MACHINE LEARNING BOOTCAMP

GDG Algiers

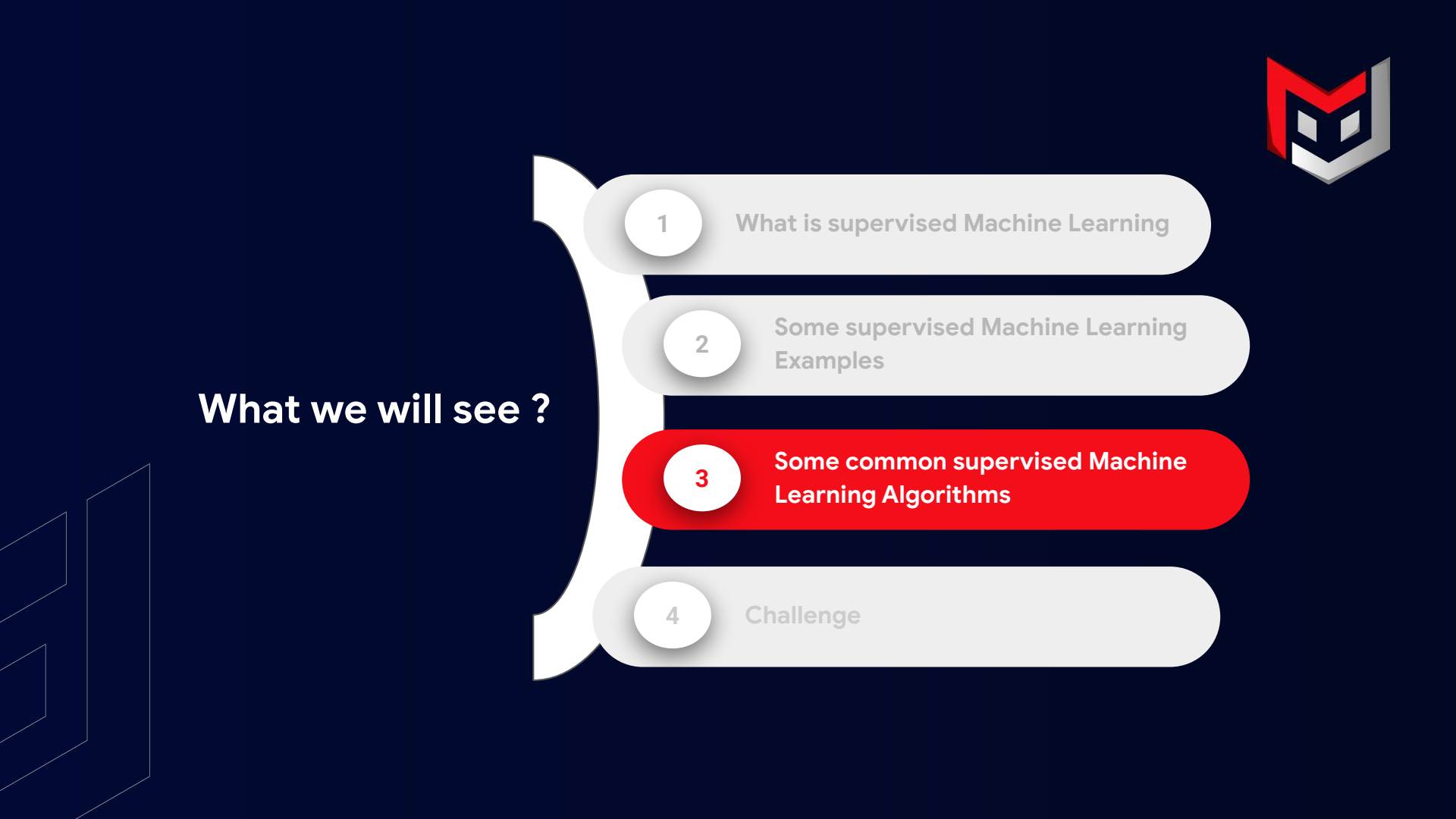
Women Techmakers Algiers

Hello there,

I am Nour Fadila Zemouri

Master 2 degree student at University of Algiers 1 - Benyousef Benkheda -

Field of study : ARTIFICIAL INTELLIGENCE


What we will see ?
1
What is supervised Machine Learning
2
Some supervised Machine Learning Examples
3
Some common supervised Machine Learning Algorithms
4
Challenge

Some common supervised Machine Learning Algorithms

k-Nearest Neighbors
Linear Regression
Logistic Regression
Support Vector Machines (SVMs)
Decision Trees and Random Forests
Neural networks

What we will see ?

1. What is supervised Machine Learning
2. Some supervised Machine Learning Examples
3. Some common supervised Machine Learning Algorithms
4. Challenge

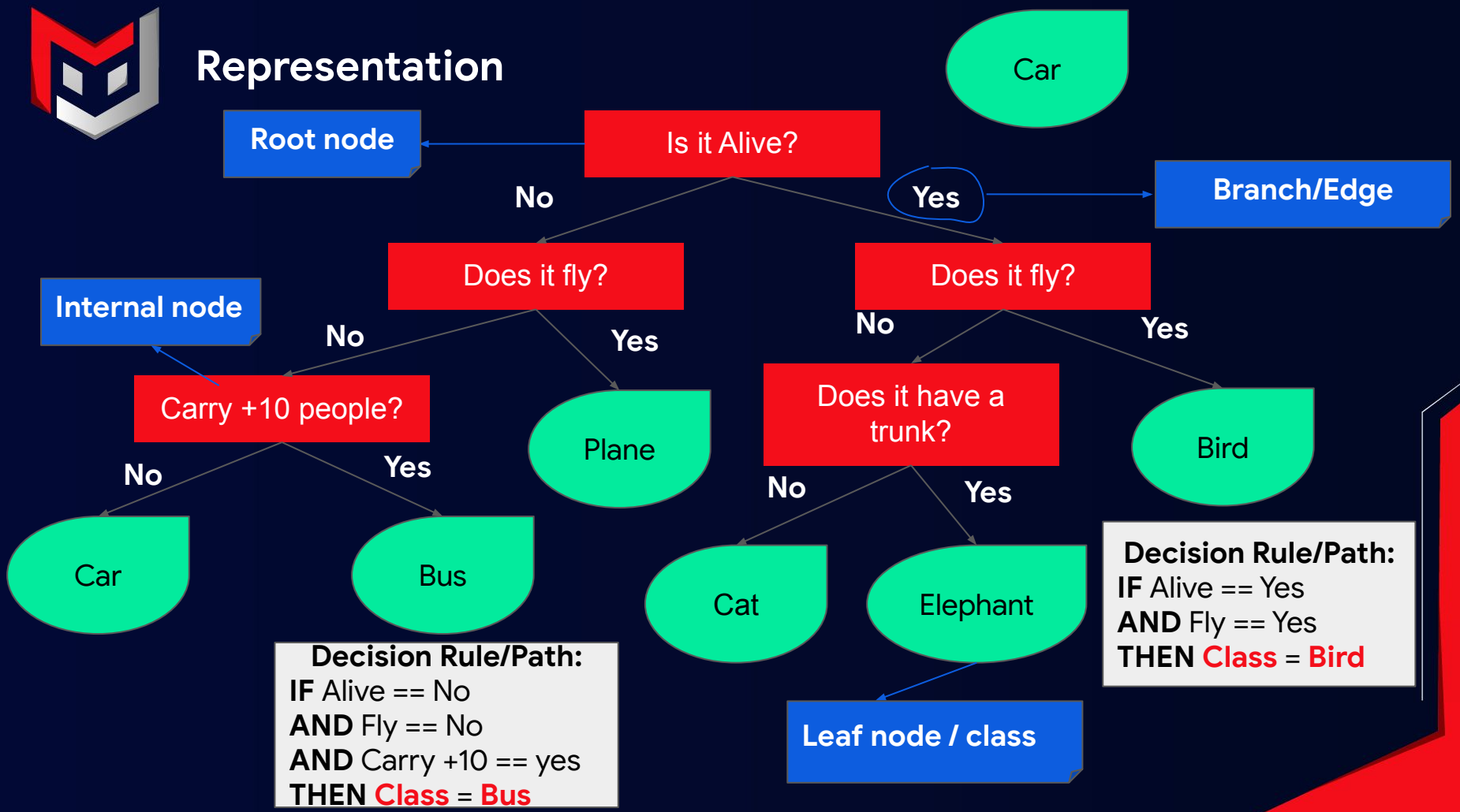
Decision Trees

Decision Tree?

Decision trees (DTs) non-parametric supervised learning method that is commonly used in data mining, it is also widely used in machine learning for both Classification and Regression.

The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features.

Representation



How does it work?

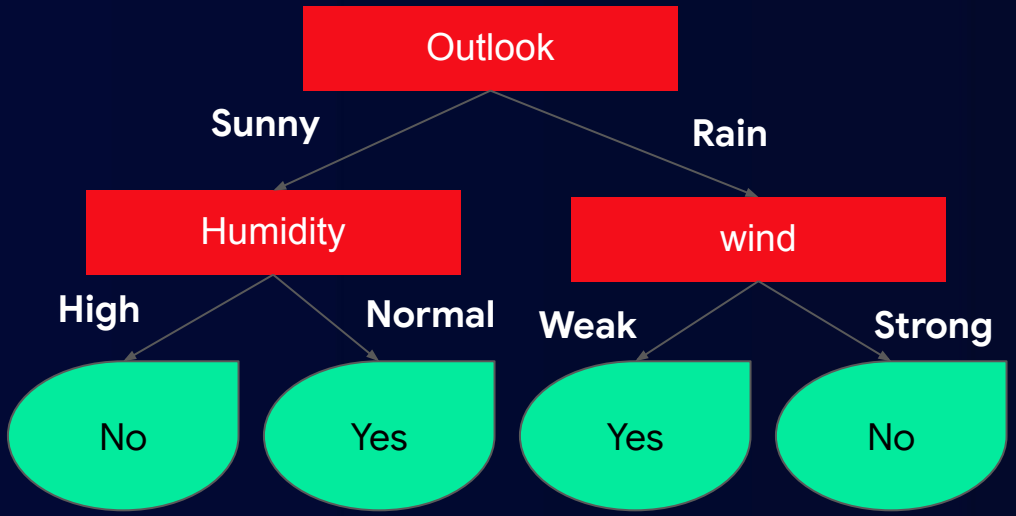
- The tree in the previous example is called Classification tree as the target is to classify objects in 6 classes.
- Regression trees are represented in the same manner, just they predict continuous values like price of a house.

- The tree is built in a Descending manner (from root to leaves) by choosing at each level an attribute to partition the data.

- The tree generation process takes place in two phases:
  1. Recursive to-down construction of the tree
  2. Pruning/Removing branches that introduce noise to improve classification quality (reduce error rate).

How does it work?

Growing a tree involves deciding on which feature to choose and what conditions to use for splitting, along with knowing when to stop.



Day	Outlook	Humidity	Wind	Play
1	Sunny	High	Weak	No
2	Sunny	High	Strong	No
3	Overcast	High	Weak	Yes
4	Rain	High	Weak	Yes
5	Rain	Normal	Weak	Yes
6	Rain	Normal	Strong	No
7	Overcast	Normal	Strong	Yes
8	Sunny	High	Weak	No
9	Sunny	Normal	Weak	Yes
10	Sunny	Normal	Strong	Yes
11	Overcast	High	Strong	Yes
12	Overcast	Normal	Weak	Yes
13	Rain	High	Strong	No
14	**Rain**	**High**	**Weak**	**?**

New Example : Day 14: Rain / High / Weak ?
⟶ Play = Yes

Construction: which attribute to split on?

Want to measure« purity » of the split

- More certain about Yes/No after the split:

→ Pure set (4 yes /0 No) => completely certain (100%)
→ Impure (3yes /3 No) => completely uncertain (50%)
→ must be symmetric: 4 Yes/ 0 No as pure as 0 Yes / 4 No

Construction: How to select an attribute?

Entropy H(S) : a measure of the amount of uncertainty in the dataset .

- Mathematical Representation of Entropy (Binary Classification):

$$H(S) = - p_{(+)} \log_2 p_{(+)} - p_{(-)} \log_2 p_{(-)}$$

S ... subset of training examples

$p_{(+)}$ / $p_{(-)}$ ... % of positive / negative examples in S

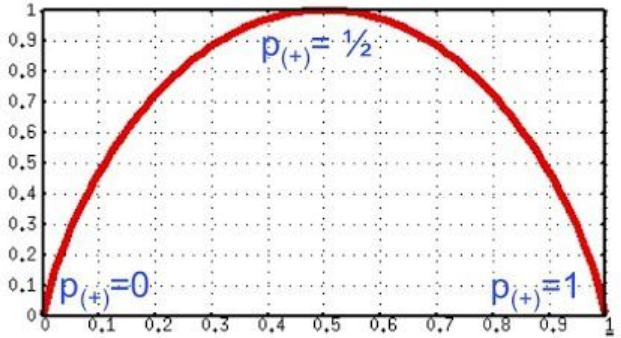
Construction: Example on how to calculate Entropy?

impure (3 yes / 3 no):

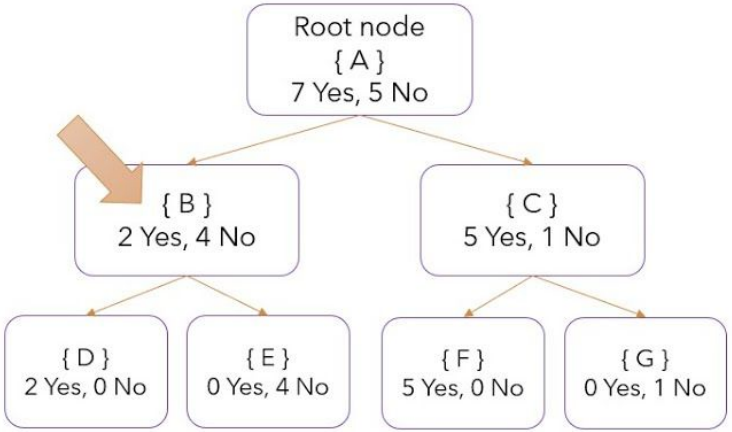
$$H(S) = -\frac{3}{6}\log_2\frac{3}{6} - \frac{3}{6}\log_2\frac{3}{6} = 1 \text{ bits}$$

pure set (4 yes / 0 no):

$$H(S) = -\frac{4}{4}\log_2\frac{4}{4} - \frac{0}{4}\log_2\frac{0}{4} = 0 \text{ bits}$$



Construction: Entropy In general



Entropy

$$Entropy(S) = \sum_{i=1}^{c} -pi\ log_2(pi)$$

$p(Yes) = \frac{P(Yes)}{P(Yes)+P(No)}$, for "Yes" component of node A

$$Entropy(B) = -\frac{2}{2+4}\ log_2\left(\frac{2}{2+4}\right) - \frac{4}{4+2}\ log_2\left(\frac{4}{4+2}\right)$$

$$= -\frac{2}{6}\ log_2\left(\frac{2}{6}\right) - \frac{4}{6}\ log_2\left(\frac{4}{6}\right)$$

$$= 0.92\ bits$$

$$Entropy(C) = -\frac{5}{6}\ log_2\left(\frac{5}{6}\right)\ -\frac{1}{6}\ log_2\left(\frac{1}{6}\right)$$

$$= 0.65 \text{ bits}$$

Construction: How to select an attribute?

Information GAIN: is the measure of uncertainty in S that was reduced after splitting set S on attribute A.

- Mathematical Representation:

$$Gain(S,A) = H(S) - \sum_{V \in Values(A)} \frac{|S_V|}{|S|} H(S_V)$$

A : Attribute in the Dataset

V : Possible Values of A

S : Set of Examples X

S_v: Subset where $X_A = V$

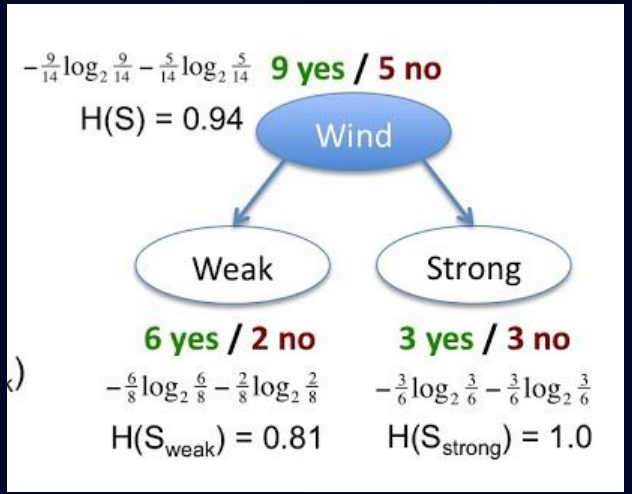
Construction: Example on how to calculate Information GAIN?

Gain (S, Wind)

$= H(S) - {}^{8}/_{14}\, H(S_{weak}) - {}^{6}/_{14}\, H(S_{weak})$

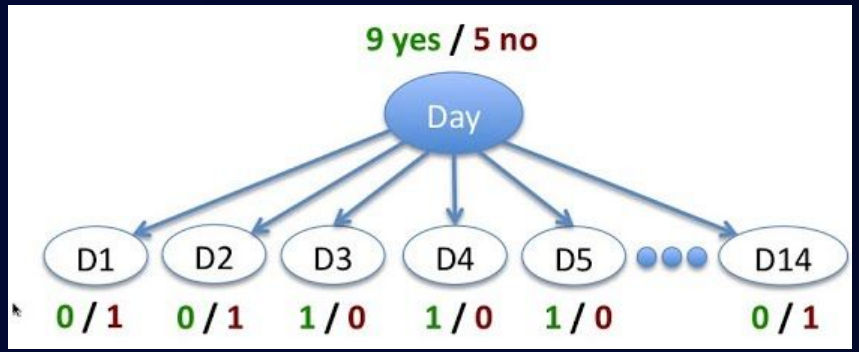
$= 0.94 - {}^{8}/_{14} * 0.81 - {}^{6}/_{14} * 1.0$

$= 0.049$



Construction: Problem with Information Gain?

- Information Gain Biased towards attributes with many values



- The attribute Day has many values, because it uniquely identifies each day, but we do not want to include it in the decision tree.
- Deciding whether or not to play tennis based on the day number is unlikely to generalize to days not seen before (overfitting).

Solution ?

Information Gain Ratio

Construction: Information Gain Ratio

- To counter this problem, Ross Quinlan proposed to choose the attribute with highest information gain ratio from among the attributes whose information gain is average or higher.

$$SplitEntropy(S,A) = -\sum_{V \in Values(A)} \frac{|S_V|}{|S|} \log \frac{|S_V|}{|S|}$$

$$GainRatio(S,A) = \frac{Gain(S,A)}{SplitEntropy(S,A)}$$

A : Candidate Attribute

V : Possible Values of A

S : Set of Examples X

S_v: Subset where $X_A = V$

Construction: ID3 Algorithm

- ID3 (Iterative Dichotomiser 3) is an algorithm invented by Ross Quinlan used to generate a decision tree from a dataset.
- ID3 iteratively (repeatedly) dichotomizes(divides) features into two or more groups at each step.
- At each iteration the best feature is selected to create a node.
- Most generally ID3 is only used for classification problems with nominal features only
- ID3 uses the concepts of entropy and gain to determine which attribute to test in each node.
- ID3 cannot support examples in which attribute values are missing.

Construction: ID3 Algorithm

- The steps of ID3 Algorithm:

1. The ID3 algorithm begins with the original set S as the root node.
2. On each iteration, the algorithm iterates through every unused attribute of the set S and calculates the entropy or the information gain of that attribute.
3. It then selects the attribute which has the smallest entropy (or largest information gain) value.
4. The set S is then partitioned by the selected attribute to produce subsets of the data.
5. Repeat for the remaining features until we run out of all features, or the decision tree has all leaf nodes

Construction: Overfitting

- Sometimes it looks like the tree memorized the training data set, which will give 100% accuracy on the training data set

→ Our tree has grown much !!
→ Solution?

Pruning Decision Trees

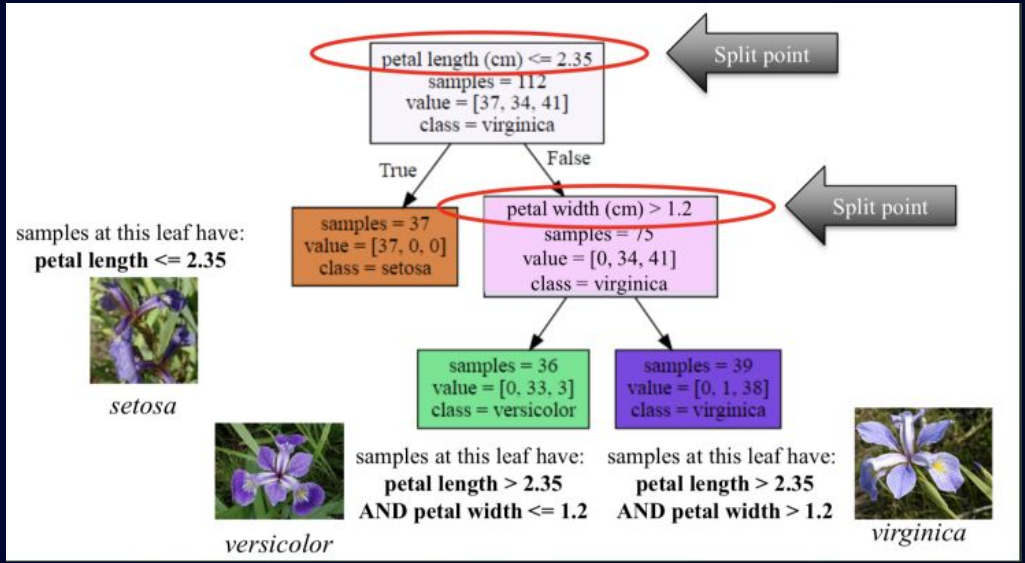
Construction: Pruning

- Pruning can start at either root or the leaves.
- The simplest method of pruning starts at leaves and removes each node with most popular class in that leaf, this change is retained if it does not deteriorate accuracy. It is also called reduced error pruning.

- This is done by separating the actual training set into two sets: Training data set D and Validation data set V.
- Prepare the decision tree using the training data set D.
- Then continue prune the tree accordingly to optimize the accuracy of the validation data set V

Continuous Attribute

- To deal with continuous values attributes, create a split
- Example: (Temperature > 50,4) = True/ False
- To take into account numeric attributes, ID3 has been extended in C4.5.

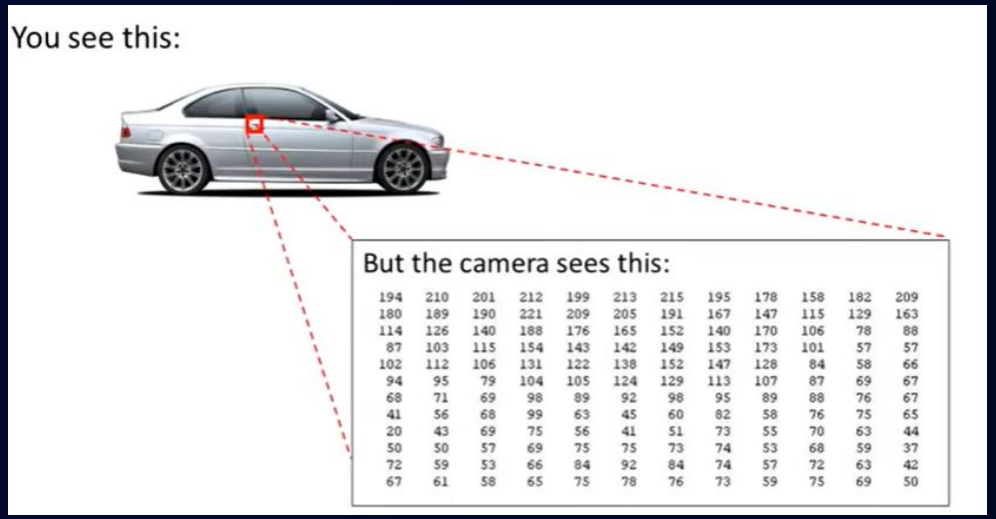


Let practice!!

Neural Networks

Problem?

Computer vision



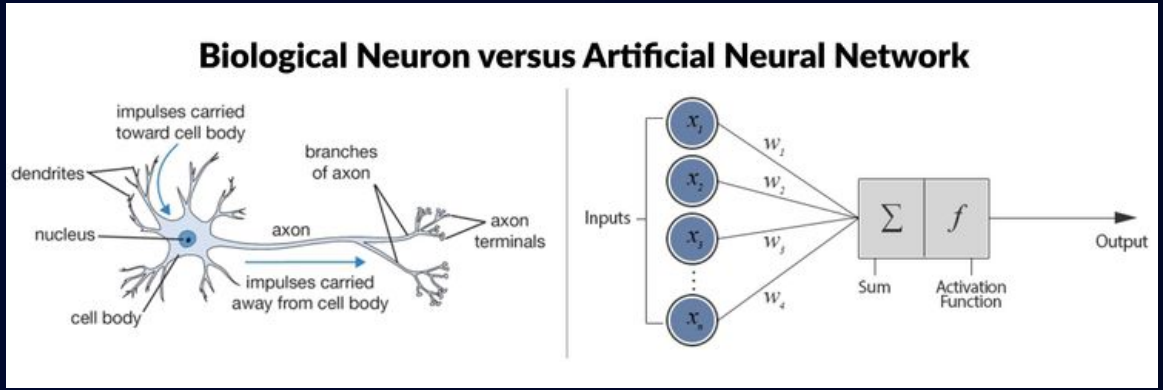
39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40	United-States	<=50K
50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	13	United-States	<=50K
38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0	40	United-States	<=50K
53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0	40	United-States	<=50K
28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0	40	Cuba	<=50K
37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	White	Female	0	0	40	United-States	<=50K
49	Private	160187	9th	5	Married-spouse-absent	Other-service	Not-in-family	Black	Female	0	0	16	Jamaica	<=50K
52	Self-emp-not-inc	209642	HS-grad	9	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	45	United-States	>50K
31	Private	45781	Masters	14	Never-married	Prof-specialty	Not-in-family	White	Female	14084	0	50	United-States	>50K
42	Private	159449	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	5178	0	40	United-States	>50K
37	Private	280464	Some-college	10	Married-civ-spouse	Exec-managerial	Husband	Black	Male	0	0	80	United-States	>50K
30	State-gov	141297	Bachelors	13	Married-civ-spouse	Prof-specialty	Husband	Asian-Pac-Islander	Male	0	0	40	India	>50K
23	Private	122272	Bachelors	13	Never-married	Adm-clerical	Own-child	White	Female	0	0	30	United-States	<=50K
32	Private	205019	Assoc-acdm	12	Never-married	Sales	Not-in-family	Black	Male	0	0	50	United-States	<=50K
40	Private	121772	Assoc-voc	11	Married-civ-spouse	Craft-repair	Husband	Asian-Pac-Islander	Male	0	0	40	?	>50K
34	Private	245487	7th-8th	4	Married-civ-spouse	Transport-moving	Husband	Amer-Indian-Eskimo	Male	0	0	45	Mexico	<=50K
25	Self-emp-not-inc	176756	HS-grad	9	Never-married	Farming-fishing	Own-child	White	Male	0	0	35	United-States	<=50K
32	Private	186824	HS-grad	9	Never-married	Machine-op-inspct	Unmarried	White	Male	0	0	40	United-States	<=50K
38	Private	28887	11th	7	Married-civ-spouse	Sales	Husband	White	Male	0	0	50	United-States	<=50K
43	Self-emp-not-inc	292175	Masters	14	Divorced	Exec-managerial	Unmarried	White	Female	0	0	45	United-States	>50K
40	Private	193524	Doctorate	16	Married-civ-spouse	Prof-specialty	Husband	White	Male	0	0	60	United-States	>50K
54	Private	302146	HS-grad	9	Separated	Other-service	Unmarried	Black	Female	0	0	20	United-States	<=50K
35	Federal-gov	76845	9th	5	Married-civ-spouse	Farming-fishing	Husband	Black	Male	0	0	40	United-States	<=50K
43	Private	117037	11th	7	Married-civ-spouse	Transport-moving	Husband	White	Male	0	2042	40	United-States	<=50K
59	Private	109015	HS-grad	9	Divorced	Tech-support	Unmarried	White	Female	0	0	40	United-States	<=50K

Big database with a lot of features and instances

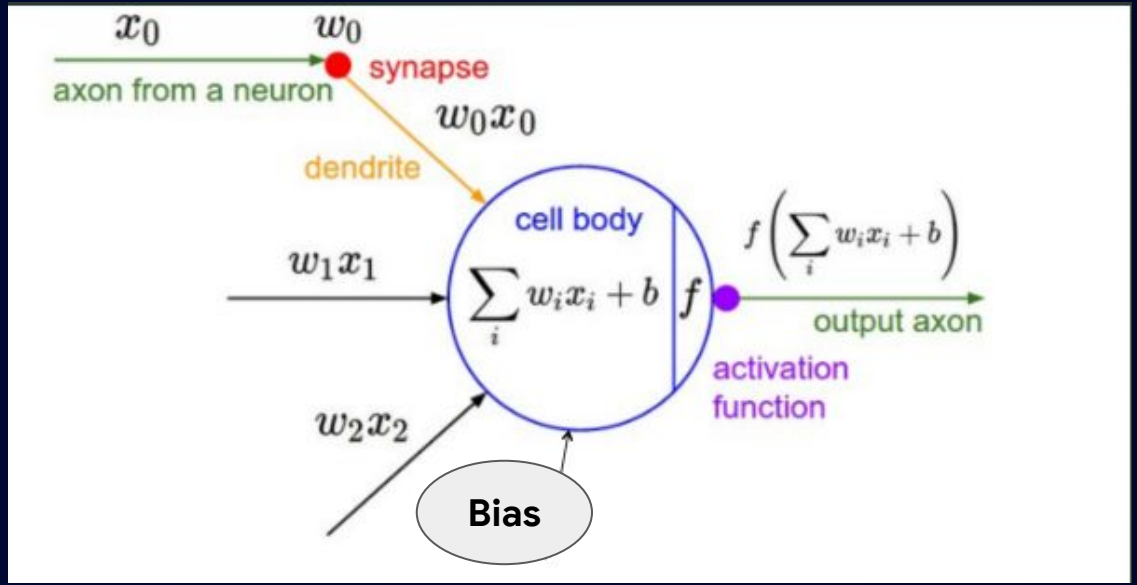
What is neural network

A neural network is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates.

- Nodes (neurons)
- Adaptive Weights (synaptic strength)
- Interconnection pattern between layers of neurons



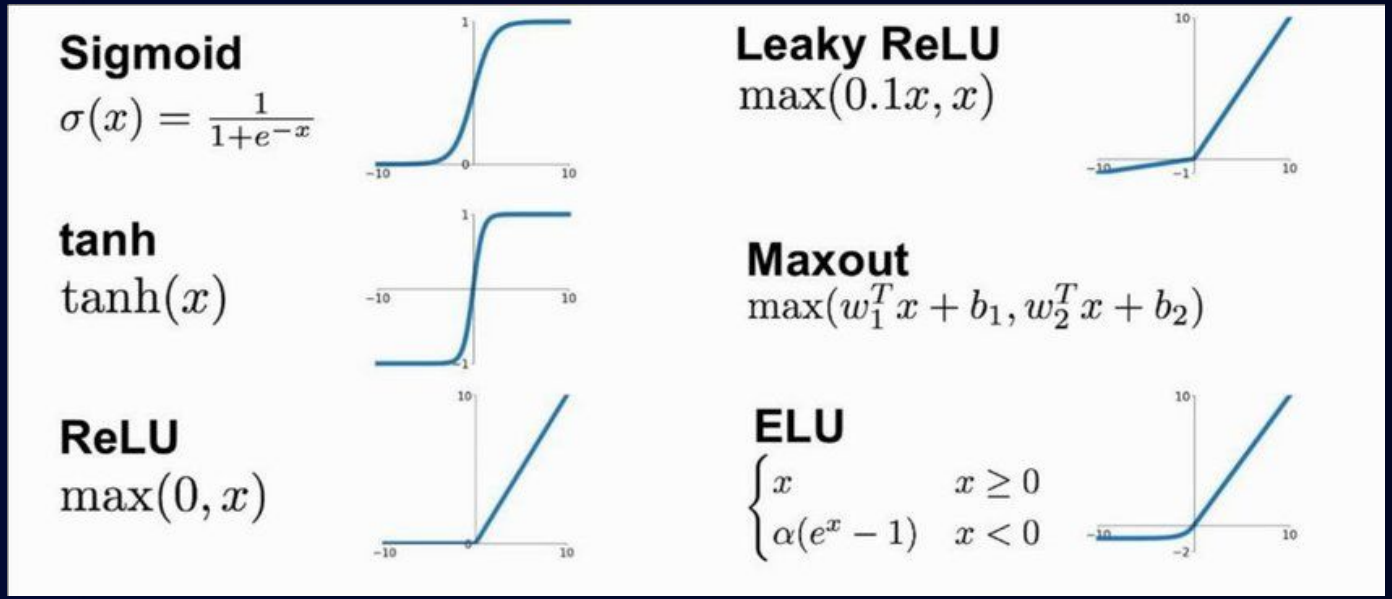
Let us break down our neuron



Activation function

The activation function is a mathematical “gate” in between the input feeding the current neuron and its output going to the next layer. It is function of transformation that maps the input signals into output signals that are needed for the neural network to function.

- Most commonly used activation functions:

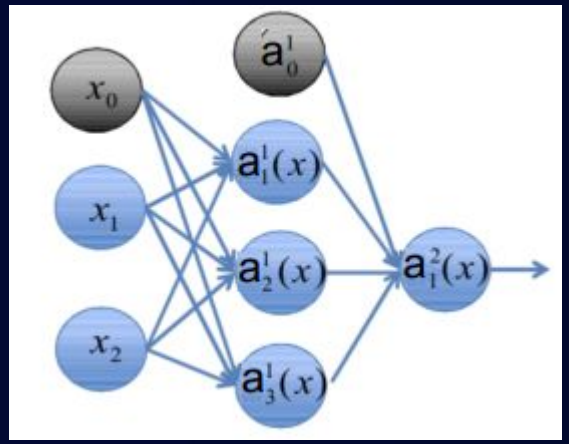


Neural networks tasks

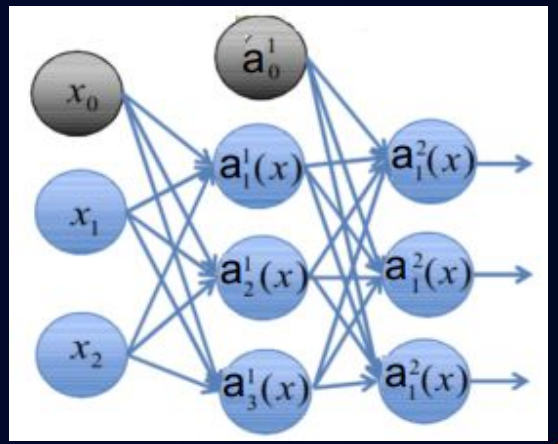
Neural Network Can Do Many Tasks by changing the Activation Function of Output Layer

- Binary Classification
- Multiclass Classification.
- Single Output Regression
- Multiple Output Regression
- Binary Multi-Label Classification.

Binary Classification/ Single Output Regression

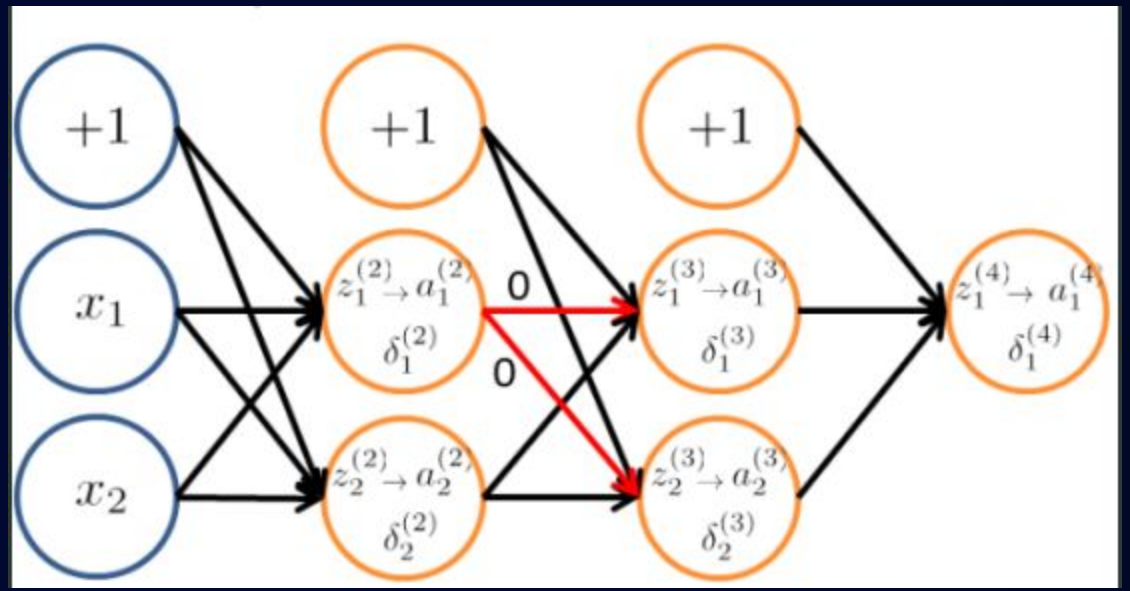


Multiclass Classification/ Multiple Output Regression



How does it work?

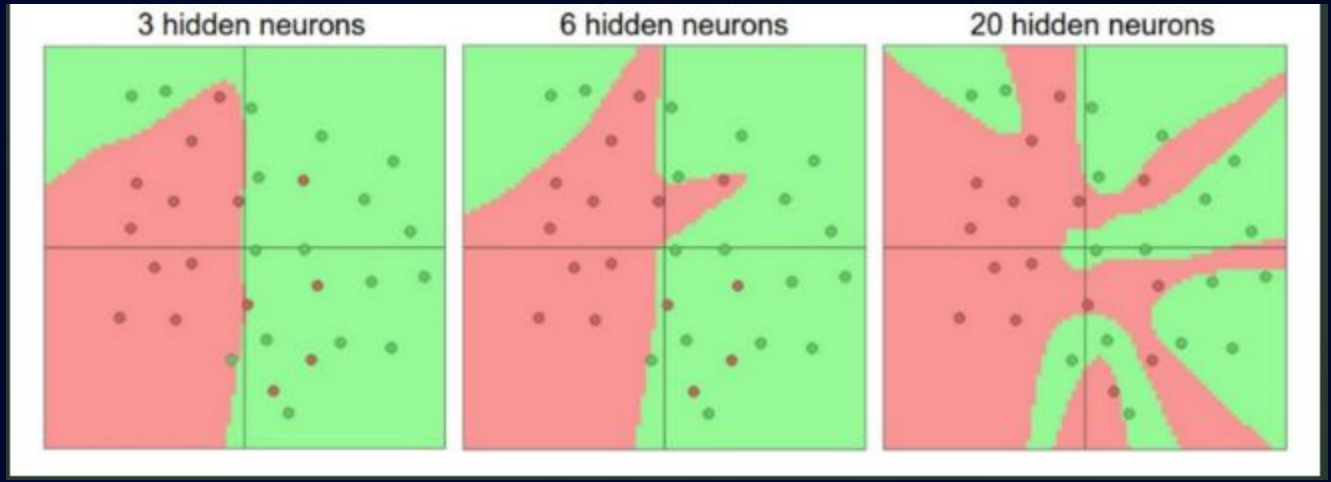
Forward Propagation



Backpropagation

Overfitting in neural networks

- The capacity of the network increases with more hidden units and more hidden layers: Neural Nets can have lots of parameters (weights) 10s of millions in practice!
- BUT “ With great power comes great overfitting.” – Boris Ivanovic, 2016
- Overfitting is a risk, especially if limited data.



Preventing Overfitting in neural networks

- Regularization.
- Standard ways to limit the capacity of a neural net:
  - ❖ Limit the number of hidden units.
  - ❖ Weight Decay: Limit the size of the Weights.
  - ❖ Early Stopping : Stop the learning before it has time to overfit.
    - At each iteration of gradient descent, check the network cost on validation set.
    - Stop once validation error starts to increase (although train error will still be decreasing)

Let practice!!

What we will see ?

1. What is supervised Machine Learning
2. Some supervised Machine Learning Examples
3. Some common supervised Machine Learning Algorithms
4. Challenge

In this challenge you will have to implement Linear Regression algorithm for one variable using what we've seen in the first session.

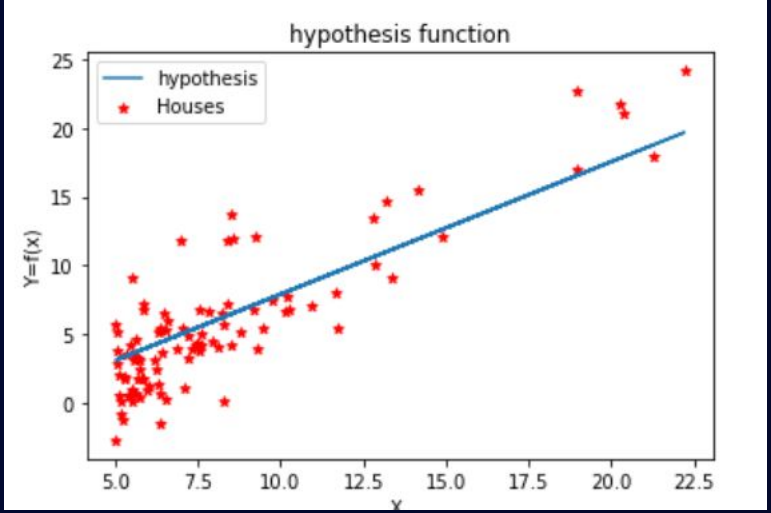
Database:

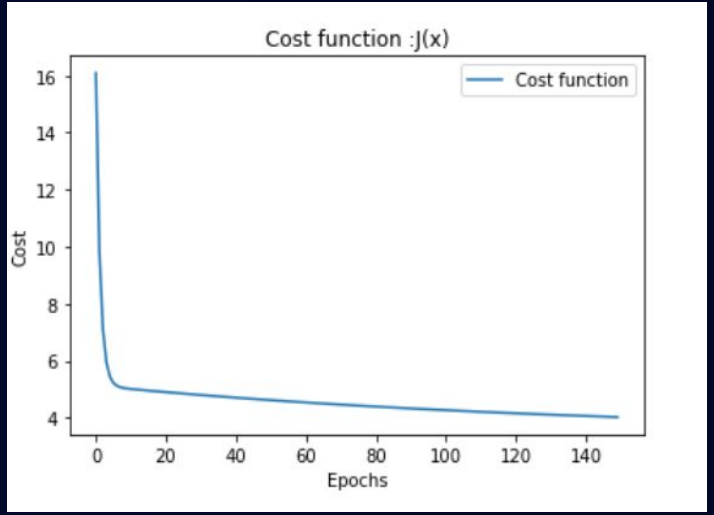
- https://drive.google.com/drive/folders/1POXWinRoZmj3KOymX_X87Bm9hvF_uzOK?usp=sharing

Script skeleton:

- https://colab.research.google.com/drive/1mG8aRXDYFGGo6K_3FkZfJ45PBKp2OfNI?usp=sharing

The desired Output:





Resources

- [Machine Learning | Andrew Ng]: https://www.youtube.com/watch?v=PPLop4L2eGk&list=PLLssT5z_DsK-h9vYZkQkYNWcItqhlRJLN

Questions ?

Thank you

MACHINE LEARNING

BOOTCAMP