Marketing and Business Analytics
A&M-COMMERCE

Machine Learning

Dr. Vinayaka Gude

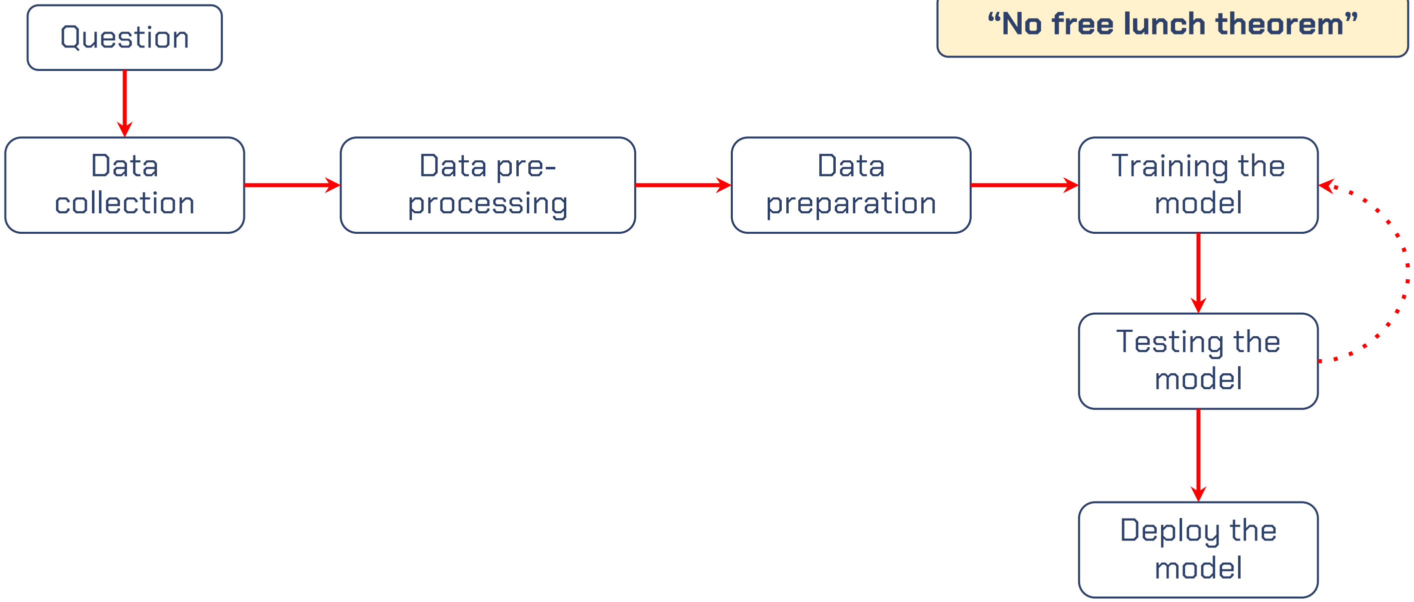
Agenda

- Machine Learning process
- Types of Machine Learning
- Algorithms
- Trends and Challenges of ML

Topic 1

Machine Learning Process

Machine Learning process



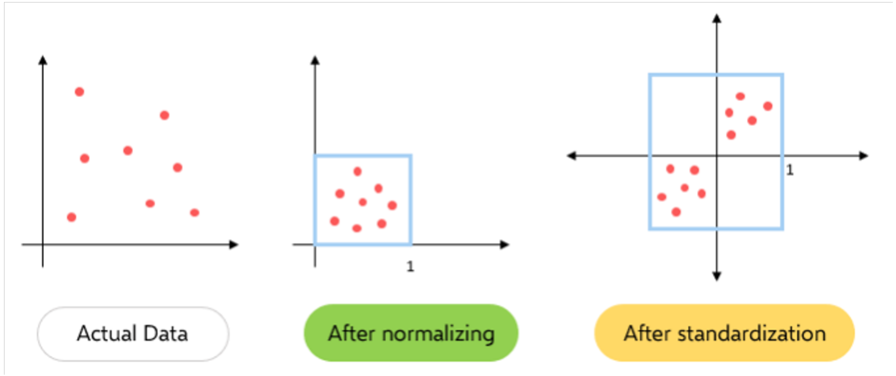
“No free lunch theorem”

Data pre-processing

Duplicates, Errors & Missing values

Reclassifying labels

Standardization and normalization



https://becominghuman.ai/what-does-feature-scaling-mean-when-to-normalize-data-and-when-to-standardize-data-c3de654405ed

Exploratory analysis

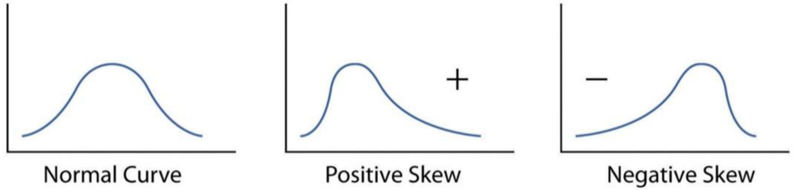
Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to ***discover patterns***, to ***spot anomalies***, to ***test hypothesis*** and to ***check assumptions*** with the help of **summary statistics** and **graphical representations**.

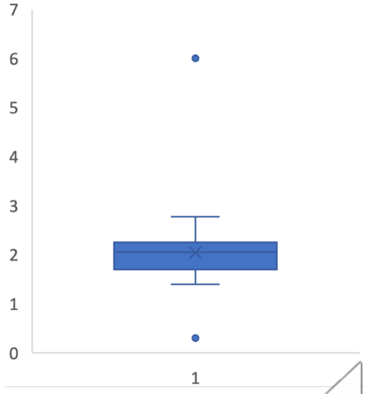
Methods of central tendency and deviation

Correlation → *remove correlated variables to improve your model*

Outliers → *Box plot, histogram, scatter plot*

Skewness → *Distribution*



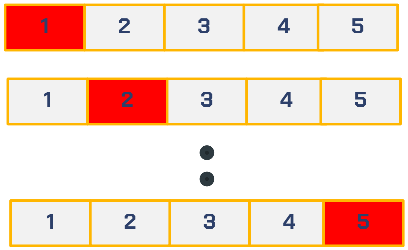


Data preparation

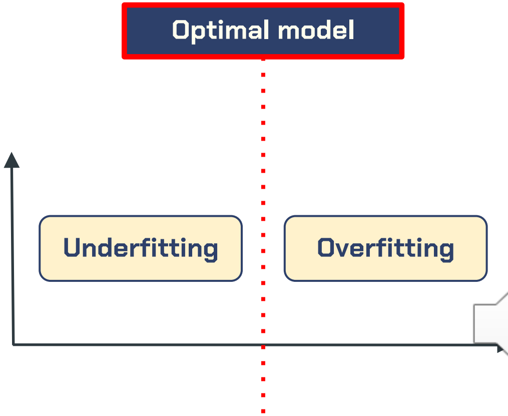
Dataset

Training	**Validation**	**Testing**
Train the models **60%**	Tune hyper parameters and select the best model **20%**	Evaluate the model performance **20%**

If the model performance is significantly better on the validation set than the testing data → **model is overfitting**



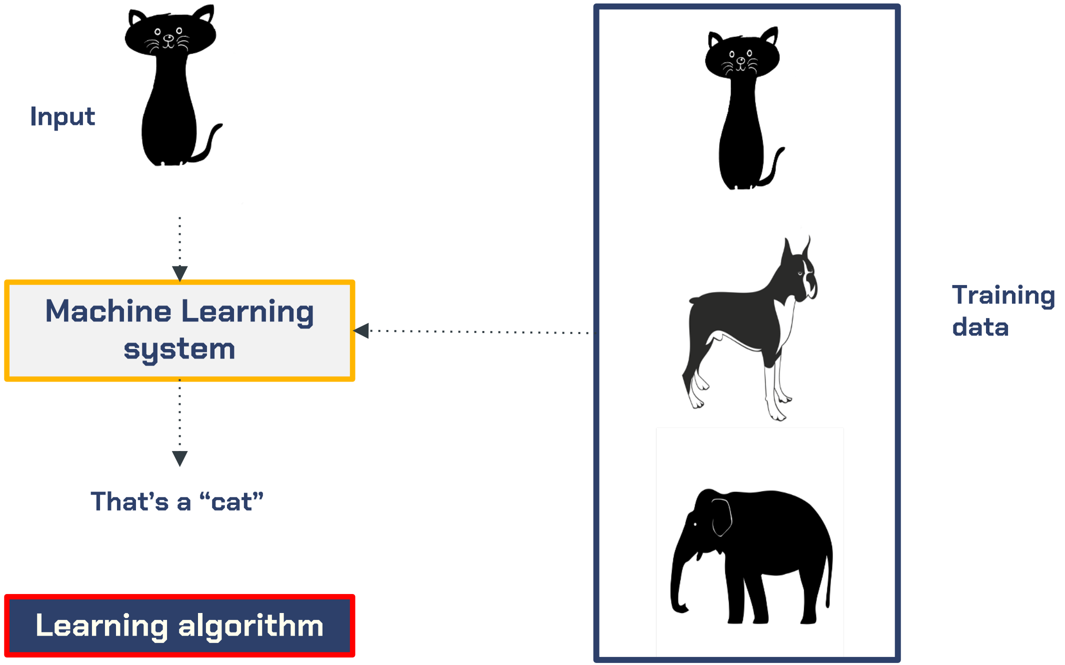
Cross-validation



Topic 2

What is Machine Learning?

Machine Learning



Types of learning

Supervised learning

Unsupervised learning

Semi-supervised learning

Reinforcement learning

Transfer learning

Supervised learning

Function

Cat

Dog

Elephant

inputs

outputs

Logistic Regression

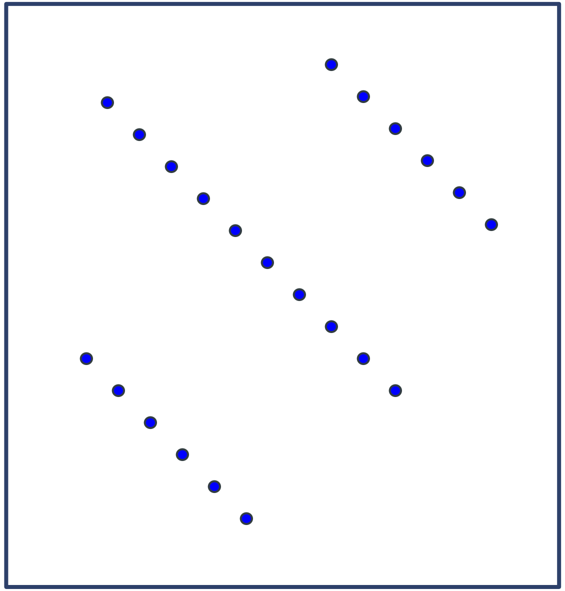
Support vector Machine

Random Forest

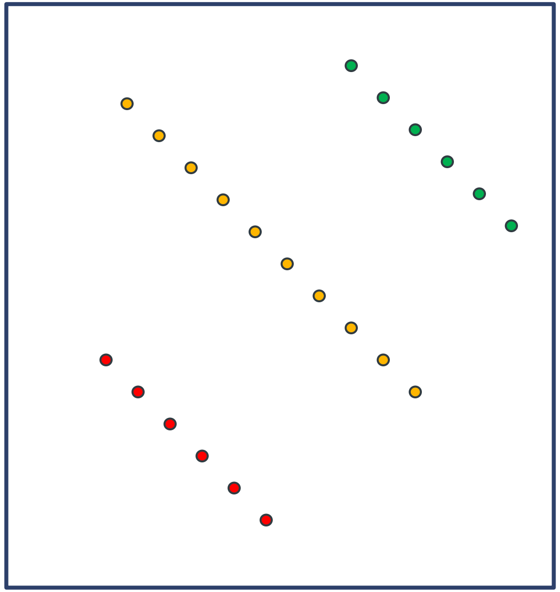
Neural Network

And many more..

Unsupervised learning

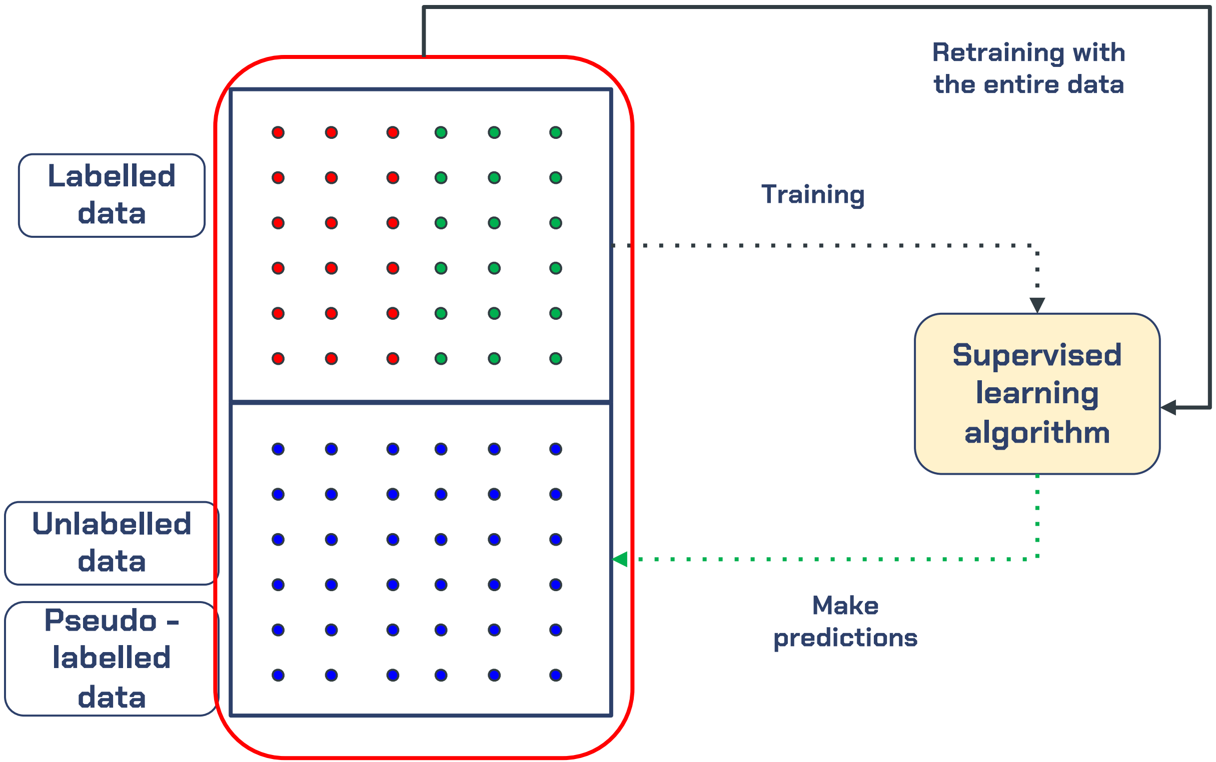


No labels

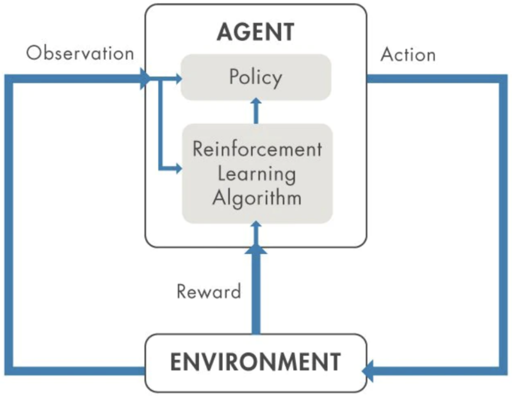


Clustering

Semi-supervised learning

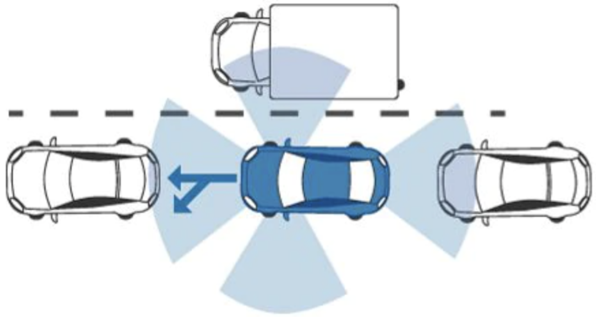


Reinforcement learning



- Observe
- Select a policy
- Action
- Get reward
- Update policy
- Iterate until optimal policy is found

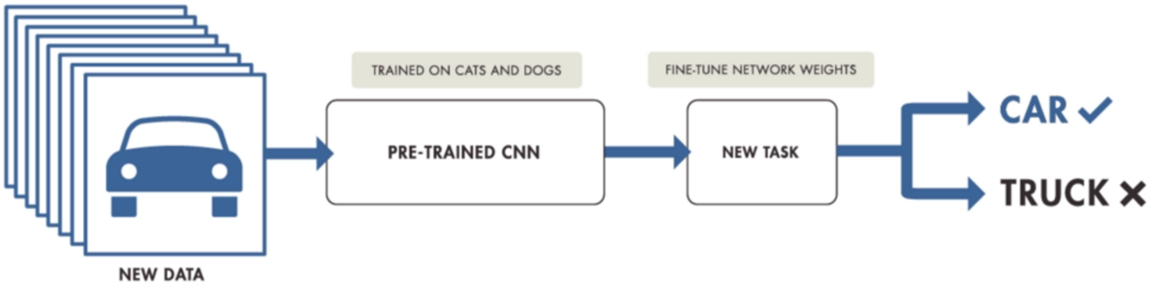
Autonomous driving



Deep neural networks

https://fr.mathworks.com/company/newsletters/articles/reinforcement-learning-a-brief-guide.html

Transfer learning



Saves times - no need to build models from scratch.

Eliminates the need for **huge training dataset**

Improves **generalizability**

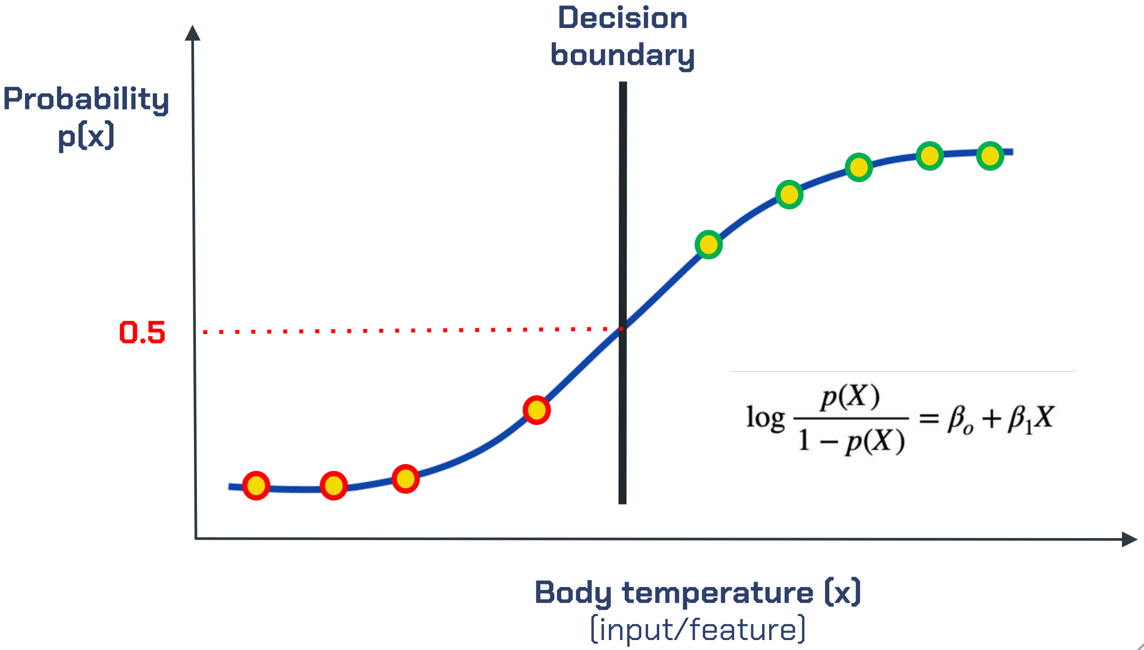
https://medium.datadriveninvestor.com/introducing-transfer-learning-as-your-next-engine-to-drive-future-innovations-5e81a15bb567

Topic 3

Logistic Regression

Logistic Regression

Binomial classification



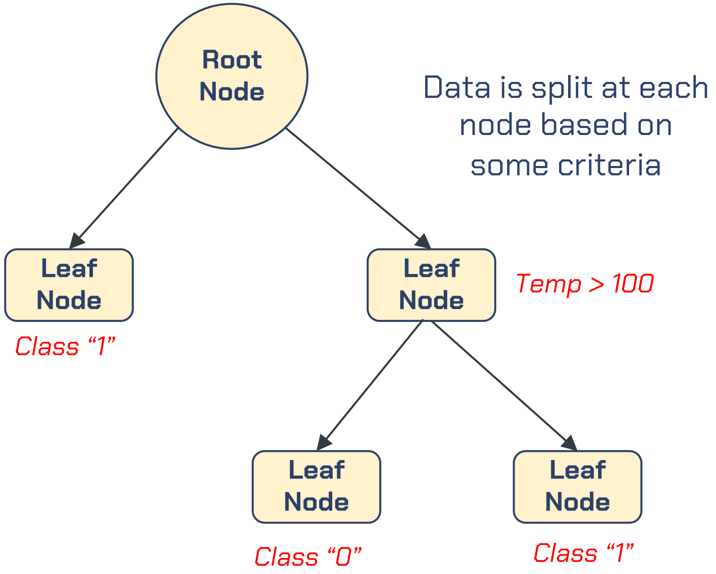
$$\log \frac{p(X)}{1 - p(X)} = \beta_o + \beta_1 X$$

Topic 4

Random forest

Random forest

Objective - minimize dissimilarity in terminal notes

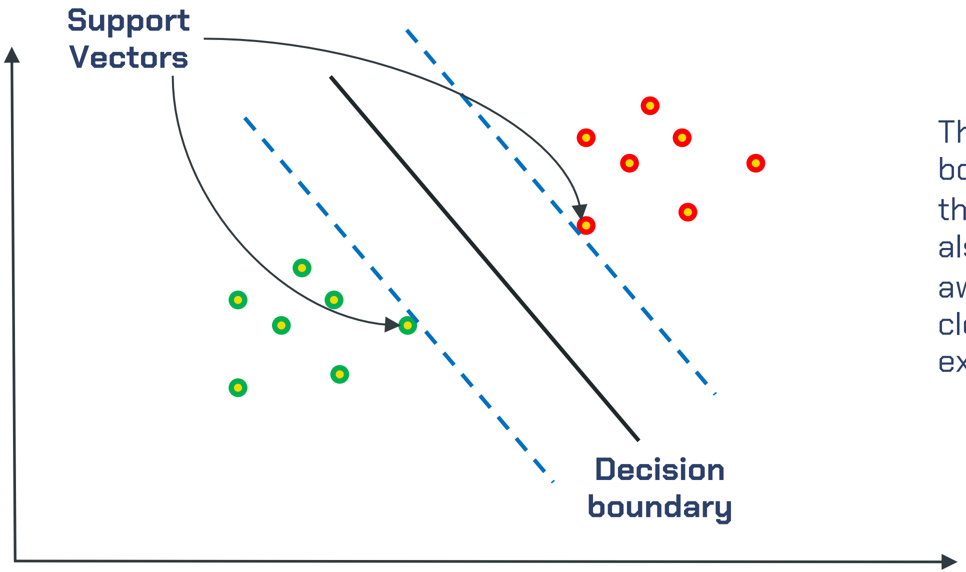


Topic 5

Support Vector Machine

Support Vector Machine

A **linear SVM classifier** fits the "***widest possible street***" between the classes.



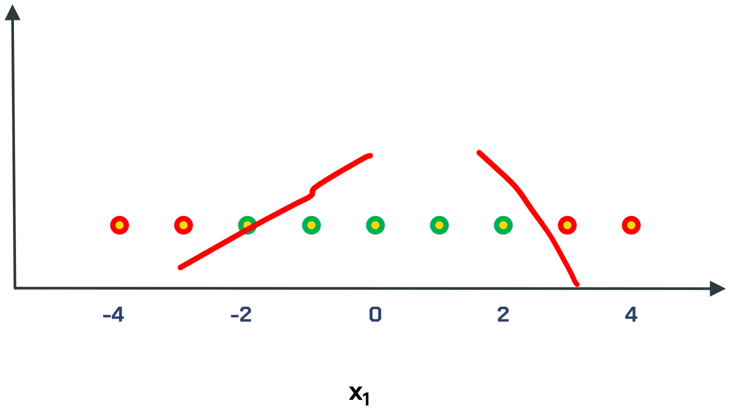
The decision boundary separates the two classes, but also stays as far away from the closest training example as possible.

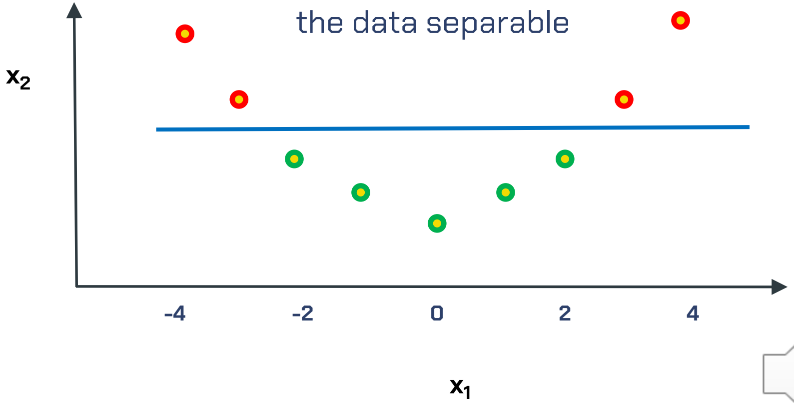
Non-linear Support Vector Machine

Many datasets are not linearly separable.

Adding additional features can make it linearly separable.

Kernel trick

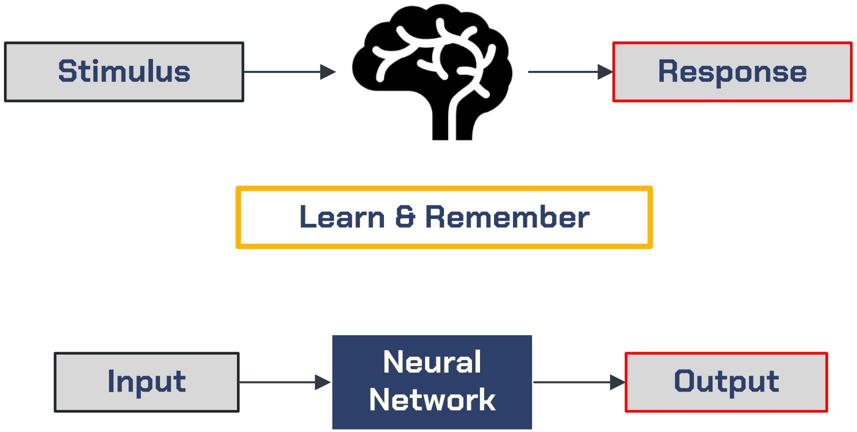




Topic 6

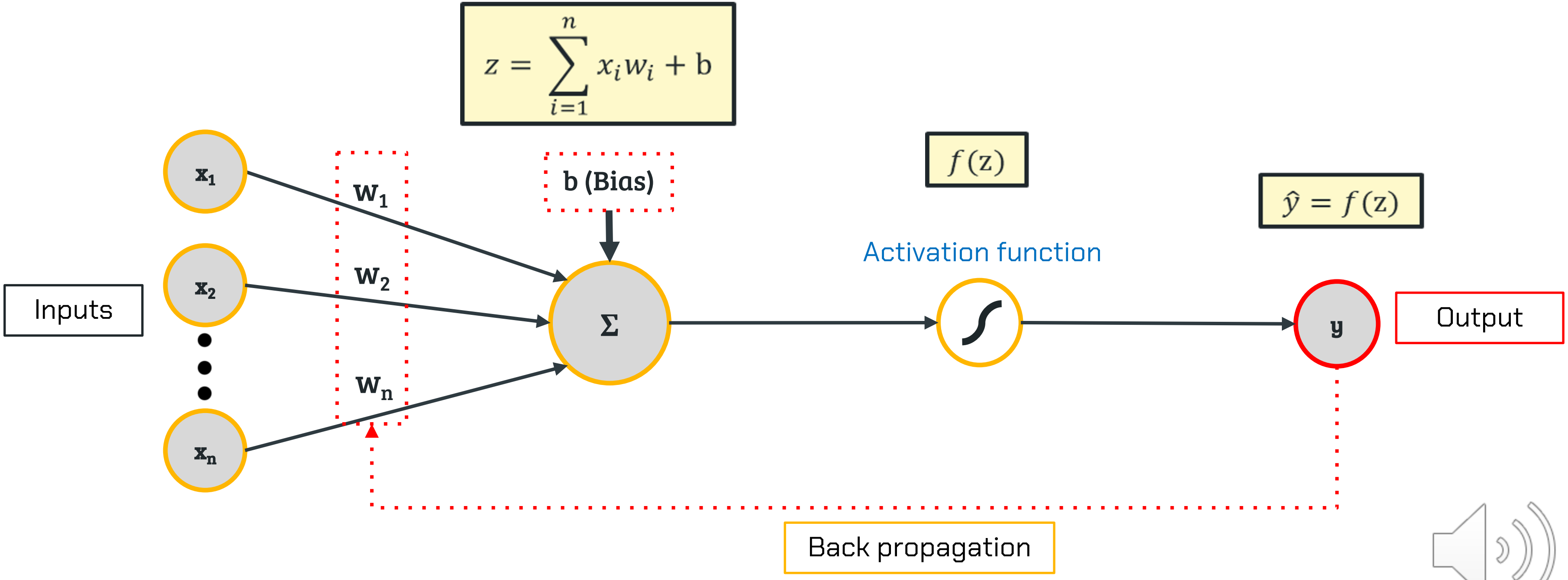
Artificial Neural Network

Artificial Neural Network



- **Neuron** - receive and transmit signals
- **Network** - Group of things connected to each other
- **Artificial Neural Network**
  - Group of connected neurons
  - Replicate the behavior of brain - training

Training a Neuron



Backpropagation

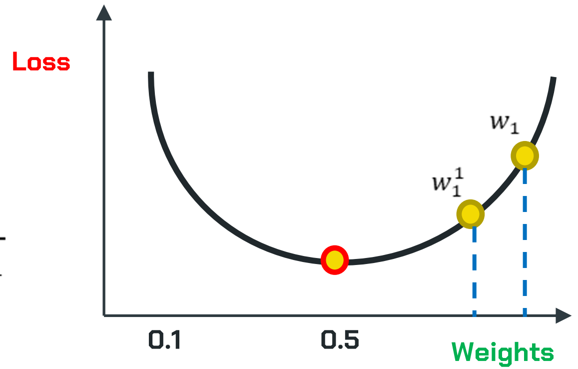
- **Step 1** - How off is the prediction?

  Errors

- **Step 2** - How much to correct

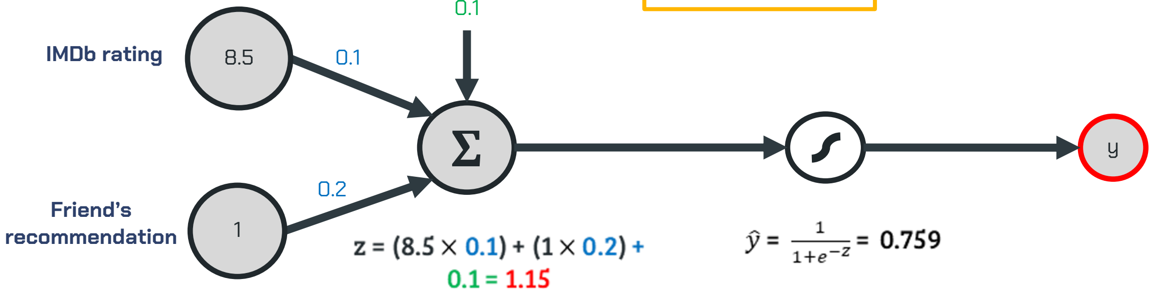
  Minimize "C"

$$new\ weight(w_1^1) = old\ weight(w_1) - learning\ rate(\alpha) \times \frac{\partial C}{\partial w_1}$$



To watch or not to watch, that is the question..

Forward pass



$\hat{y}$ = 0.759

y = 0

C = $(y - \hat{y})^2$ = 0.576

z = (8.5 × 0.1) + (1 × 0.2) + 0.1 = 1.15

$\hat{y} = \frac{1}{1+e^{-z}}$ = 0.759

$\frac{\partial C}{\partial \hat{y}} = -2\,(y - \hat{y})$ = - 2 (0 – 0.759) = 1.518

$\frac{\partial \hat{y}}{\partial z} = \hat{y}(1 - \hat{y})$ = 0.759 × 0.241 = 0.183

$\frac{\partial z}{\partial w_1} = x_1$ = 8.5 $\quad (z = w_1 x_1 + w_2\, x_2 + b)$

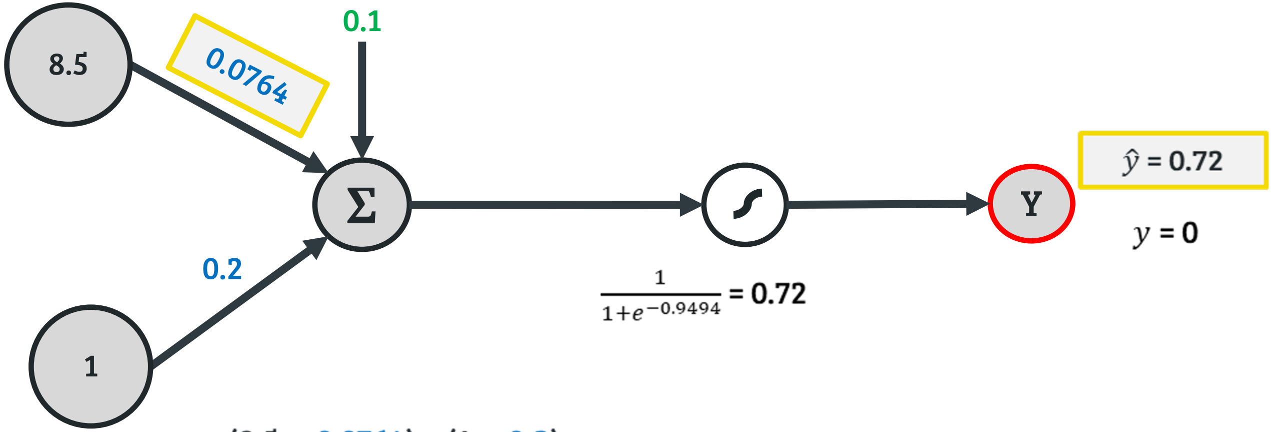
$$\boxed{\frac{\partial C}{\partial w_1} = 2.36}$$

$$\boxed{\frac{\partial C}{\partial w_i} = \frac{\partial C}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial z}{\partial w_i}}$$

Learning rate (α) = 0.01

New weight, $w_1^1 = w_1 - \alpha \times \frac{\partial C}{\partial w_1}$ = 0.0764

Updating weights → Learning



z = (8.5 × 0.0764) + (1 × 0.2) + 0.1 = 0.9494

Topic 7

Evaluating Classification performance

Classification - performance measures

Confusion matrix

Predicted values	Actual values	
	Positive (1)	Negative (0)
Positive (1)	100 (TP)	30 (FP)
Negative (0)	50 (FN)	120 (TN)

Sensitivity = TP/(TP+FN) → 100/(100+50) = **0.667**

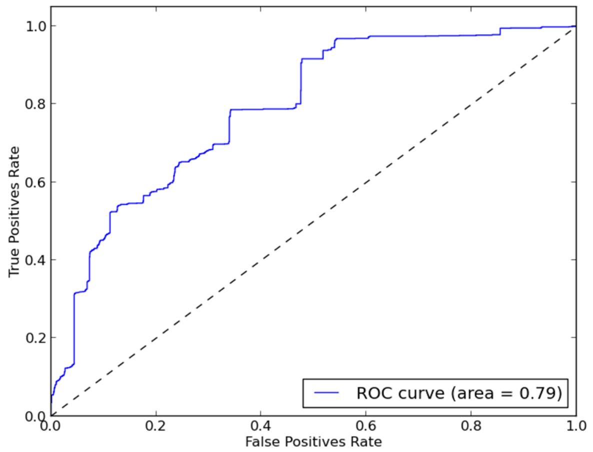
Specificity = TN/(TN+FP) → 120/(120+30) = **0.80**

Precision = TP/(TP+FP) → 100/(100+30) = **0.769**

Accuracy = TP+TN/(TP+FN+TN+FP) → (100+120)/300= **0.733**

Classification - performance measures

Receiver Operator Characteristic



- True positive rate (Sensitivity)
- False positive rate (1-Specificity)

Plot → Sensitivity vs. (1 – Specificity)

Best operating point

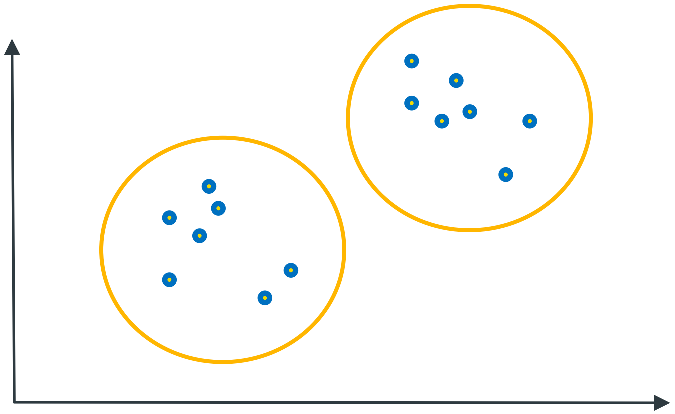
Topic 8

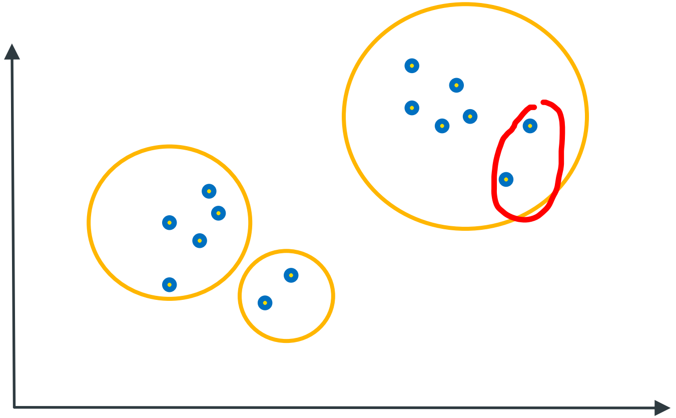
Clustering

Clustering

Unlabelled data

Example: Given a set of restaurants, group them into **good, average** and **bad**.





K-means clustering

Dataset, $D = \{\boldsymbol{x_1}, \ldots, \boldsymbol{x_m}\}$ Clusters, $\mathcal{C} = \{C_1, \ldots, C_k\}$

Objective: Minimize → $E = \sum_{i=1}^{k} \sum_{\boldsymbol{x} \in C_i} ||\boldsymbol{x} - \boldsymbol{\mu_i}||_2^2$ where, $\boldsymbol{\mu_i} = \frac{1}{|C_i|} \sum_{\boldsymbol{x} \in C_i} \boldsymbol{x}$

Steps:

Given a value of *k*, the *k*-means algorithm randomly assigns each observation to one of the *k* clusters.

After all observations have been assigned to a cluster, the resulting cluster centroids are calculated.

Using the cluster centroids, all observations are reassigned to the cluster with the closest centroid.

Topic 9

Discussion

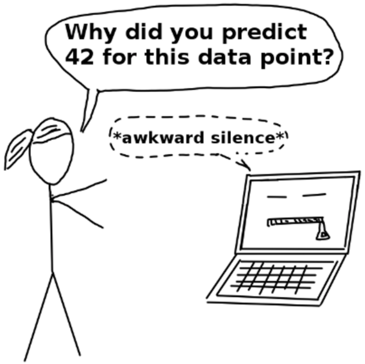
Advanced Machine Learning

Deep learning

Interpretable machine learning

Attention and transformers

Generative Adversarial Networks (GAN)



https://christophm.github.io/interpretable-ml-book/terminology.html

Challenges of Machine Learning

Bad Quantity of Training Data

- Bad samples
- Outliers, noise and errors
- Irrelevant features

Bad algorithms

- Overfitting & Underfitting
- Model Selection

Errors are Okay!

Any questions?