EXAMPLE 1**

  you want to predict whether tennis is played or not . this depend on the condition how we know? data collected: predict based on the observed properties of the people?

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

CONT…

Day	Outlook	Temp	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Attribute: Outlook

$Values\ (Outlook) = Sunny, Overcast, Rain$

$S = [9+, 5-]$ $\quad Entropy(S) = -\frac{9}{14}\log_2\frac{9}{14} - \frac{5}{14}\log_2\frac{5}{14} = 0.94$

$S_{Sunny} \leftarrow [2+, 3-]$ $\quad Entropy(S_{Sunny}) = -\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5} = 0.971$

$S_{Overcast} \leftarrow [4+, 0-]$ $\quad Entropy(S_{Overcast}) = -\frac{4}{4}\log_2\frac{4}{4} - \frac{0}{4}\log_2\frac{0}{4} = 0$

$S_{Rain} \leftarrow [3+, 2-]$ $\quad Entropy(S_{Rain}) = -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5} = 0.971$

$$Gain\ (S, Outlook) = Entropy(S) - \sum_{v \in \{Sunny, Overcast, Rain\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S, Outlook) = Entropy(S) - \frac{5}{14} Entropy(S_{Sunny}) - \frac{4}{14} Entropy(S_{Overcast}) - \frac{5}{14} Entropy(S_{Rain})$$

$$Gain(S, Outlook) = 0.94 - \frac{5}{14}0.971 - \frac{4}{14}0 - \frac{5}{14}0.971 = 0.2464$$

CONT…

Day	Outlook	Temp	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Attribute: Temp

$Values\ (Temp) = Hot, Mild, Cool$

$S = [9+, 5-]$ $\qquad Entropy(S) = -\frac{9}{14}log_2\frac{9}{14} - \frac{5}{14}log_2\frac{5}{14} = 0.94$

$S_{Hot} \leftarrow [2+, 2-]$ $\qquad Entropy(S_{Hot}) = -\frac{2}{4}log_2\frac{2}{4} - \frac{2}{4}log_2\frac{2}{4} = 1.0$

$S_{Mild} \leftarrow [4+, 2-]$ $\qquad Entropy(S_{Mild}) = -\frac{4}{6}log_2\frac{4}{6} - \frac{2}{6}log_2\frac{2}{6} = 0.9183$

$S_{Cool} \leftarrow [3+, 1-]$ $\qquad Entropy(S_{Cool}) = -\frac{3}{4}log_2\frac{3}{4} - \frac{1}{4}log_2\frac{1}{4} = 0.8113$

$$Gain\ (S, Temp) = Entropy(S) - \sum_{v \in \{Hot, Mild, Cool\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S, Temp)$$

$$= Entropy(S) - \frac{4}{14}Entropy(S_{Hot}) - \frac{6}{14}Entropy(S_{Mild})$$

$$-\frac{4}{14}Entropy(S_{Cool})$$

$$\text{Gain(S, Temp)} = 0.94 - \frac{4}{14} * 1.0 - 6/14*0.9183 - 4/14*0.8113 = 0.0289$$

CONT...

Day	Outlook	Temp	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Attribute: Humidity

$Values\ (Humidity) = High, Normal$

$S = [9+, 5-]$ $\quad Entropy(S) = -\frac{9}{14}\log_2\frac{9}{14} - \frac{5}{14}\log_2\frac{5}{14} = 0.94$

$S_{High} \leftarrow [3+, 4-]$ $\quad Entropy(S_{High}) = -\frac{3}{7}\log_2\frac{3}{7} - \frac{4}{7}\log_2\frac{4}{7} = 0.9852$

$S_{Normal} \leftarrow [6+, 1-]$ $\quad Entropy(S_{Normal}) = -\frac{6}{7}\log_2\frac{6}{7} - \frac{1}{7}\log_2\frac{1}{7} = 0.5916$

$$Gain\ (S, Humidity) = Entropy(S) - \sum_{v \in \{High, Normal\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S, Humidity)$$

$$= Entropy(S) - \frac{7}{14} Entropy(S_{High}) - \frac{7}{14} Entropy(S_{Normal})$$

$$Gain(S, Humidity) = 0.94 - \frac{7}{14} 0.9852 - \frac{7}{14} 0.5916 = 0.1516$$

CONT…

Day	Outlook	Temp	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Attribute: Wind

$Values\ (Wind) = Strong, Weak$

$S = [9+, 5-]$ $\quad Entropy(S) = -\frac{9}{14}\log_2\frac{9}{14} - \frac{5}{14}\log_2\frac{5}{14} = 0.94$

$S_{Strong} \leftarrow [3+, 3-]$ $\quad Entropy(S_{Strong}) = 1.0$

$S_{Weak} \leftarrow [6+, 2-]$ $\quad Entropy(S_{Weak}) = -\frac{6}{8}\log_2\frac{6}{8} - \frac{2}{8}\log_2\frac{2}{8} = 0.8113$

$$Gain\ (S, Wind) = Entropy(S) - \sum_{v \in \{Strong, Weak\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S, Wind) = Entropy(S) - \frac{6}{14} Entropy(S_{Strong}) - \frac{8}{14} Entropy(S_{Weak})$$

$$Gain(S, Wind) = 0.94 - \frac{6}{14}\,1.0 - \frac{8}{14}\,0.8113 = 0.0478$$

CONT…

Day	Outlook	Temp	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

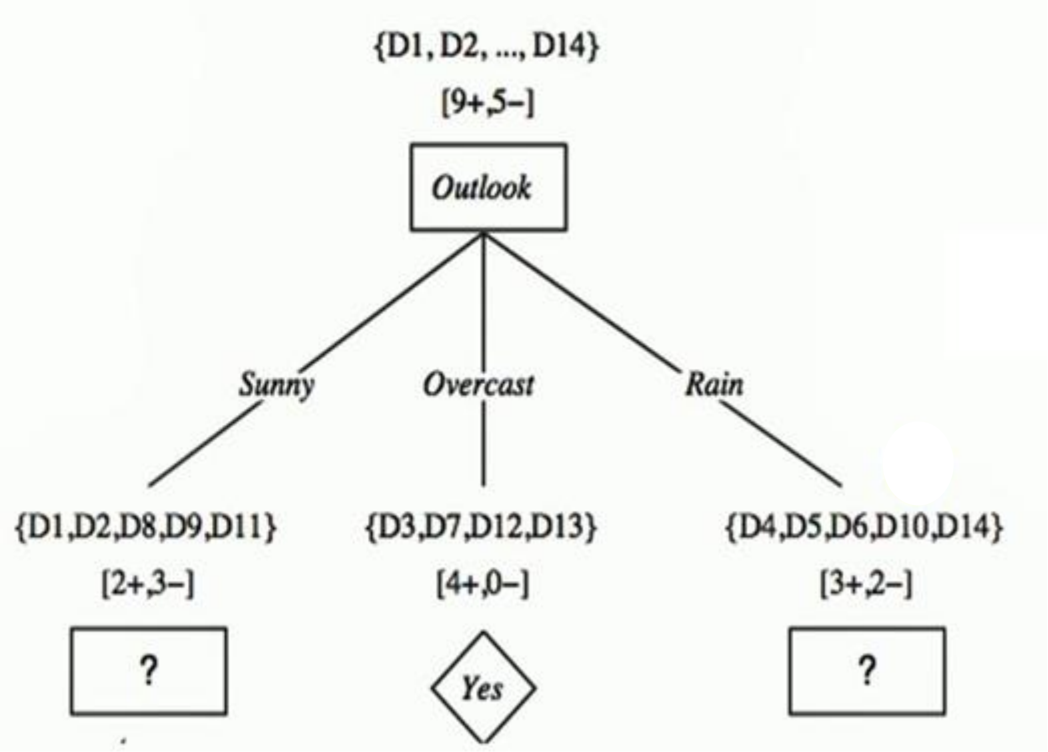
$$Gain(S, Outlook) = 0.2464$$

$$Gain(S, Temp) = 0.0289$$

$$Gain(S, Humidity) = 0.1516$$

$$Gain(S, Wind) = 0.0478$$

CONT..



Day	Temp	Humidity	Wind	Play Tennis
D1	Hot	High	Weak	No
D2	Hot	High	Strong	No
D8	Mild	High	Weak	No
D9	Cool	Normal	Weak	Yes
D11	Mild	Normal	Strong	Yes

Attribute: Temp

Values (Temp) = Hot, Mild, Cool

$S_{Sunny} = [2+, 3-]$ $\quad Entropy(S_{Sunny}) = -\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5} = 0.97$

$S_{Hot} \leftarrow [0+, 2-]$ $\quad Entropy(S_{Hot}) = 0.0$

$S_{Mild} \leftarrow [1+, 1-]$ $\quad Entropy(S_{Mild}) = 1.0$

$S_{Cool} \leftarrow [1+, 0-]$ $\quad Entropy(S_{Cool}) = 0.0$

$$Gain(S_{Sunny}, Temp) = Entropy(S) - \sum_{v \in \{Hot, Mild, Cool\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S_{Sunny}, Temp)$$

$$= Entropy(S) - \frac{2}{5}Entropy(S_{Hot}) - \frac{2}{5}Entropy(S_{Mild})$$

$$-\frac{1}{5}Entropy(S_{Cool})$$

$$Gain(S_{sunny}, Temp) = 0.97 - \frac{2}{5}0.0 - \frac{2}{5}1 - \frac{1}{5}0.0 = 0.570$$

CONT...

Day	Temp	Humidity	Wind	Play Tennis
D1	Hot	High	Weak	No
D2	Hot	High	Strong	No
D8	Mild	High	Weak	No
D9	Cool	Normal	Weak	Yes
D11	Mild	Normal	Strong	Yes

Attribute: Humidity

$Values(Humidity) = High, Normal$

$S_{Sunny} = [2+, 3-]$ $\quad Entropy(S) = -\frac{2}{5}log_2\frac{2}{5} - \frac{3}{5}log_2\frac{3}{5} = 0.97$

$S_{high} \leftarrow [0+, 3-]$ $\quad Entropy(S_{High}) = 0.0$

$S_{Normal} \leftarrow [2+,$ 0] entropy(Snormal)=0.0

$$Gain(S_{Sunny}, Humidity) = Entropy(S) - \sum_{v \in \{High, Normal\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S_{Sunny}, Humidity) = Entropy(S) - \frac{3}{5}Entropy(S_{High}) - \frac{2}{5}Entropy(S_{Normal})$$

$$Gain(S_{sunny}, Humidity) = 0.97 - \frac{3}{5}0.0 - \frac{2}{5}0.0 = 0.97$$

Day	Temp	Humidity	Wind	Play Tennis
D1	Hot	High	Weak	No
D2	Hot	High	Strong	No
D8	Mild	High	Weak	No
D9	Cool	Normal	Weak	Yes
D11	Mild	Normal	Strong	Yes

Attribute: Wind

$Values(Wind) = Strong, Weak$

$S_{Sunny} = [2+, 3-]$ $\quad Entropy(S) = -\frac{2}{5}log_2\frac{2}{5} - \frac{3}{5}log_2\frac{3}{5} = 0.97$

S. ← [1+ 1-] $\quad Entropy(S_{Strong}) = 1.0$

$Entropy(S_{Weak}) = -\frac{1}{3}log_2\frac{1}{3} - \frac{2}{3}log_2\frac{2}{3} = 0.9183$

$$Gain(S_{Sunny}, Wind) = Entropy(S) - \sum_{v \in \{Strong, Weak\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

Gain(Ssunny,wind)= Entropy(S)-2/5*Entropy(S strong)-3/5*Entropy(S weak)

0.97-(2/5)*1-(3/5)*0.9183=0.0192

CONT…

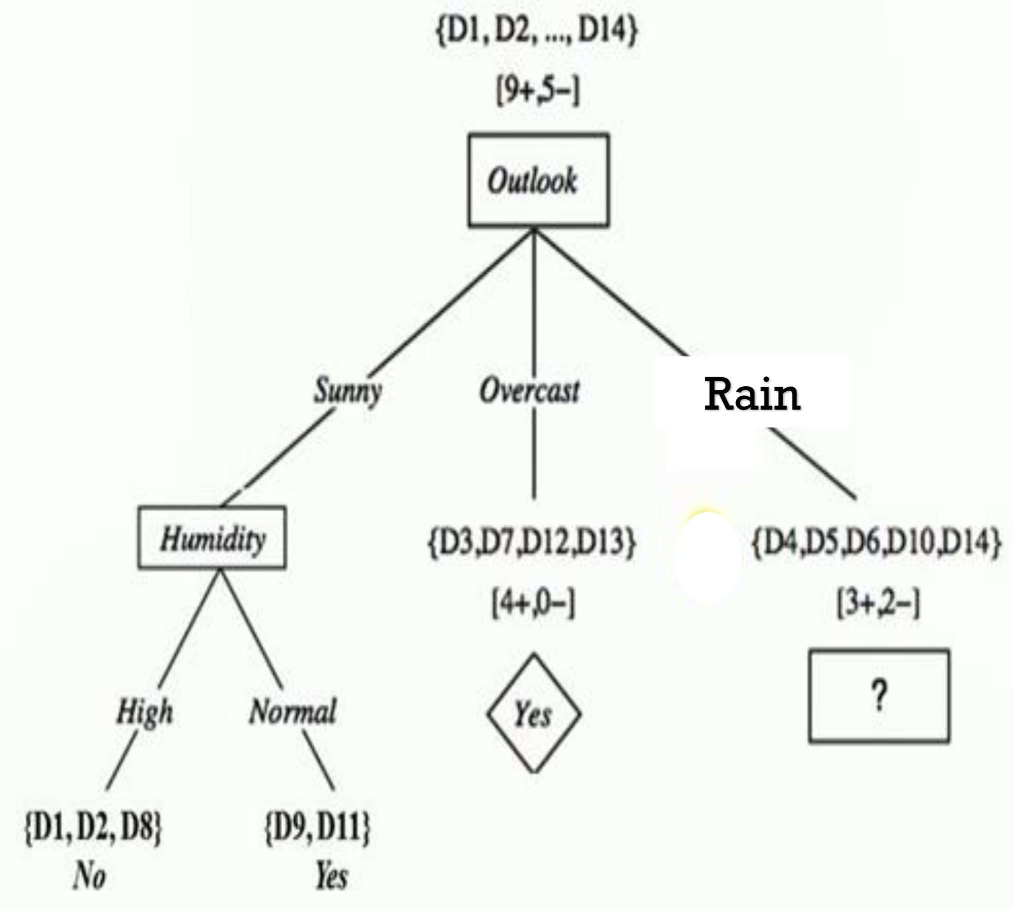
Day	Temp	Humidity	Wind	Play Tennis
D1	Hot	High	Weak	No
D2	Hot	High	Strong	No
D8	Mild	High	Weak	No
D9	Cool	Normal	Weak	Yes
D11	Mild	Normal	Strong	Yes

Gain(Ssunny, Temp) =0.570

$$Gain(S_{sunny}, Humidity) = 0.97$$

$$Gain(S_{sunny}, Wind) = 0.0192$$

CONT…



Day	Temp	Humidity	Wind	Play Tennis
D4	Mild	High	Weak	Yes
D5	Cool	Normal	Weak	Yes
D6	Cool	Normal	Strong	No
D10	Mild	Normal	Weak	Yes
D14	Mild	High	Strong	No

Attribute: Temp

$Values(Temp) = Hot, Mild, Cool$

$S_{Rain} = [3+, 2-]$ Entropy(S_{rain}) $= -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5} = 0.97$

$S_{Hot} \leftarrow [0+, 0-]$ $Entropy(S_{Hot}) = 0.0$

$S_{Mild} \leftarrow [2+, 1-]$ $Entropy(S_{Mild}) = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} = 0.9183$

$S_{Cool} \leftarrow [1+, 1-]$ $Entropy(S_{Cool}) = 1.0$

$$Gain(S_{Rain}, Temp) = Entropy(S) - \sum_{v \in \{Hot, Mild, Cool\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S_{Rain}, Temp)$$

$$= Entropy(S) - \frac{0}{5}Entropy(S_{Hot}) - \frac{3}{5}Entropy(S_{Mild})$$

$$-\frac{2}{5}Entropy(S_{Cool})$$

$$Gain(S_{Rain}, Temp) = 0.97 - \frac{0}{5}0.0 - \frac{3}{5}0.918 - \frac{2}{5}1.0 = 0.0192$$

CONT...

Day	Temp	Humidity	Wind	Play Tennis
D4	Mild	High	Weak	Yes
D5	Cool	Normal	Weak	Yes
D6	Cool	Normal	Strong	No
D10	Mild	Normal	Weak	Yes
D14	Mild	High	Strong	No

Attribute: Humidity

$Values(Humidity) = High, Normal$

$S_{Rain} = [3+, 2-]$ $\quad Entropy(S_{Sunny}) = -\frac{3}{5}log_2\frac{3}{5} - \frac{2}{5}log_2\frac{2}{5} = 0.97$

$S_{High} \leftarrow [1+, 1-]$ $\quad Entropy(S_{High}) = 1.0$

$S_{Normal} \leftarrow [2+, 1$ $\quad Entropy(S_{Normal}) = -\frac{2}{3}log_2\frac{2}{3} - \frac{1}{3}log_2\frac{1}{3} = 0.9183$

$$Gain(S_{Rain}, Humidity) = Entropy(S) - \sum_{v \in \{High, Normal\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S_{Rain}, Humidity) = Entropy(S) - \frac{2}{5}Entropy(S_{High}) - \frac{3}{5}Entropy(S_{Normal})$$

$$Gain(S_{Rain}, Humidity) = 0.97 - \frac{2}{5}1.0 - \frac{3}{5}0.918 = 0.0192$$

Day	Temp	Humidity	Wind	Play Tennis
D4	Mild	High	Weak	Yes
D5	Cool	Normal	Weak	Yes
D6	Cool	Normal	Strong	No
D10	Mild	Normal	Weak	Yes
D14	Mild	High	Strong	No

Attribute: Wind

$Values(wind) = Strong, Weak$

$S_{Rain} = [3+, 2-]$ $\quad Entropy(S_{Sunny}) = -\frac{3}{5}log_2\frac{3}{5} - \frac{2}{5}log_2\frac{2}{5} = 0.97$

$S_{Strong} \leftarrow [0+, 2-]$ $\quad Entropy(S_{Strong}) = 0.0$

$S_{Weak} \leftarrow [3+, 0-]$ $\quad Entropy(S_{Weak}) = 0.0$

$$Gain(S_{Rain}, Wind) = Entropy(S) - \sum_{v \in \{Strong, Weak\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S_{Rain}, Wind) = Entropy(S) - \frac{2}{5}Entropy(S_{Strong}) - \frac{3}{5}Entropy(S_{Weak})$$

$$Gain(S_{Rain}, Wind) = 0.97 - \frac{2}{5}0.0 - \frac{3}{5}0.0 = 0.97$$

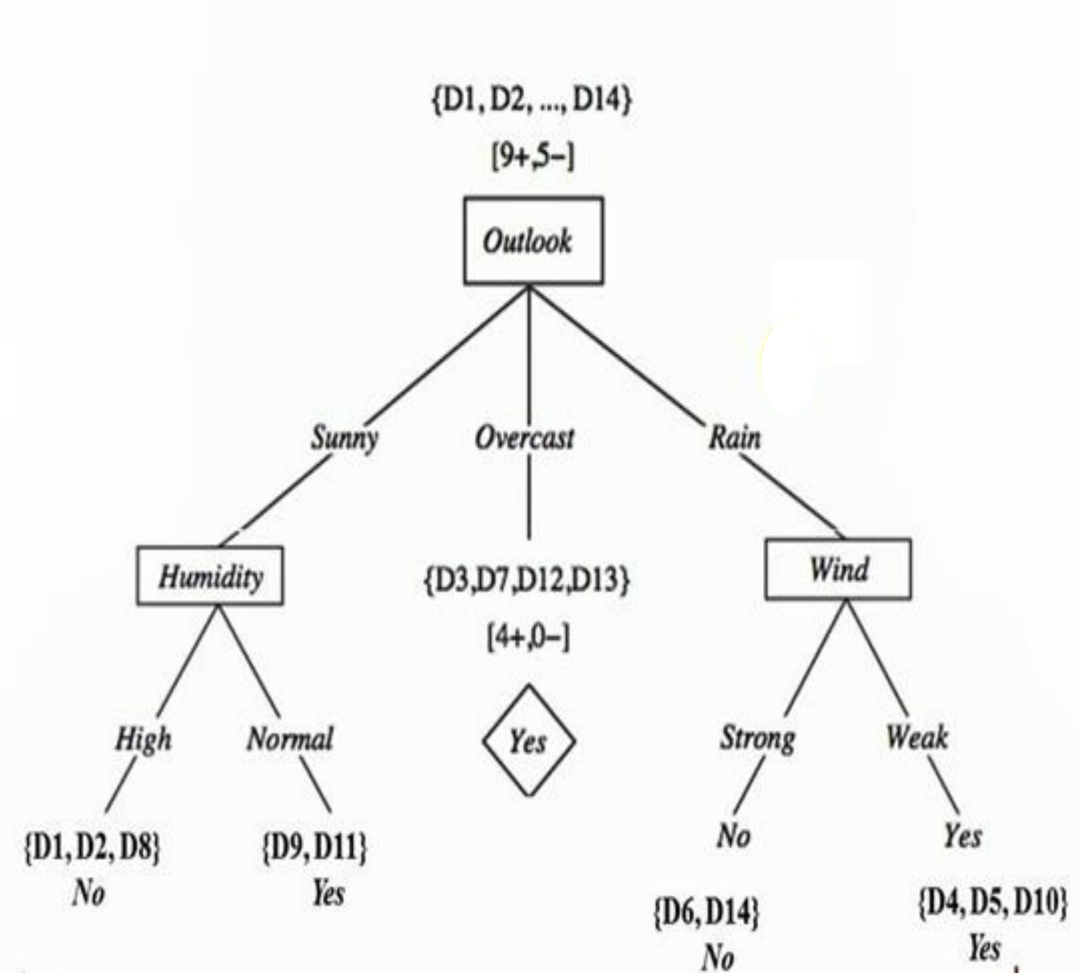
CONT...

Day	Temp	Humidity	Wind	Play Tennis
D4	Mild	High	Weak	Yes
D5	Cool	Normal	Weak	Yes
D6	Cool	Normal	Strong	No
DI0	Mild	Normal	Weak	Yes
DI4	Mild	High	Strong	No

$$Gain(S_{Rain}, Temp) = 0.0192$$

$$Gain(S_{Rain}, Humidit\ )=0.0192$$

$$Gain(S_{Rain}, Wind) = 0.97$$



PROS AND CONS OF DECISION TREES

- **Pros**

+ Reasonable training time
+ Fast application
+ Easy to interpret
+ Easy to implement
+ Can handle large

- **Cons**

- Cannot handle complicated relationship between features
- simple decision boundaries
- problems with lots of missing

- Why decision tree induction in data mining?
  - relatively faster learning speed (than other classification methods)
  - convertible to simple and easy to understand classification rules
- comparable classification accuracy with other methods

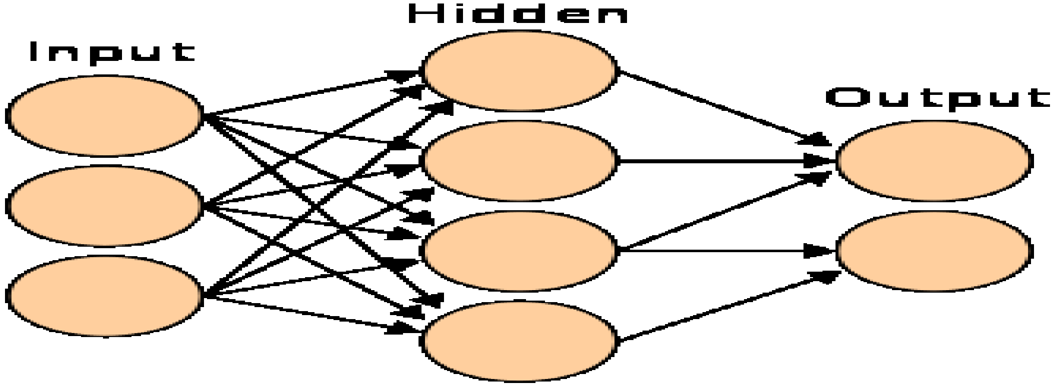
NEURAL NETWORK

NEURAL NETWORK

- Complex networks of simple computing elements
- Capable of learning from examples
  - with appropriate learning methods
- Collection of simple elements performs high-level operations
  - Thought
  - Reasoning
  - consciousness

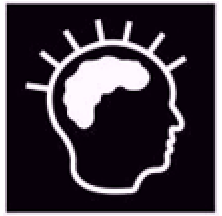
NEURAL NETWORK

- It is represented as a layered set of interconnected processors.
- These processor nodes has a relationship with the neurons of the brain.
- Each node has a weighted connection to several other nodes in adjacent layers.
- Individual nodes take the input received from connected nodes and use the weights together to compute output values.
- The inputs are fed simultaneously into the input layer.
- The weighted outputs of these units are fed into hidden layer.
- the weighted outputs of the last hidden layer are inputs to units making up the output layer.



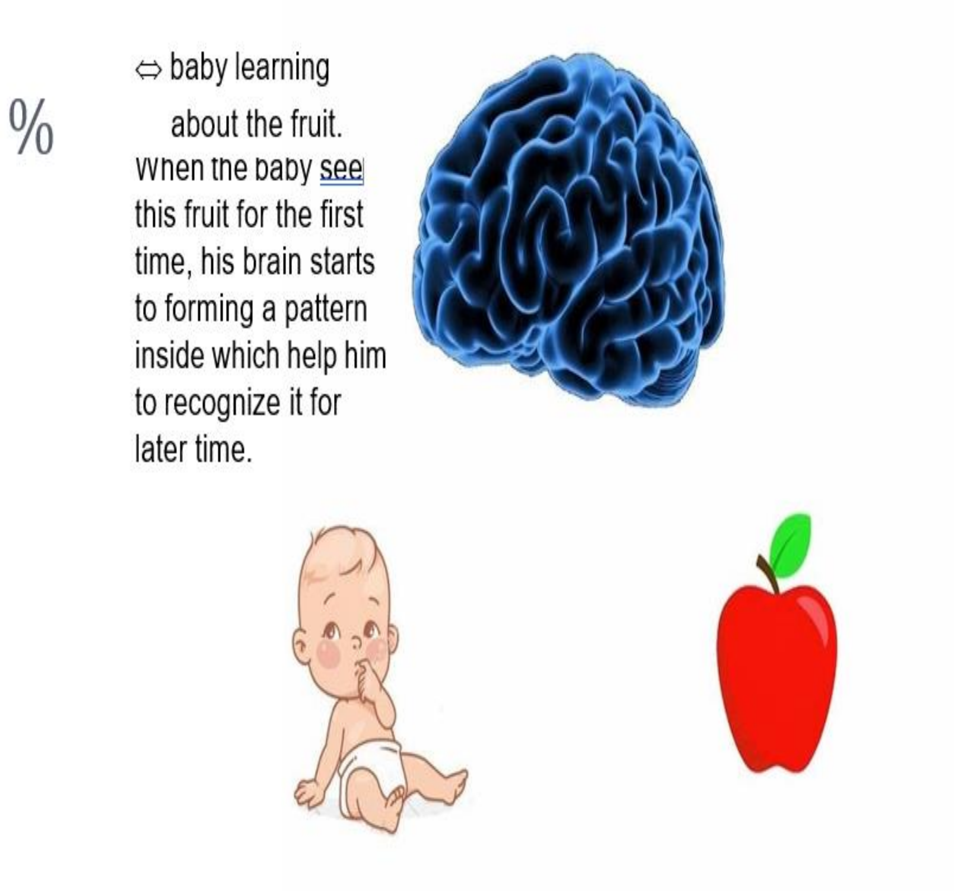
CONT…

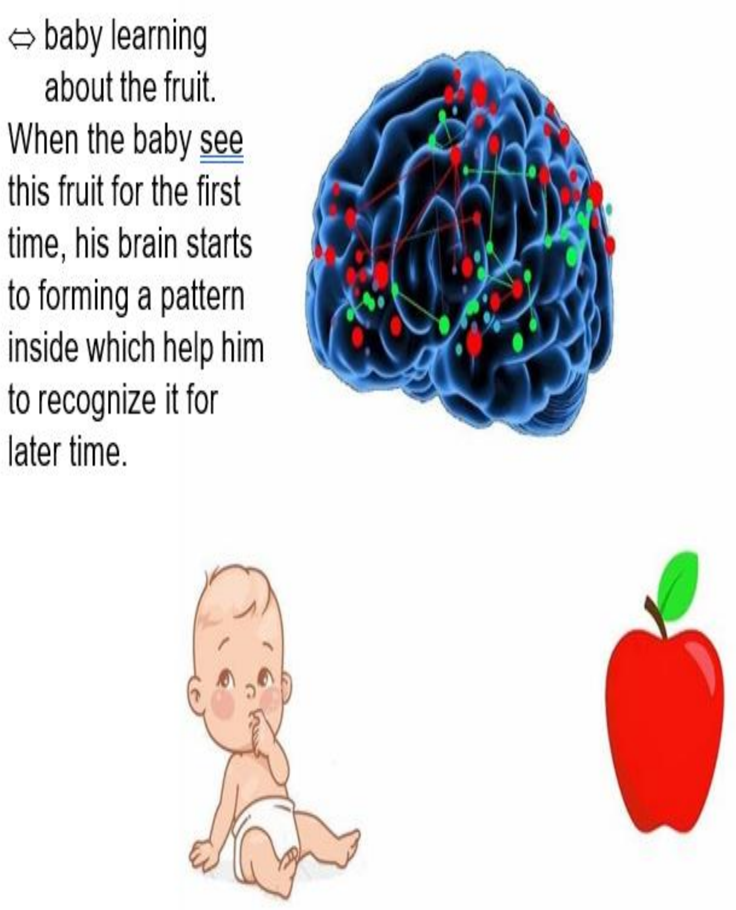
Features of the Brain



- There are around ten billion (1010) neurons in our brain
- Neuron switching time >10-3secs.
- Face Recognition ~0.1secs.
- On average, each neuron has several thousand connections.
- Hundreds of operations per second.
- High degree of parallel computation.
- Distributed representations.
- Die off frequently (never replaced).
- Compensated for problems by massive parallelism.

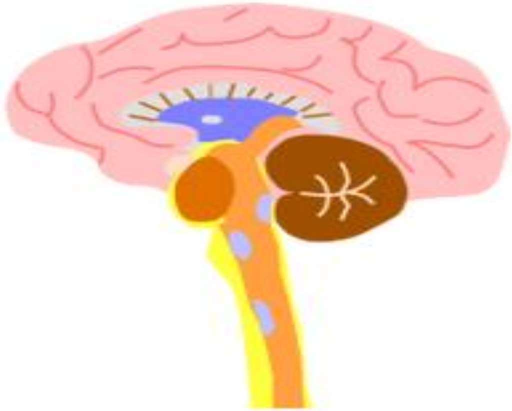
CONT...


%
⇔ baby learning
about the fruit.
When the baby see
this fruit for the first
time, his brain starts
to forming a pattern
inside which help him
to recognize it for
later time.


⇔ baby learning
about the fruit.
When the baby see
this fruit for the first
time, his brain starts
to forming a pattern
inside which help him
to recognize it for
later time.

MACHIN VS HUMAN

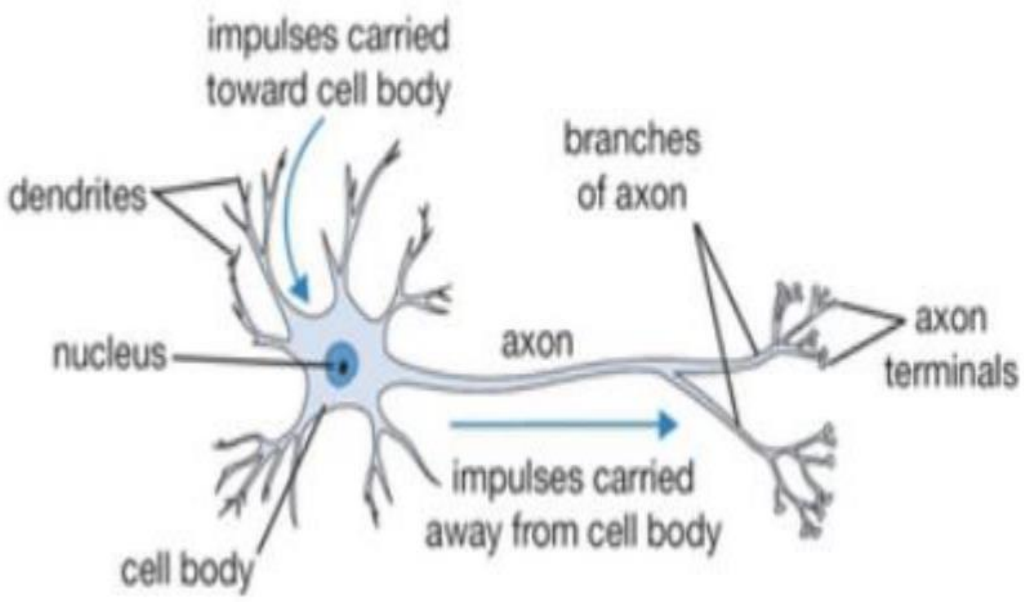
- The Brain
  - Pattern Recognition
  - Association
  - Complexity
  - Noise Tolerance

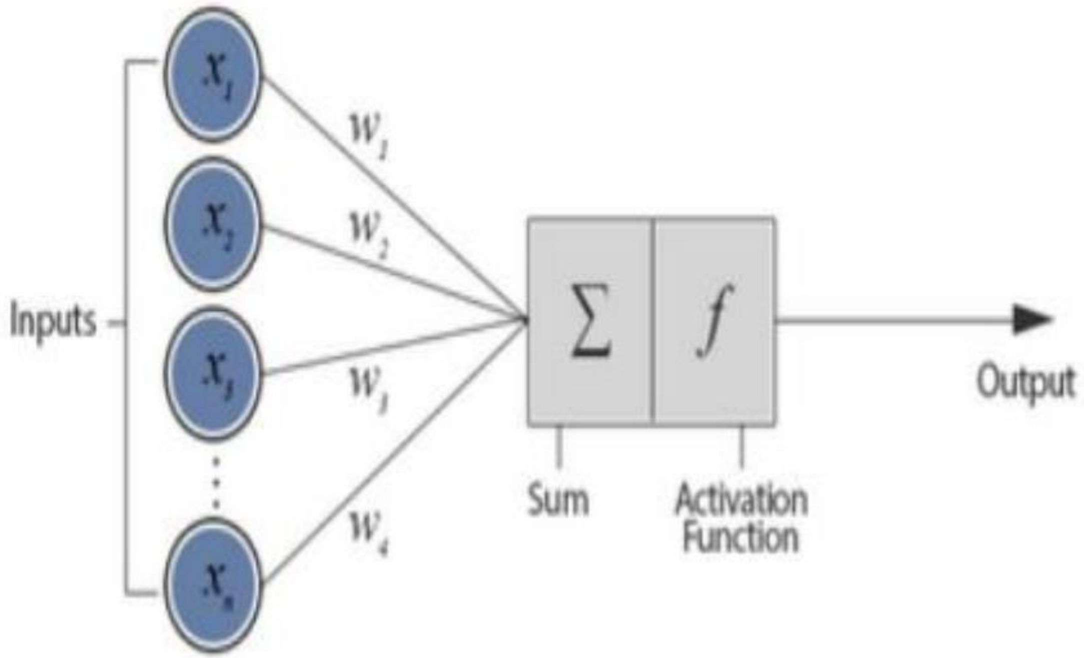


- The Machine
  - Calculation
  - Precision
  - Logic

CONT…

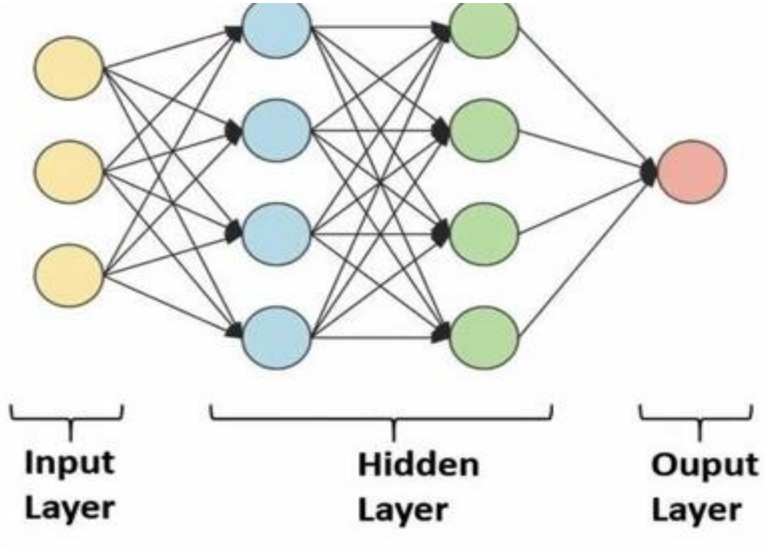
Biological neurons vs Artificial neural network

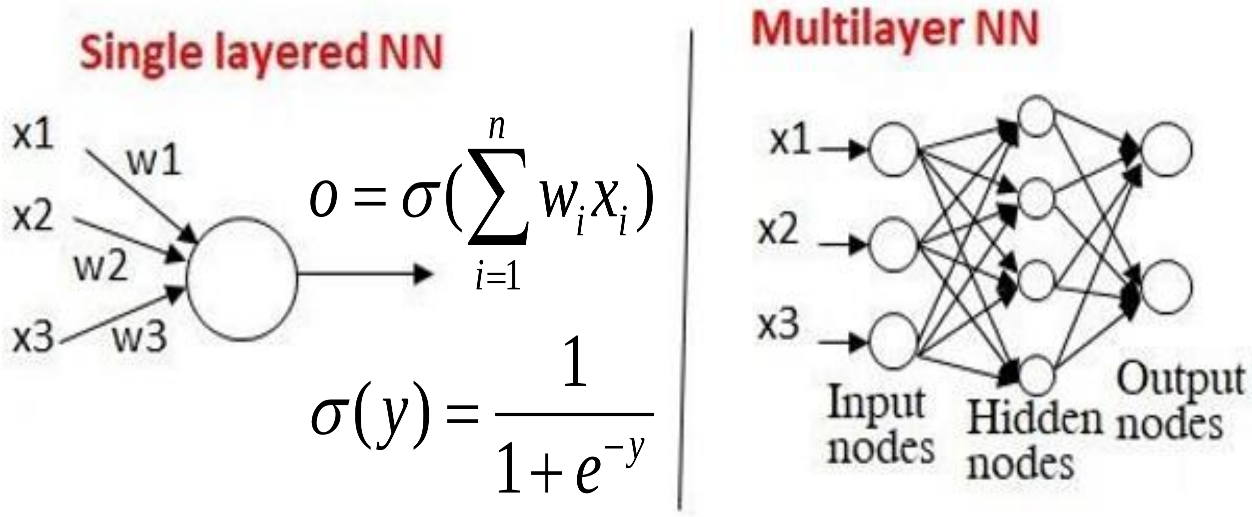




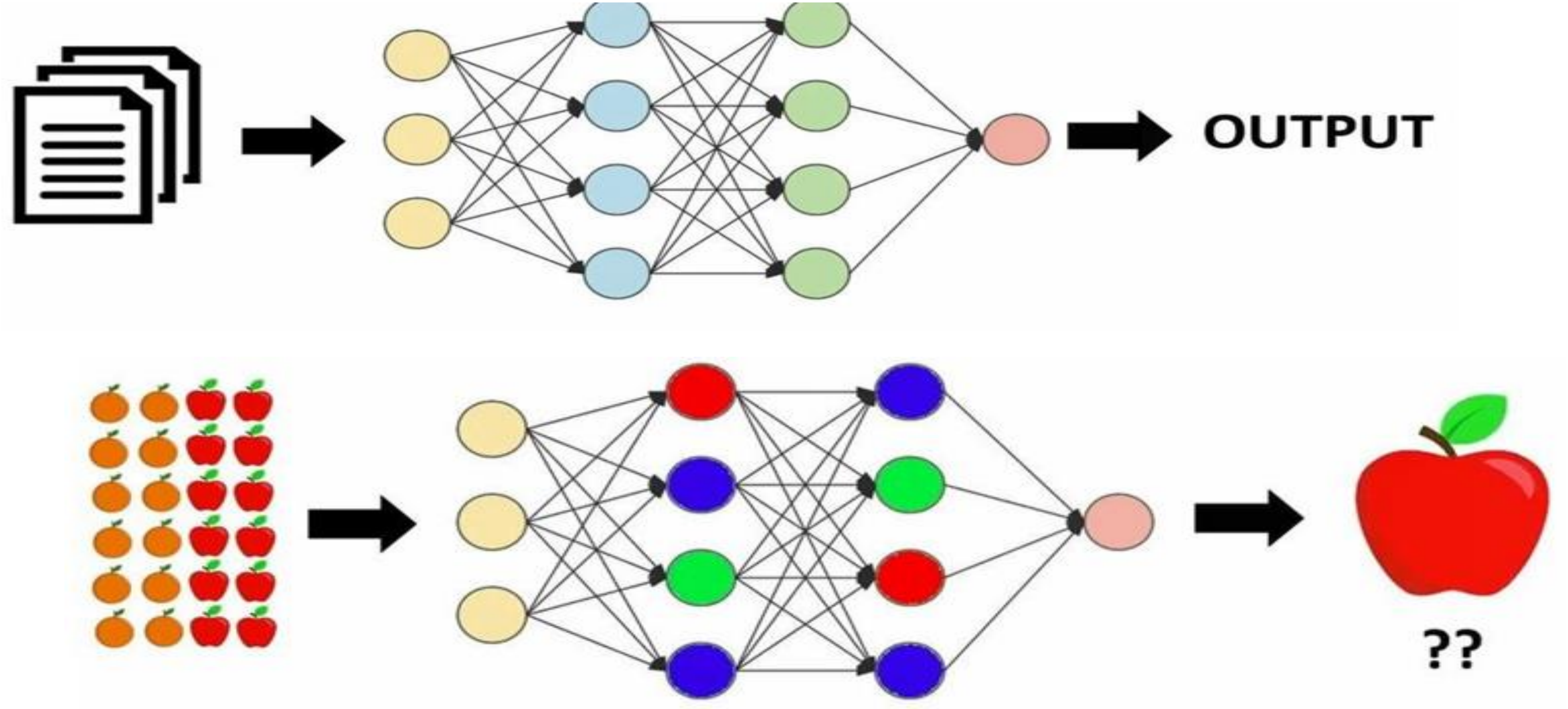
ARCHITECTURE OF NEURAL NETWORK

- **Neural networks** are used **to look for patterns in data**, **learn these patterns,** and then **classify new patterns** & **make forecasts.**
- A network with **the input** and **output layer** only is called single-layered neural network. Whereas, a multilayer neural network is a generalized one with one or more hidden layer.
- A network containing two hidden layers is called a three-layer neural network, and so on.

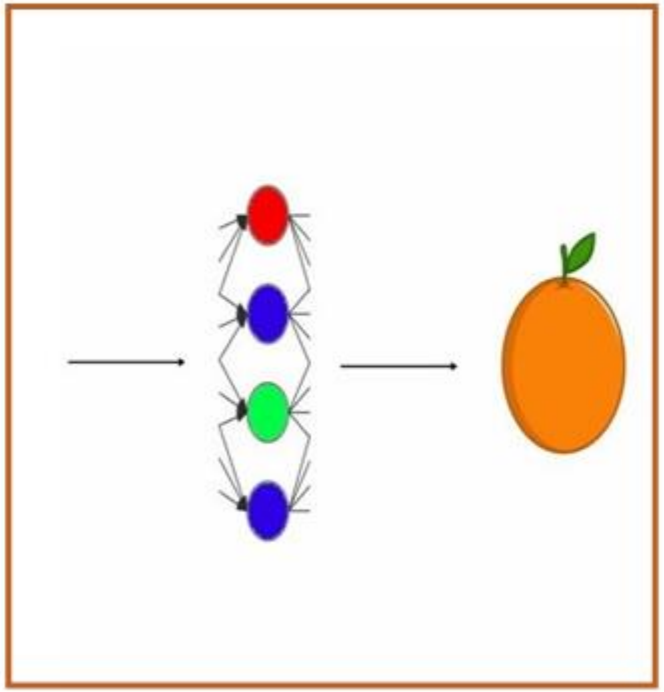


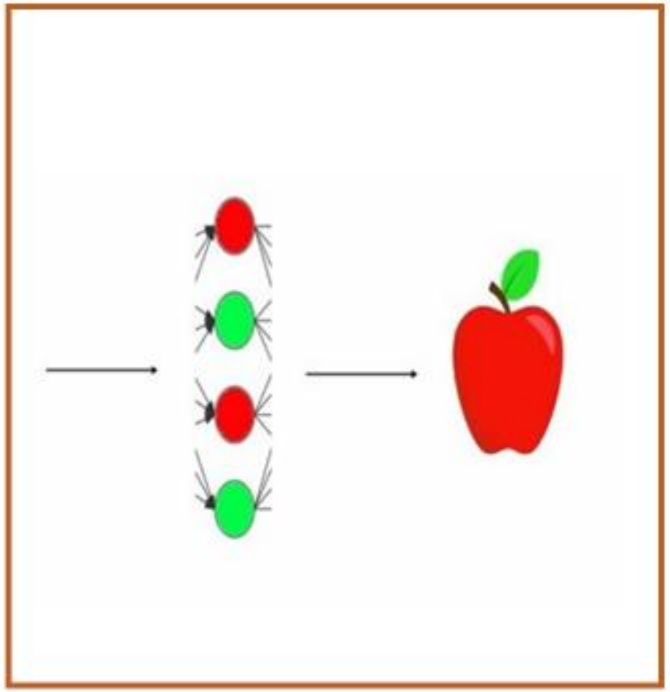


CONT…



CONT…



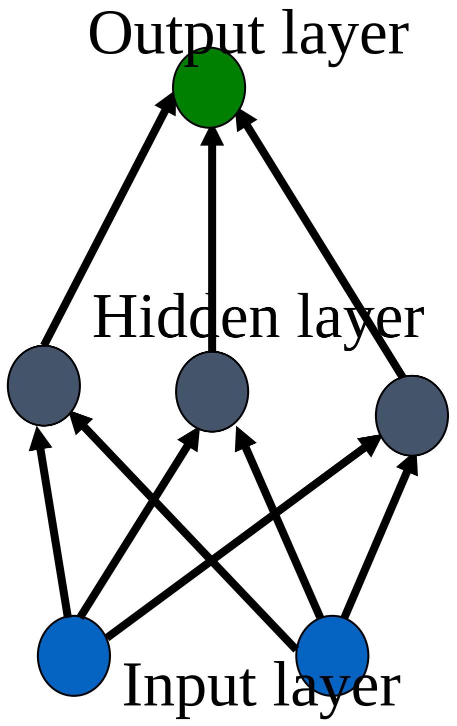


CONT...

- INPUT: records with class attribute with normalized attributes values.
  - INPUT VECTOR: $X = \{ x_1, x_2, \ldots. x_m\}$, where n is the number of attributes.
  - INPUT LAYER – there are as many nodes as class attributes i.e. as the length of the input vector.

- HIDDEN LAYER – neither its input nor its output can be observed from outside.
  - The number of nodes in the hidden layer and the number of hidden layers depends on implementation.

- OUTPUT LAYER – corresponds to the class attribute.
  - There are as many nodes as classes (values of the class attribute).
  - Ok, where k= 1, 2,.. n, where n is number of classes.



NEURON WITH ACTIVATION

- ANN is an electronic network of neurons based on the neural structure of the brain.
- The neuron is the basic information processing unit of a NN. It consists of:

1. A set of links, describing the neuron inputs, with weights $W_1, W_2, \ldots, W_m$
2. An adder function (linear combiner): computes the weighted sum of the inputs (real numbers):

$$y = \sum_{j=1}^{m} w_j x_j$$

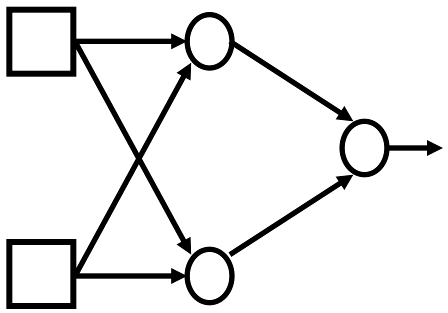
3. Activation function(also called squashing function): limits the output behavior of the neuron.

$$\sigma(y) = \frac{1}{1+e^{-y}}$$

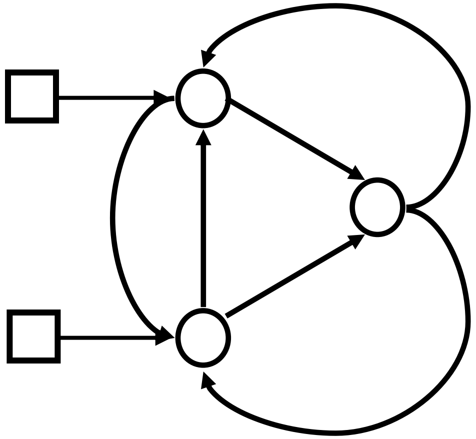
Where e = 2.718281828459045235

TWO TOPOLOGIES OF NEURAL NETWORK

- NN can be designed in a feed forward or recurrent manner
- In a **feed forward** neural network connections between the units do not form a directed cycle.
  - In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) & to the output nodes.
  - There are no cycles or loops or no feedback connections are present in the network,
  - connections extending from outputs of units to inputs of units in the same layer or previous layers.



- In **recurrent networks** data circulates back & forth until the activation of the units is stabilized.
  - Recurrent networks have a feedback loop where data can be fed back into the input at some point before it is fed forward again for further processing and final output.



TRAINING THE NEURAL NETWORK

- The purpose is to learn to generalize using a set of sample patterns where the desired output is known.

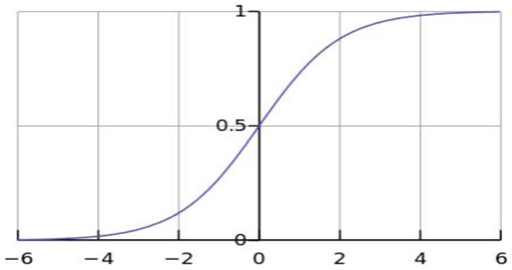
- ***Back Propagation*** is the most commonly used method for training multilayer feed forward NN.
  - *Back propagation* learns by iteratively processing a set of training data (samples).
  - For each sample, weights are modified to minimize the error between the desired output and the actual output.

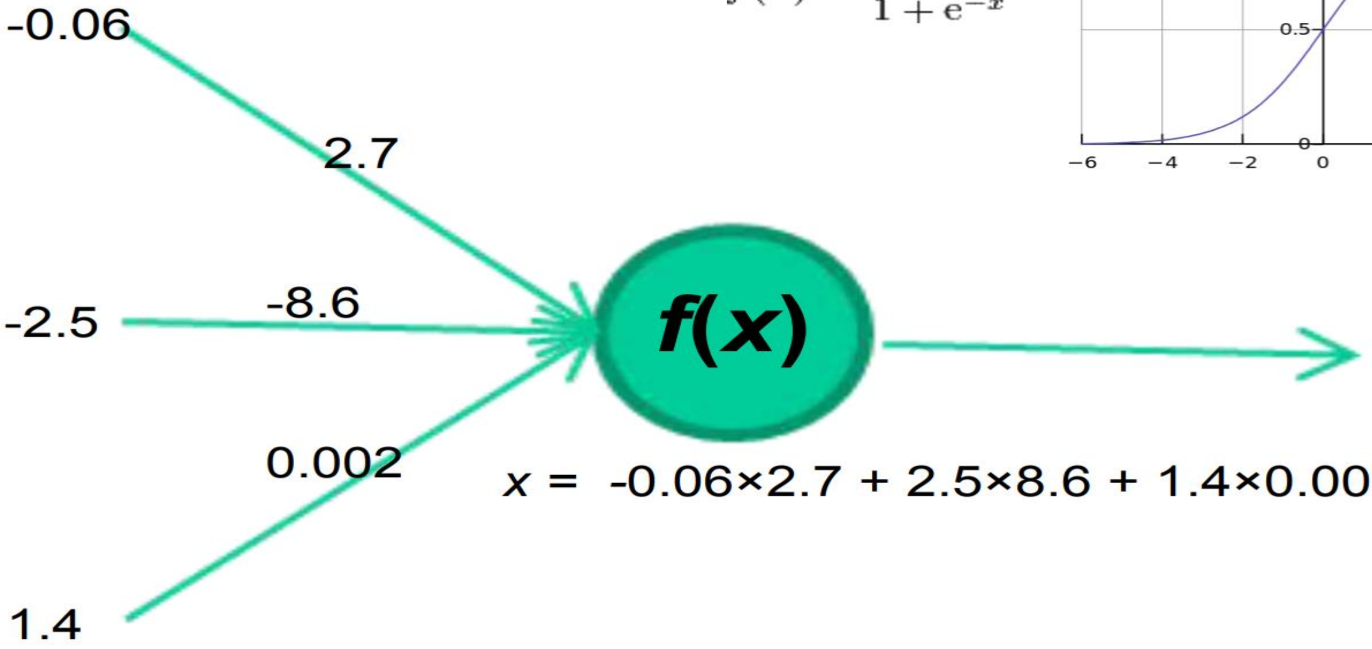
- After propagating an input through the network, the error is calculated and the error is propagated back through the network while the weights are adjusted in order to make the error smaller.

CONT…

- Example: how does NN algorithm learns?

$$f(x) = \frac{1}{1+e^{-x}}$$





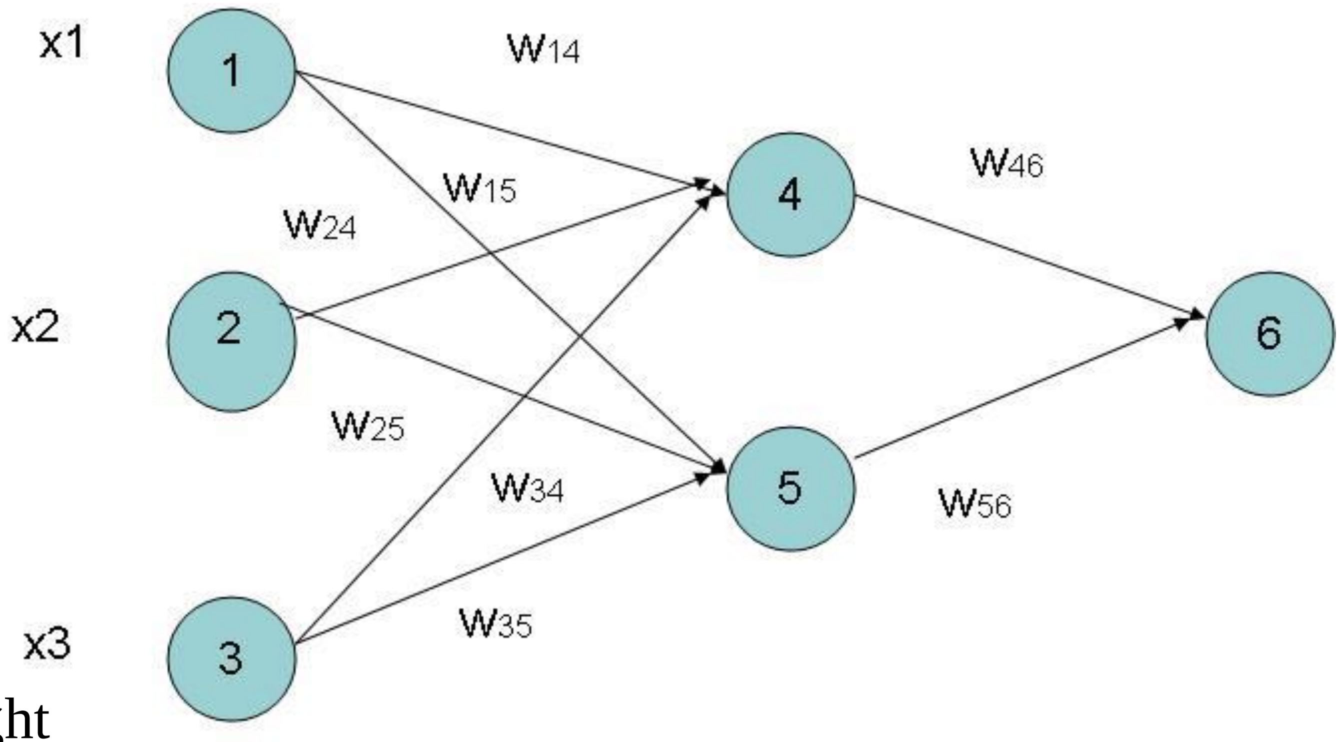
$$x = -0.06\times2.7 + 2.5\times8.6 + 1.4\times0.002 = 21.34$$

EXAMPLE OF BACK PROPAGATION

Input = 3, Hidden Neuron = 2, Output =1

Initialize weights :

Random Numbers from -1.0 to 1.0

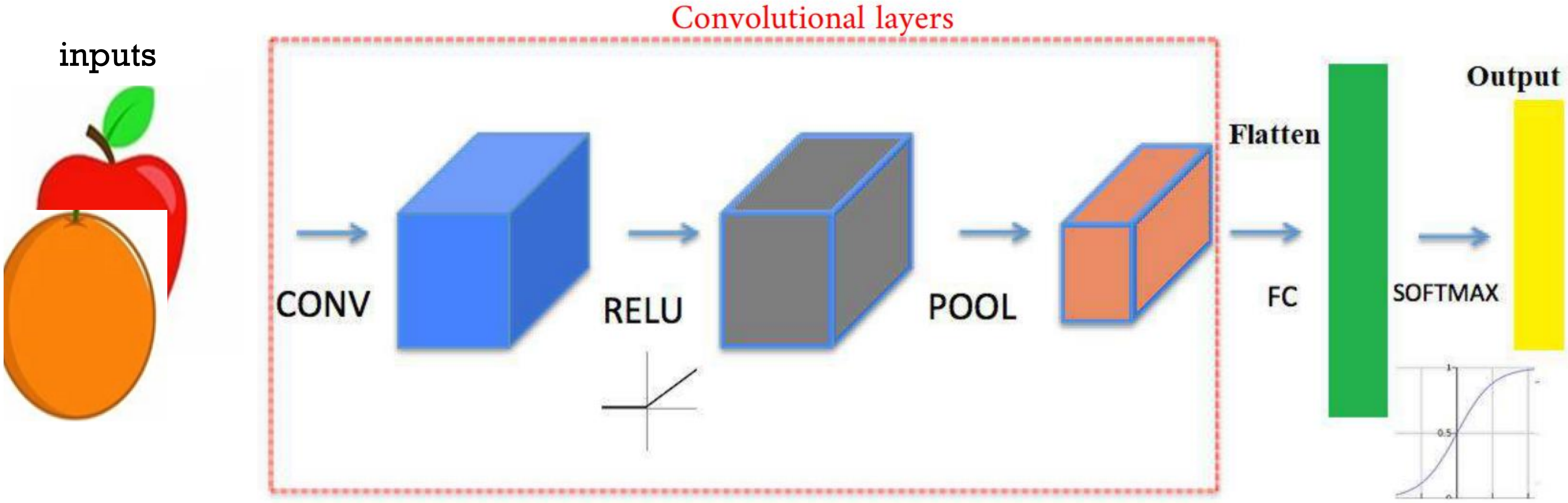


Initial Input and weight

x1	x2	x3	w14	w15	w24	w25	w34	w35	w46	w56
1	0	1	0.2	-0.3	0.4	0.1	-0.5	0.2	-0.3	-0.2

CONVOLUTIONAL NEURAL NETWORKS (CNNS)

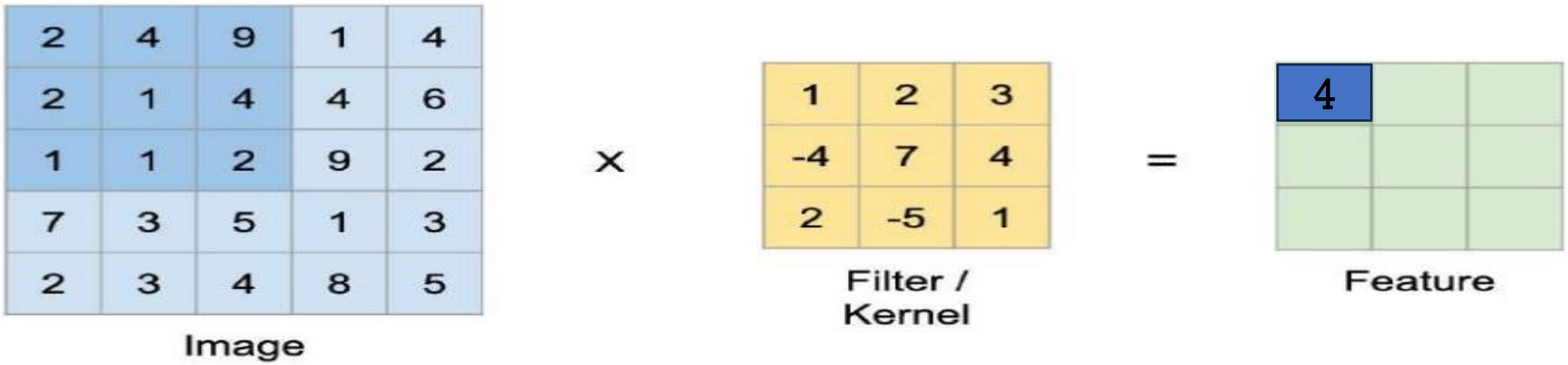
- CNNs are a special kind of multi-layer neural networks, designed to recognize visual patterns directly from pixel images with minimal preprocessing.

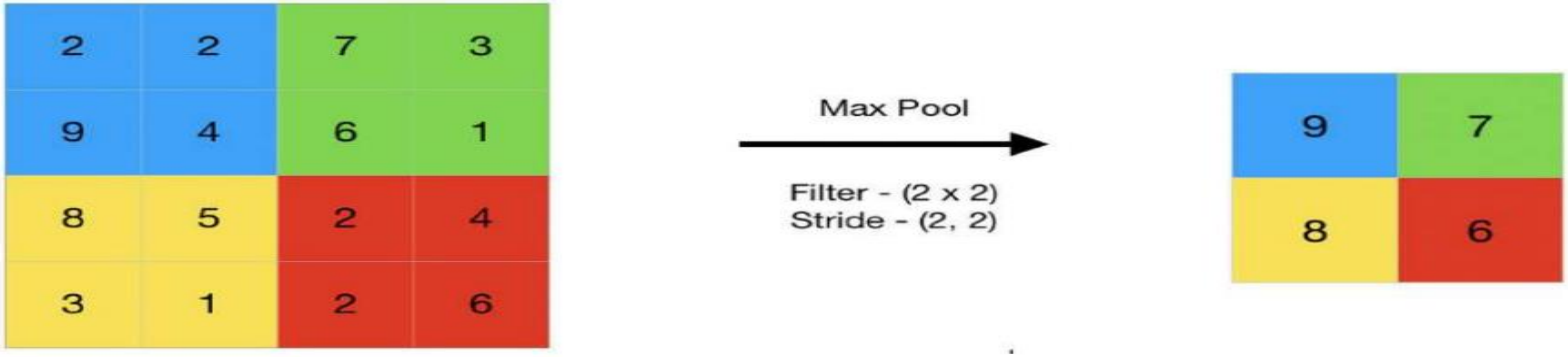


- Convolution: summarize/learn the presence of features in an input image .
- Pooling: A fixed operation that down sample the features.

CONT...

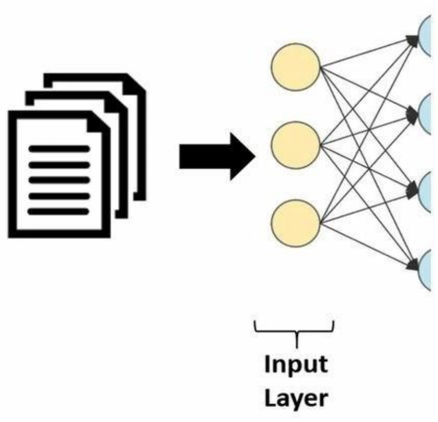
- (CONV) uses filters that perform convolution operations as it is scanning the input with respect to its dimensions.
- Its hyperparameters include the filter size and stride . The resulting output called **feature map/activation map**. **Example:**





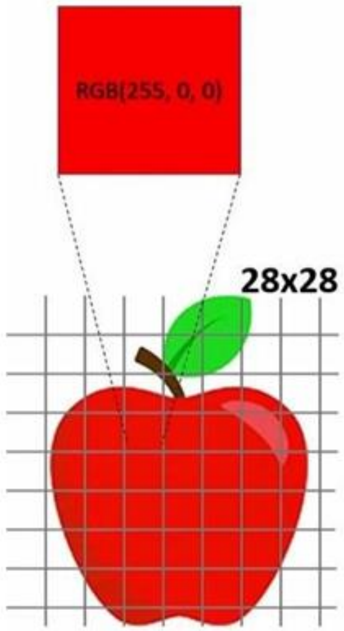
CONT…

- Responsible for holding the value from dataset.



Number of neurons = Number of features

example

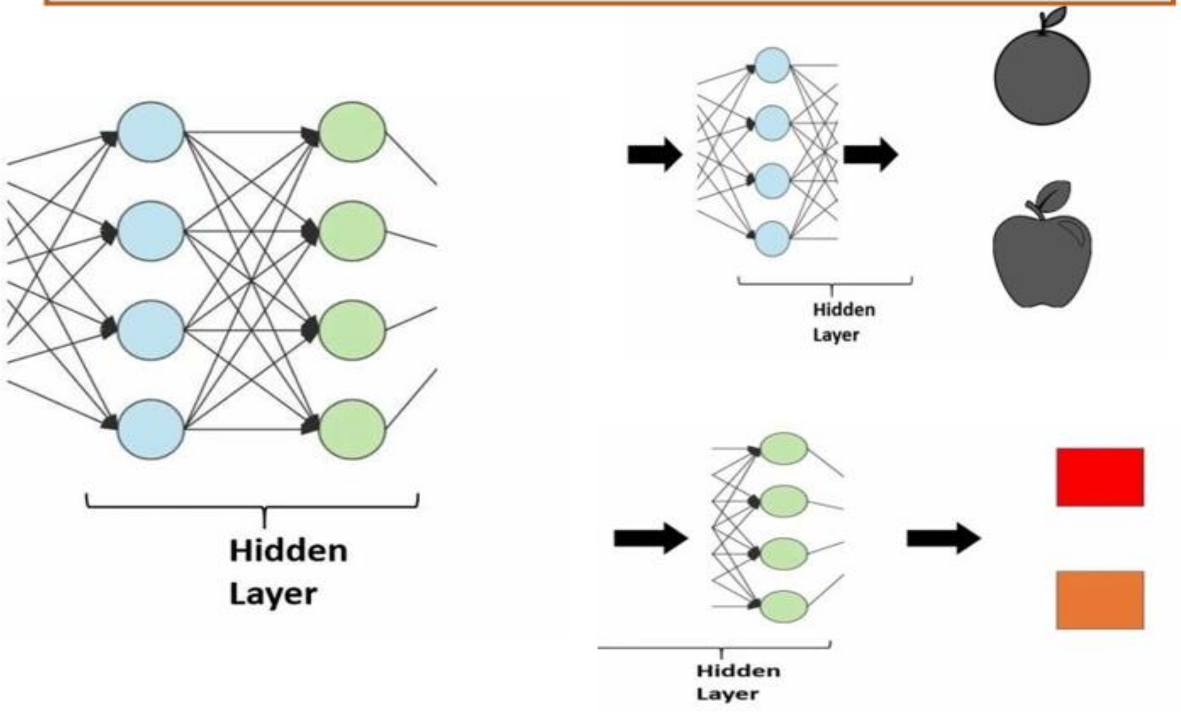


Total number of Neurons in Input layer = 28 x 28 x 3 = 2352

Each neuron hold color value of each pixel

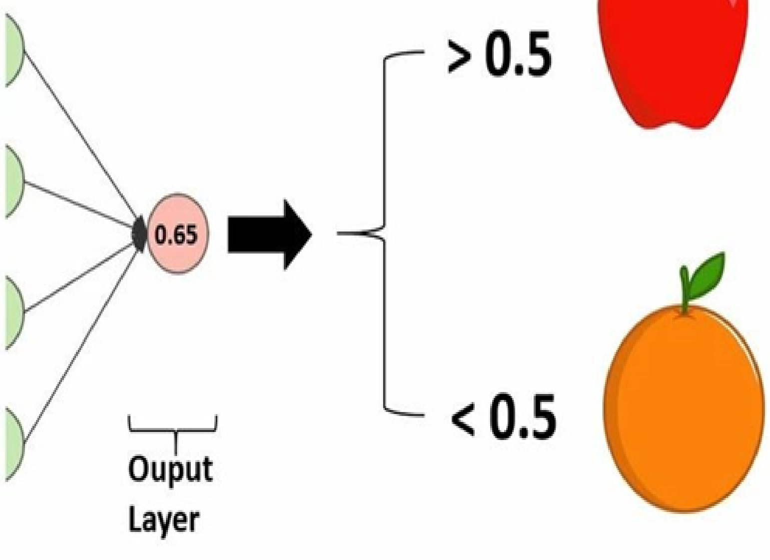
CONT…

Responsible for holding the pattern in them.



Hidden layer

The output neuron may be one which hold the value between 0 to1 representing the probability of the output



Output layer

CONT…

- Useful for learning complex data like handwriting, speech and image recognition.
    - **Pros** **Cons**

      - Hard to interpret

      + Can learn more complicated - Slow training timeclass boundaries

      + Fast application - Hard to implement: trial

      + Can handle large number of number of nodes

      and error for choosing features
    - Neural Network needs long time for training.
- Neural Network has a high tolerance to noisy and incomplete data.
- Conclusion: Use neural nets only if decision-trees fail.

THE ENDES OF CHAPTER 5