SUPERVISED AND UNSUPERVISED LEARNING

Fundamentals of Artificial Intelligence

Session 06

Pramod Sharma
pramod.sharma@prasami.com

2

Agenda

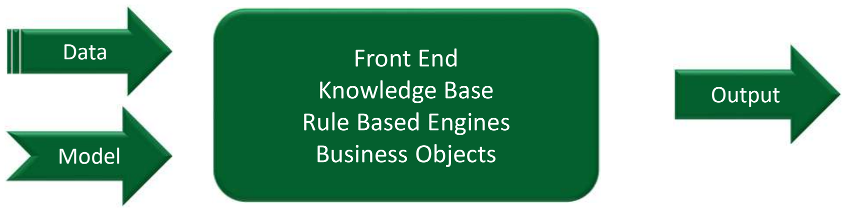
- Executing ML Projects
- Machine Learning
- Supervised and Unsupervised Learning
- Reinforcement Learning
- Descriptive Modelling

4/6/2024

pra-sâmi

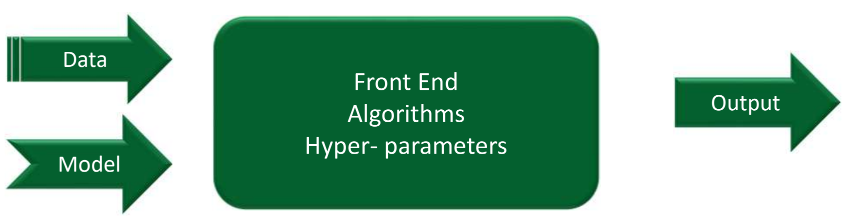
3

Traditional Programming



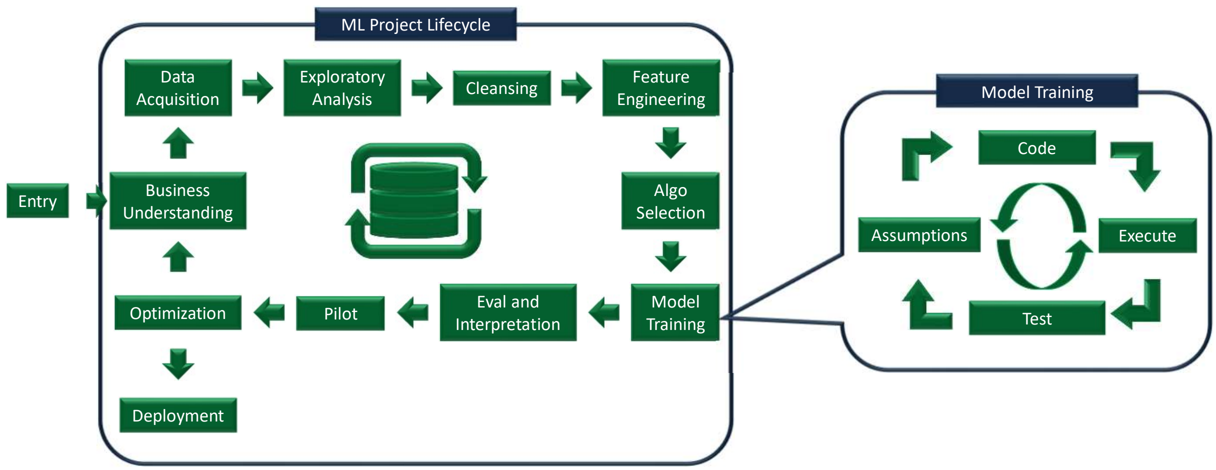
4

Machine Learning



5

Project Lifecycle

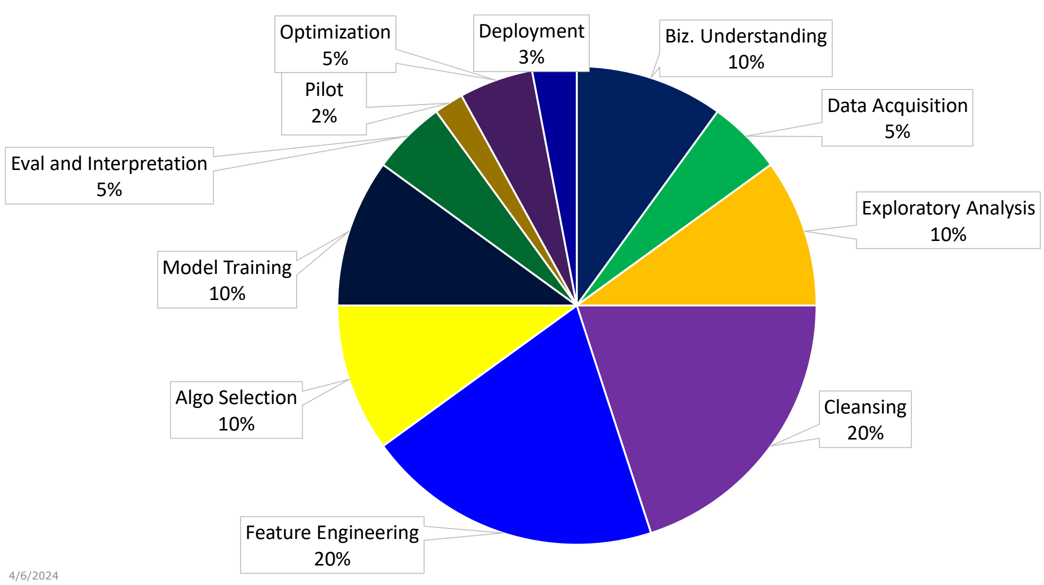


4/6/2024

pra-sâmi

6

Effort Distribution



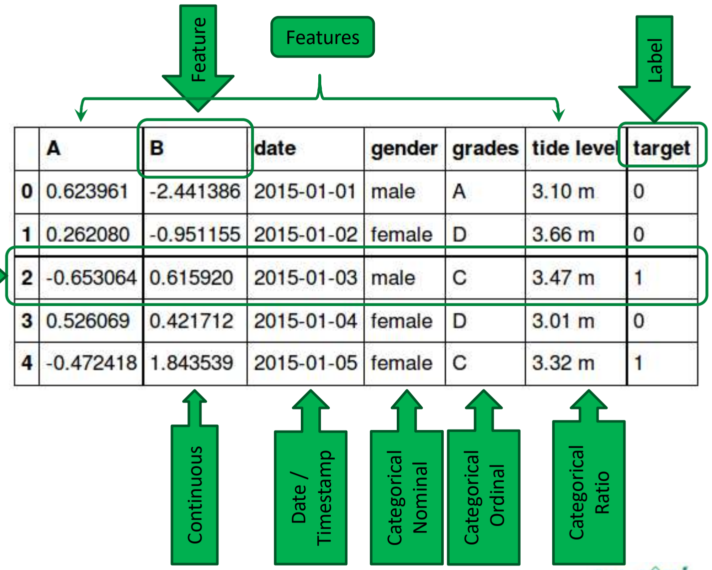
4/6/2024

pra-sâmi

7

Common Terms

- Model
  - Collection of parameters you are trying to fit
- Data : that CSV file which is given to you
- Instance
- Feature
  - Attributes of your data that will be used in prediction
- Labels
  - Value you are trying to predict with your model
- Categorical Data
  - **Nominal** : categorically discrete data
  - **Ordinal** : quantities that have a natural ordering
  - **Interval** : intervals between each value are equally split e. g. age groups, income levels
  - **Ratio:** interval data with a natural zero point. They are ordered units that have the same difference.
- Continuous Data
- Date / time stamp



	A	B	date	gender	grades	tide level	target
0	0.623961	-2.441386	2015-01-01	male	A	3.10 m	0
1	0.262080	-0.951155	2015-01-02	female	D	3.66 m	0
2	-0.653064	0.615920	2015-01-03	male	C	3.47 m	1
3	0.526069	0.421712	2015-01-04	female	D	3.01 m	0
4	-0.472418	1.843539	2015-01-05	female	C	3.32 m	1

8

Data is dirty!

- Key to success of any Data Science Project is Data.
- In ideal world, data is:
  - Representation of the population
  - Covers all relevant parameters
  - Equal number of representative samples are available
  - No missing or incorrect data
  - Clean representation of the desired process

- Real world data is dirty
  - Observations are biased
  - Merging data from different sources is challenge
  - Data Format change over time
  - Missing Data
  - Outliers, Errors

You will devote most of your time in creating right dataset to train the model!

9

Machine Learning and our Focus

- Humans learn from past experiences
- A computer does not have "experiences"
- A computer system learns from "data", which represent some "*past experiences*" of an application domain
- Our focus: learn a target function that can be used to predict the values of:
  - A discrete class attribute, e.g., car or truck; approve or not-approved; high-risk or low risk, etc.
  - An attribute having continuous value
- The task is commonly called: supervised learning, classification, unsupervised learning, regression or inductive learning

10

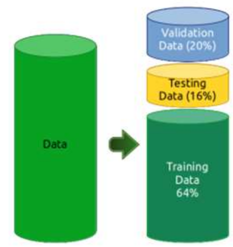
How am I Doing?

- Assumption: The distribution of training examples is identical to the distribution of test examples (including future unseen examples)
  - In practice, this assumption is often violated to certain degree
  - Strong violations will clearly result in poor classification accuracy
- Ever suffered crash and burn in Prod?
  - No one wants to see a failure in Prod
- How well a model will generalize to new cases?
  - Try it out on new cases
  - It's better to test and validate in Dev itself
- To achieve good accuracy on the test data, training examples must be sufficiently representative of the test data

Train – Test – Validate

- Split your data into Three sets

Data Set	Up to 10,000 records	From 10,000 to 1 Million	Above 1 Million
Training	64 %	80%	90%
Testing	16%	10%	5%
Validation	20%	10%	5%

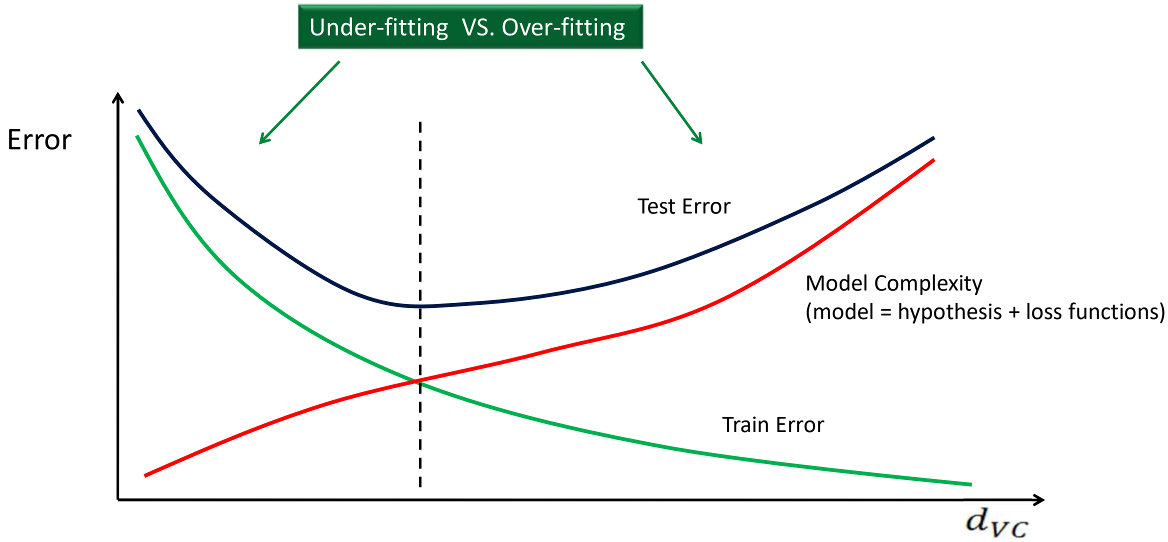


- Train your model using the training set
- Test it using the test set
- The error rate on new cases (Validation Set) is called the generalization error (or out-of-sample error)

Train – Test – Validate

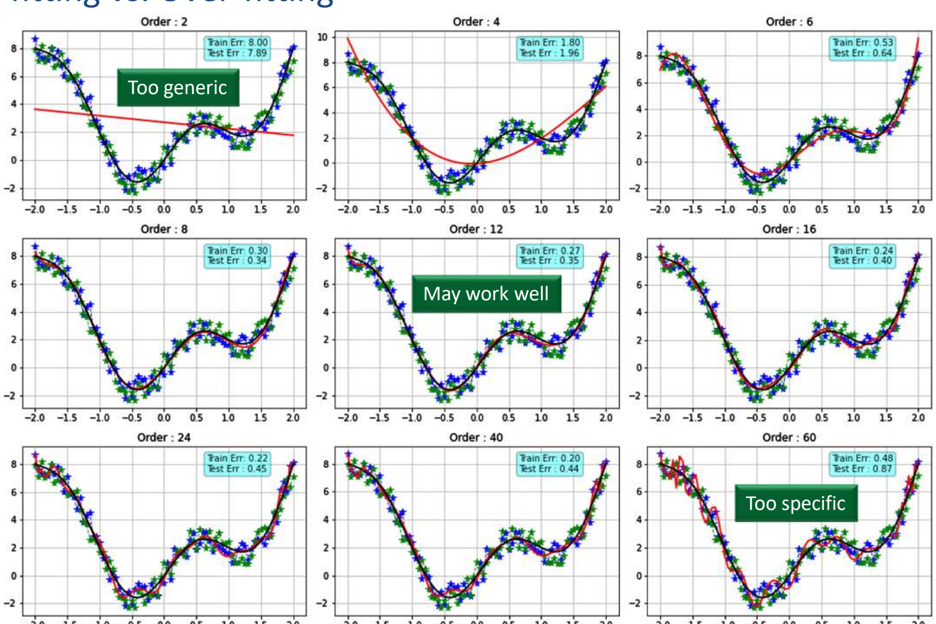
- First indication of how well model will perform
- Case 1: Training Error High, Test Error High
  - Model is under fitting
- Case 2: Training Error Low, Test Error High
  - Model is over-fitting the training data.
- In train-test cycle, hyper parameters are tuned keeping test data as target
  - Model is indirectly exposed to test set
- Results on previously "unexposed" Validation Set will still be a better indication of what to expect in Production

What are we seeking?



pra-sâmi

Under-fitting vs. Over-fitting



pra-sâmi

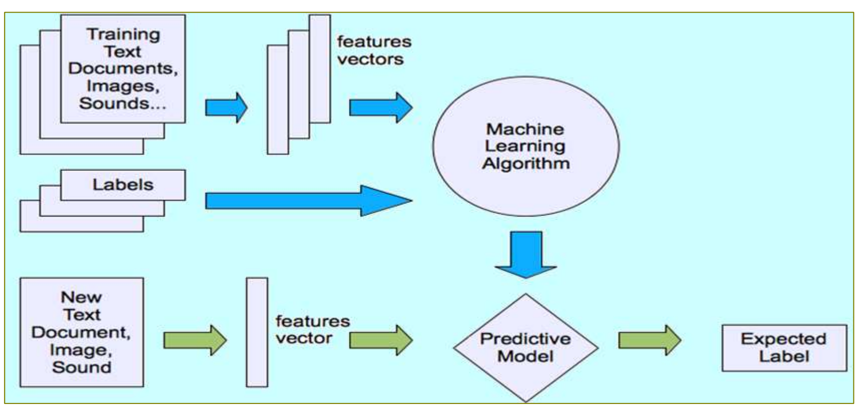
15

Cross Validation

- Worried about more than $\frac{1}{3}$ data being used in test-validate
- The training set is split into complementary subsets,
  - Each model is trained against a different combination of these subsets
  - Test against the remaining parts.
  - Happy with the model type and hyper-parameters
  - Final model is trained using these hyper-parameters on the full training set
  - The generalized error is measured on the validation set.

16

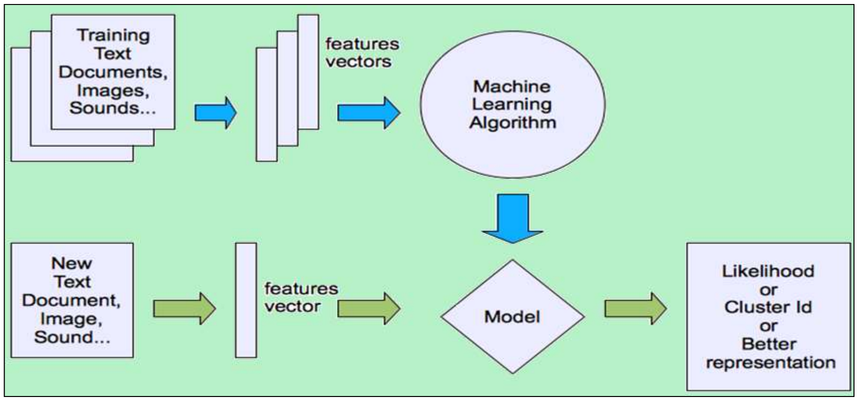
Machine Learning Structure



Supervised Learning

17

Machine Learning Structure



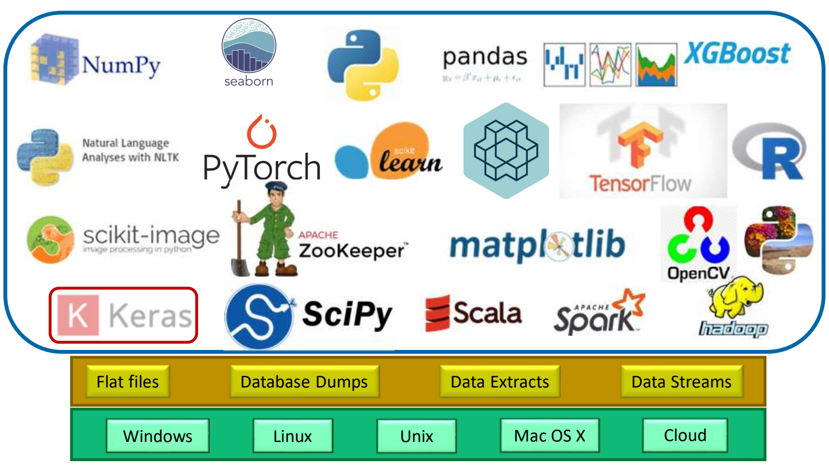
Unsupervised Learning

4/6/2024

pra-sâmi

18

AI and ML Technology Stack...



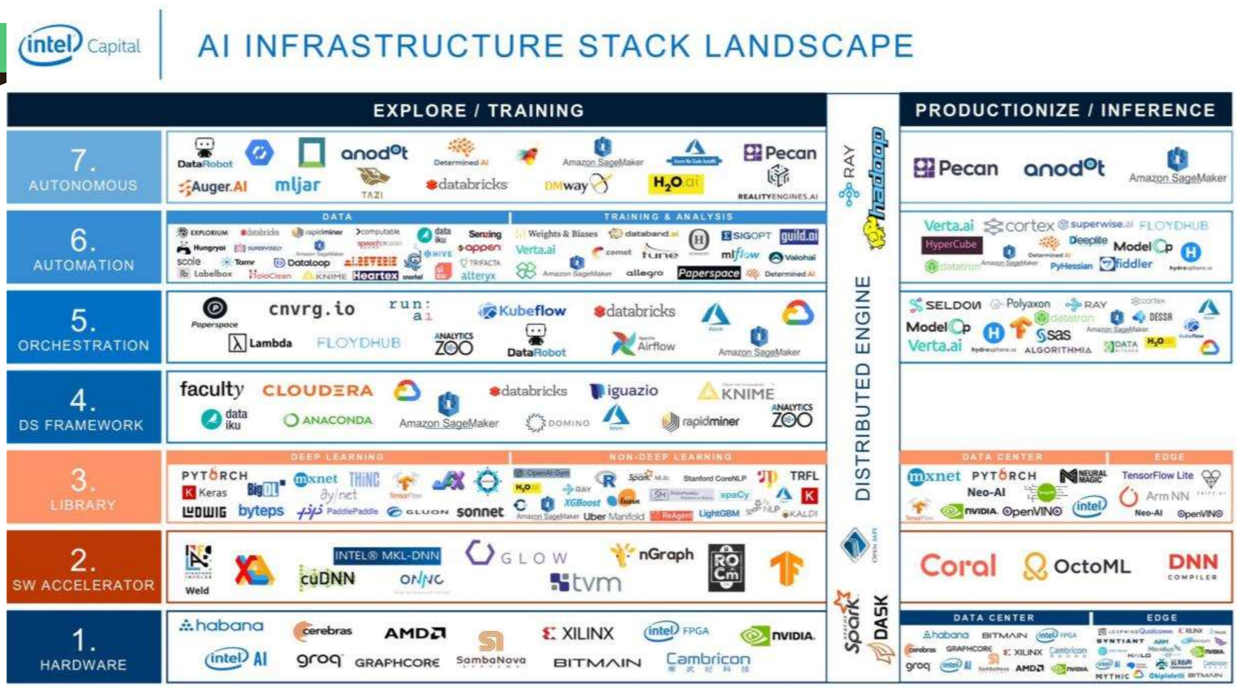
Additional

- IDE – Jupyter, Pycharm, Pydev…
- Solution design
- Implementation and execution
- Project management
- Change management

4/6/2024

pra-sâmi

19



20

Types of Artificial Intelligence

pra-sâmi

Definitions

- Weak AI
  - The idea that machines could act as if they were intelligent
- Strong AI
  - Machines can actually think consciously thinking
  - Not just simulating thinking
- Over time new level was introduced as "human-level AI" or "general AI"
  - Programs that can solve an arbitrarily wide variety of tasks, including novel ones,
  - and perform as well as a human

pra-sâmi

Classification of AI

Weak AI	Strong AI
Also known as narrow AI	Constantly learning
Focuses on performing a specific task - Answering questions based on user input - Playing chess - Can perform only one type of task	Eventually teaching itself to solve for new problems
Relies on human interference to - Define the parameters of its learning algorithms - Provide the relevant training data to ensure accuracy	While human input accelerates the growth phase of Strong AI, it is not required
	Over time, it develops a human-like consciousness instead of simulating it
Examples : Self-driving cars , games and virtual assistants, like Siri	Chess would fall in strong AI if a novice system can learn from opponents and out play them

Weak AI is learning by example and Strong AI is learning from experience
Most of current implementations fall under weak AI or a marginal improvement thereof

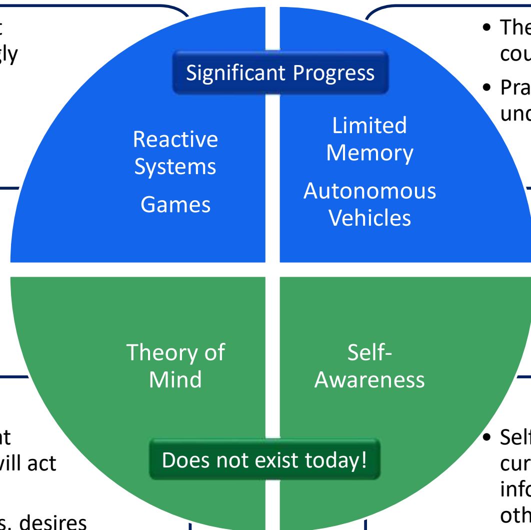
pra-sâmi

24

Another way to classify

- Games can identify current situation and act accordingly

- They use past data to decide course of action
- Practically, most use cases fall under this category



- Develop understanding that different people/systems will act differently
- They have their own beliefs, desires and intentions

- Self-aware machine understands current state and can use the information not only to infer what others are feeling but also how it should act.

25

Machine Learning

26

Why "Learn" ?

- Machine learning is programming computers to optimize a performance criterion using example data or past experience.
- There is no need to "learn" to calculate payroll
- Learning is used when:
  - Human expertise does not exist (navigating on Mars),
  - Humans are unable to explain their expertise (speech recognition)
  - Solution changes in time (routing on a computer network)
  - Solution needs to be adapted to particular cases (user biometrics)

4/6/2024

pra-sâmi

27

Machine Learning

- "Learning is any process by which a system improves performance from experience."
  - Herbert Simon
- Definition by Tom Mitchell (1998):
  - Machine Learning is the study of algorithms that
    - improve their performance P
    - at some task T
    - with experience E
  - A well-defined learning task is given by < P, T, E>
- Role of Statistics: Inference from a sample
- Role of Computer science: Efficient algorithms to
  - Solve the optimization problem
  - Representing and evaluating the model for inference

4/6/2024

pra-sâmi

28
ML Algorithms in Short

- Tens of thousands of machine learning algorithms
  - Hundreds new every year
- Every ML algorithm has three components:
  - Representation
  - Optimization
  - Evaluation

pra-sâmi

29
Various Function Representations

- Numerical functions
  - Linear regression
  - Neural networks
  - Support vector machines
- Symbolic functions
  - Decision trees
  - Rules in propositional logic
  - Rules in first-order predicate logic
- Instance-based functions
  - Nearest-neighbor
  - Case-based
- Probabilistic Graphical Models
  - Naïve Bayes
  - Bayesian networks
  - Hidden-Markov Models (HMMs)
  - Probabilistic Context Free Grammars (PCFGs)
  - Markov networks

pra-sâmi

30

Various Search / Optimization Algorithms

- Gradient descent
  - Perceptron
  - Back-propagation
- Dynamic Programming
  - Hidden Markov Model (HMM) Learning
  - Probabilistic Context Free Grammar (PCFG) Learning
- Divide and Conquer
  - Decision tree induction
  - Rule learning
- Evolutionary Computation
  - Genetic Algorithms (GAs)
  - Genetic Programming (GP)
  - Neuro-evolution

31

Evaluation

- Accuracy
- Precision and recall
- Squared error
- Likelihood
- Posterior probability
- Cost / Utility
- Margin
- Entropy
- K-L divergence
- and many more…

32

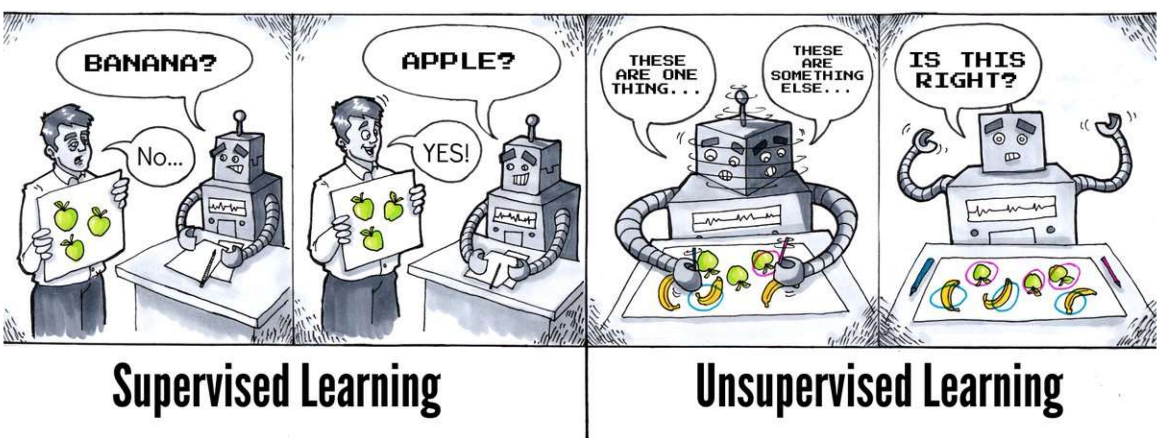
Applications

- Association
  - Basket analysis: P (*Y* | *X*) probability that somebody who buys *X* also buys *Y* where *X* and *Y* are products/services.
    - Example: *P* (chips | beer) = 0.7
- Supervised Learning
  - Given: training data + desired outputs (labels) ➔ fit a hypothesis to it
- Types of supervised learning
  - Classification
  - Regression
- Unsupervised Learning
  - Given: training data (without desired outputs); Try and determining structure in the data
  - Clustering algorithm groups data together based on data features
- Semi-supervised learning
  - Given: training data + a few desired outputs
- Reinforcement learning
  - Rewards from sequence of actions

pra-sâmi

33

Supervised vs. Unsupervised



pra-sâmi

34

Supervised Classification

âmi

35

Unsupervised Classification

4/6/2024 pra-sâmi

37

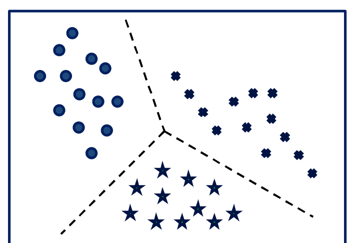
Supervised Learning

pra-sâmi

38

Supervised Learning

- Map input data to known targets
  - Optical character recognition, speech recognition, image classification, and language translation
- Mostly consists of classification and regression,
  - Classification (discrete labels), Regression (real values)
- Other Variants
  - Sequence generation—Given a picture, predict a caption describing it
  - Syntax tree prediction—Given a sentence, predict its decomposition into a syntax tree
  - Object detection—detect objects in a picture
  - Image segmentation—Given a picture, draw a pixel-level mask on a specific object
- Minimize output error or maximize probabilistic terms



Supervised learning

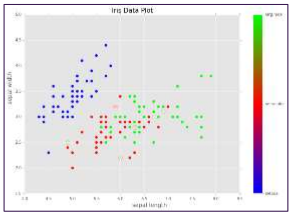
pra-sâmi

Supervised Learning Categories and Techniques

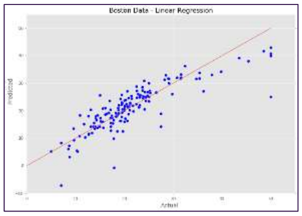
- Linear classifier (numerical functions)
- Parametric (Probabilistic functions)
  - Naïve Bayes, Gaussian discriminant analysis (GDA), Hidden Markov models (HMM), Probabilistic graphical models
- Non-parametric (Instance-based functions)
  - K-nearest neighbors, Kernel regression, Kernel density estimation, Local regression
- Non-metric (Symbolic functions)
  - Classification and regression tree (CART), decision tree
- Aggregation
  - Bagging (bootstrap + aggregation), Adaboost, Random forest, XGBOOST

Classification and Regression

- Classifications
  - Naïve Bayes
  - Support Vector Machines
  - Decision trees
  - Random Forest
  - K-Nearest Neighbor
  - Neural Networks
- Classification predicts labels (class)



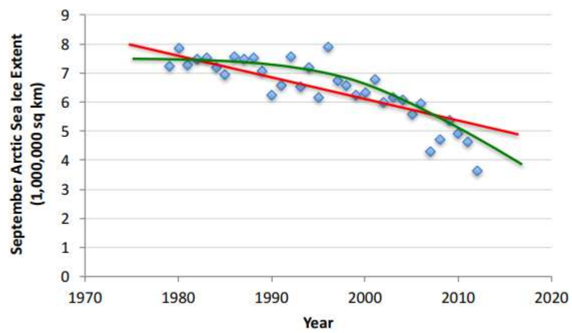
- Regression
  - Linear Regression
  - Random Forest
  - Gradient Boosting
  - Neural Networks
  - Others
- Regression predicts continuous value



41

Supervised Learning: Regression

- Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Learn a function f(x) to predict y given x
  - y is real-valued $\Rightarrow$ regression



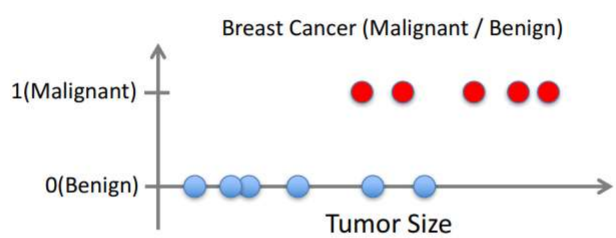
4/6/2024

pra-sâmi

42

Supervised Learning: Classification

- Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Learn a function f(x) to predict y given x
  - y is categorical $\Rightarrow$ classification



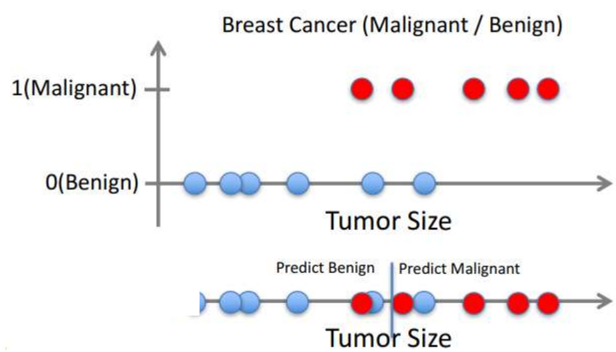
4/6/2024

pra-sâmi

43

Supervised Learning: Classification

- Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
  - y is categorical $\Rightarrow$ classification



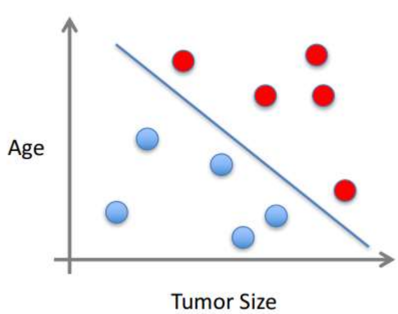
4/6/2024

pra-sâmi

44

Supervised Learning

- x can be multi-dimensional
  - Each dimension corresponds to an attribute



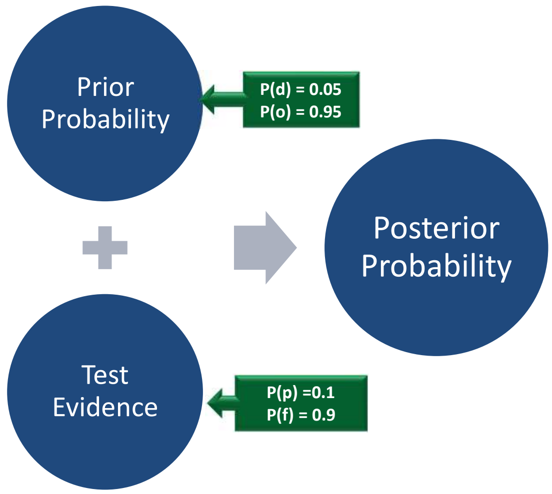
- Clump Thickness
- Uniformity of Cell Size
- Uniformity of Cell Shape
- and other observable/ measurable parameters

4/6/2024

pra-sâmi

45

Naïve Bayes



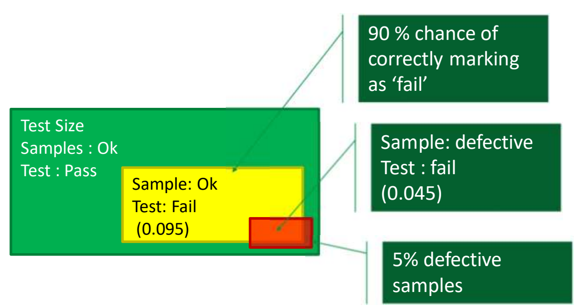
- In a factory producing mechanical parts, 5 % pieces are defective
  - P(ok) = 0.95; P(defect) = 0.05
- There is a test lab in the factory. The test is 90% accurate
  - P(c) = 0.9; P(inc) = 0.1
- There is 90% chance, that it catches failed sample – Precision
  - 10 % chance It can show failed even if the sample is ok
- There is 90% chance, it shows pass if the sample is ok – Recall
  - 10 % chance it shows pass even when the sample is not ok

4/6/2024 pra-sâmi

46

Let's Take an Example

- Suppose the test result show "fail", what is the probability that it is actually defective
- Joint p(X) =P(d|f) = 0.05*0.9 = 0.045
- Joint p(Y) = P(0|f) = 0.95* 0.1 = 0.095
  - Remember there is 90 % chance of fail for defective pieces
- Normalize the probabilities
  - Sum is not 1, the values are normalized to 1.
  - Divide both by (0.095 + 0.045) = 0.14
  - P(X) = 0.32
  - P(Y) = 0.68



Exercise:
Try with test being 99 % accurate

4/6/2024 pra-sâmi

47

Advantages and Disadvantages

Advantages	Disadvantages
❑ Fast and space efficient ❖ Look up all the probabilities with a single scan of the database ❑ NOT sensitive to irrelevant features... ❑ Handles real and discrete data well ❑ Handles streaming data well	❑ Assumes independence of features

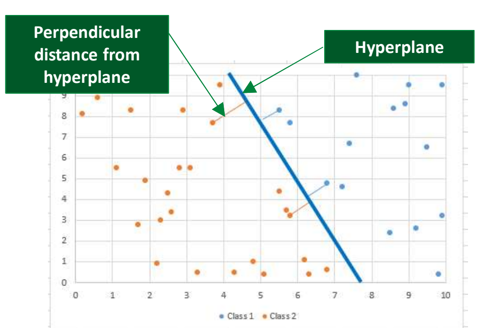
4/6/2024

pra-sâmi

48

Support Vector Machines

- ❑ Assumes that these classes are linearly separable
- ❑ Support Vector Machines (SVM) finds a hyperplane to be as far as possible from the closest members of both classes.

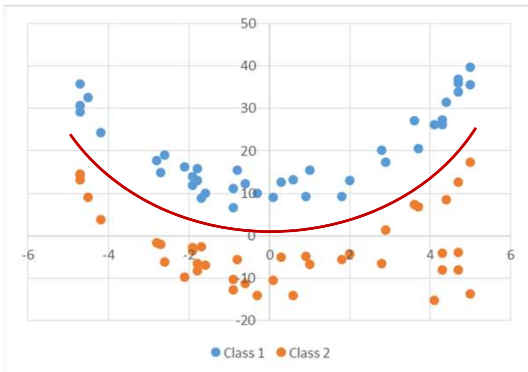


4/6/2024

pra-sâmi

49

Support Vector Machines



- What if two classes are not linearly separable?
- We can still find a n^{th} order plane which can separate the two classes
  - Use of Kernel

pra-sâmi

50

Advantages and Disadvantages

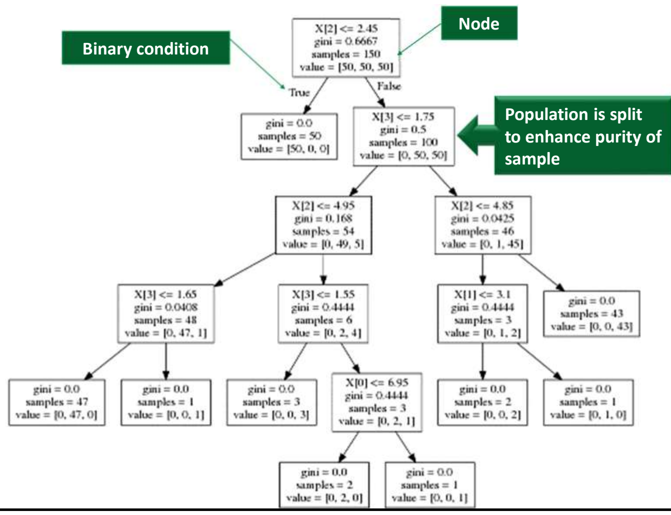
Advantages	Disadvantages
Effective in high dimensional spaces. Still effective in cases where number of dimensions is greater than the number of samples. Uses a subset of training points in the decision function (called support vectors), so it is also memory efficient Versatile: different Kernel functions can be specified for the decision function. Common kernels are provided, but it is also possible to specify custom kernels.	If the number of features is much greater than the number of samples, the method is likely to give poor performances. SVMs do not directly provide probability estimates, these are calculated using an expensive five-fold cross-validation

pra-sâmi

51

Decision Trees

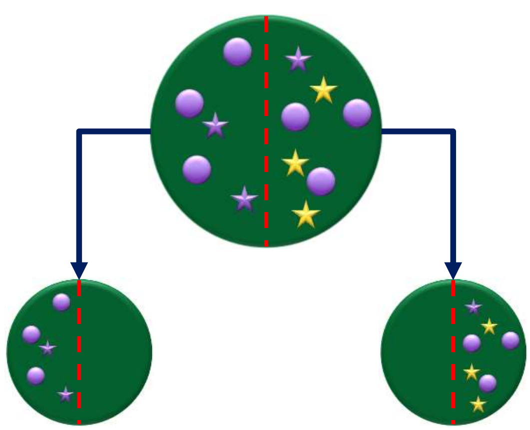
- Decision tree is most intuitive and most popular.
- Decision Tree is like “20 questions” game… based on yes/no responses, one can find the answer.



pra-sâmi

52

Split of sample



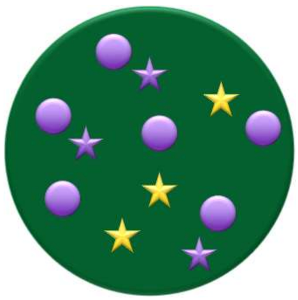
Decision tress involve greedy, recursive partitioning.

pra-sâmi

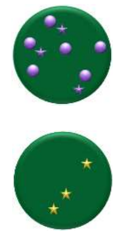
53

Entropy

- Entropy is measure of impurity of a collection.
  - All the samples are of same type
    $\Rightarrow$ low impurity
    $\Rightarrow$ entropy = 0
  - Evenly spread sample
    $\Rightarrow$ high impurity
    $\Rightarrow$ entropy = 1
- Circles = 6
- Stars = 6



- Entropy = - $\Sigma\, p_i \log_2\, p_i$
  - Example: $= -0.5.\log_2 0.5\text{-}0.5.\log_2 0.5$
    $= 0.5 + 0.5$
    $= 1.0$
  - If we split them based on color
    $\Rightarrow$ 6 circles and 3 stars
    pCircles = 6/9
    pStars = 3/9



  - Entropy = $-\frac{6}{9}.\log_2\frac{6}{9} + \text{-}\frac{3}{9}.\log_2\frac{3}{9}$
    $= 0.918$
  - Information Gain $=$ *entropy of parent* $-$ [*weighted average*]*entropy of children*
  - Information Gain $= 1\ - \frac{9}{12} * 0.918 - \frac{3}{12} * 0.00$
    $= 0.3115$

pra-sâmi

54

Advantages and Disadvantages

Advantages	Disadvantages
Easy to understand. – Human inspection (physical) is possible	Extremely sensitive to change in data : can result in a drastically different tree
Can give valuable results even with little data	Prone to overfitting : Can be negated by validation methods and pruning.
Scalable : addition of new possible scenarios	Difficulty in dealing with multicollinearity: when two variables both explain the same thing, a decision tree will greedily choose the best one, other is ignored.
Help determine worst, best and expected values for different scenarios	Lack of a principled probabilistic framework.
Can be combined with other decision techniques	Poor resolution if data has complex relationships among the features
	Practically Limited to Classification – Poor resolution with continuous variables

pra-sâmi

55

Wisdom of Crowds

- Ensemble methods
  - Combine multiple classifiers to make "better " classifier
  - Predictions are averaged out – reducing error
  - Can use weighted combinations
  - Can use even different classifiers

- Two types of ensemble:
  - Bagging
    - Weak learners are trained in parallel
    - Used on weak learners that exhibit high variance and low bias
    - Avoids overfitting
    - Being leveraged for loan approval processes and statistical genomics
  - Boosting
    - They learn sequentially→ a series of models are constructed and with each new model iteration, the weights of the misclassified data in the previous model are increased.
    - Redistribution of weights helps to focus on the parameters which will lead to performance improvement
    - Leveraged when low variance and high bias is observed
    - Prone to overfitting
    - Boosting has been used more within image recognition apps and search engines

4/6/2024

pra-sâmi

56

Random Forest

- Random Forests are ensembles of decision trees

- Random Record Selection
  - Each tree is built from a separate random sample (bootstrap sample)
  - Create hundreds of trees each - gain diversity from examining different examples
  - Over fits the data extensively

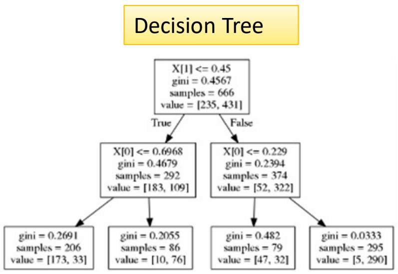
- Random Feature Selection
  - A random subset of variables
  - Number of variables to consider is √(features) - its configurable though
  - Forces the trees to find alternative ways to predict the target

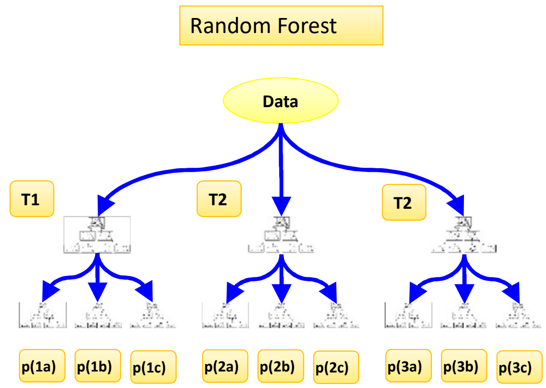
4/6/2024

pra-sâmi

57

Random Forest





- Split at point of maximum info gain from random sample
- Prediction = average probabilities

4/6/2024

pra-sâmi

58

Advantages and Disadvantages

Advantages

- Most accurate learning algorithms available.
  - For many data sets, it produces a highly accurate classifier.
- Efficient even on large datasets.
- Can handle thousands of features.
- Estimates feature importance as well.
- Unbiased estimate of the generalization error.
- Effective method for estimating missing data
  - Maintains accuracy even if significant chunk of data is missing.
- Balances out the error in class population unbalanced data sets.

Disadvantages

- Tend to over-fit for some datasets
  - especially noisy classification/regression tasks.
- If dataset has significantly higher instances of missing some feature in categorical targets, random forests are biased against these features.
  - Feature importance may not be reliable

4/6/2024

pra-sâmi

59

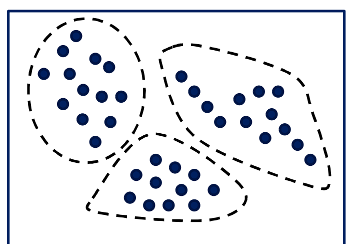
Unsupervised Learning

pra-sâmi

60

Unsupervised Learning

- Transformations of the input data
- Helpful in data visualization, data compression, or data de-noising,
- Explore correlations
- Better understanding a dataset before attempting to solve
- Minimum quantization error, Minimum distance, MAP, MLE(maximum likelihood estimation)



Unsupervised learning

pra-sâmi

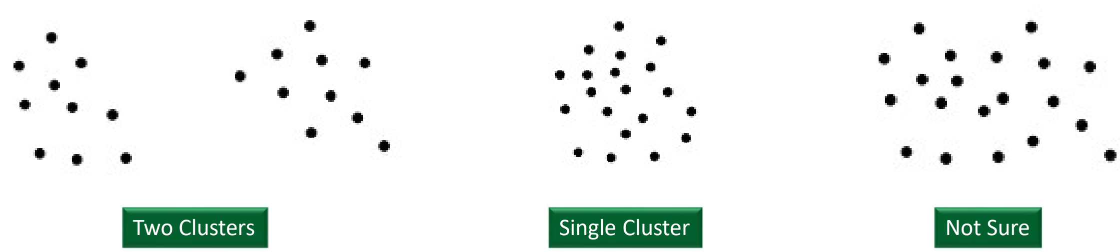
61

Unsupervised Learning

- Clustering
  - K-means clustering
  - Spectral clustering
- Density Estimation
  - Gaussian mixture model (GMM)
  - Graphical models
- Dimensionality reduction
  - Principal component analysis (PCA)
  - Factor analysis
- Finding association (in features)
  - Correlations

62

Clusters

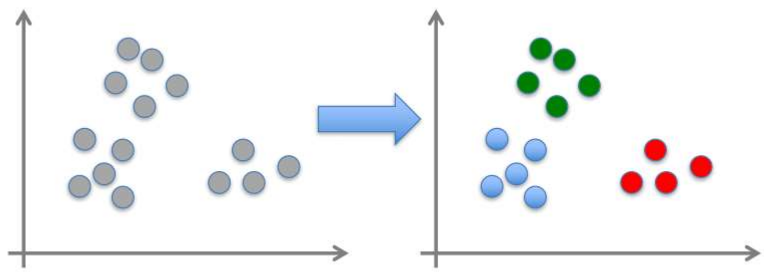


Unsupervised learning uses procedures that attempt to find natural partitions of patterns.

63

Unsupervised Learning

- Given $x_1, x_2, \ldots, x_n$ (without labels)
- Output hidden structure behind the x's
  - e.g., clustering



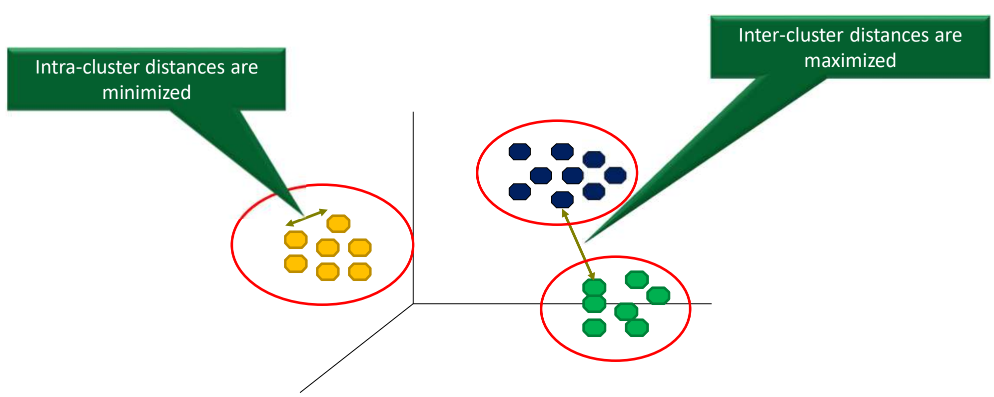
4/6/2024

pra-sâmi

64

What is Cluster Analysis?

- Finding groups of objects in data such that the objects in a group will be similar (or related) to one another and different from (or unrelated to) the objects in other groups



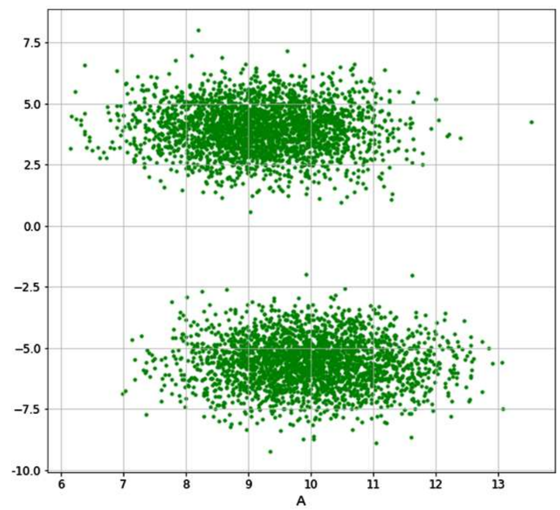
4/6/2024

pra-sâmi

65

K-means Algorithm

- Randomly allocate two points as the cluster centroids
- There can be as many cluster centroids as needed (K cluster centroids, in fact)
- Consider data as shown

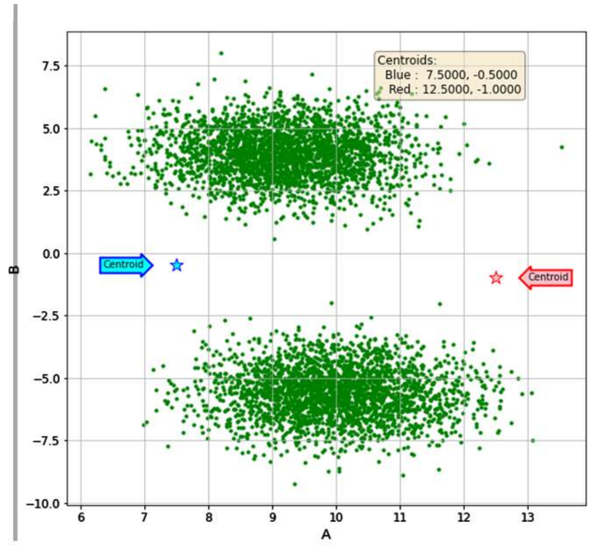


pra-sâmi

66

Iterations

- Our data has two groups
- Consider centroids for two groups arbitrarily
- Calculate Euclidean distance of each point from the two centroid.
- Allocate the point to centroid with minimum distance

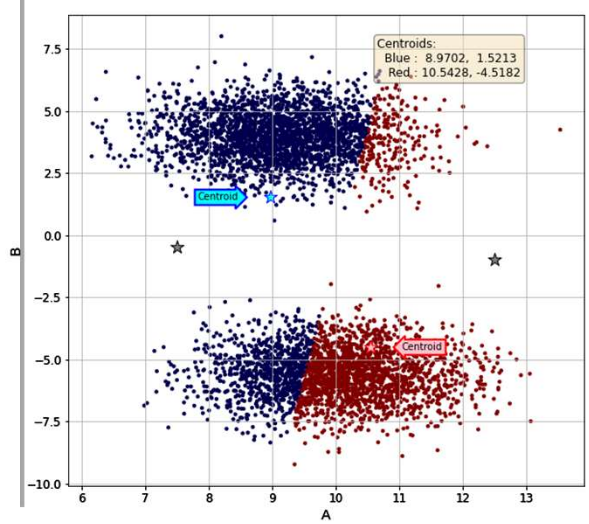


pra-sâmi

67

Iterations

- After allotment, calculate coordinates of centroid of each of the group
- Move the centroid to new location.



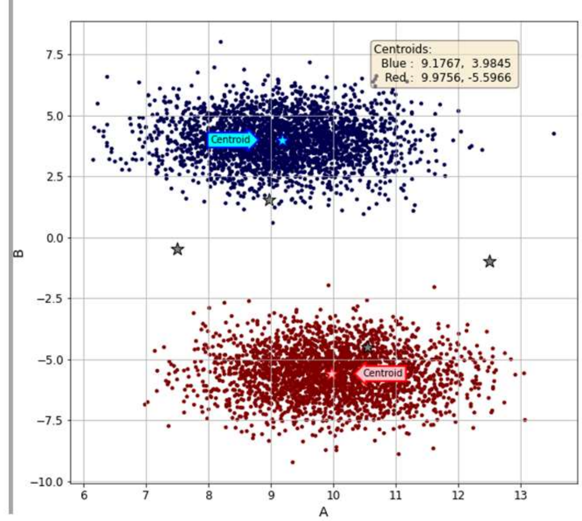
4/6/2024

pra-sâmi

68

Iterations

- Once again calculate Euclidean distance of each point from the two centroid
- Allot the points to closest centroid
- And we have identified the clusters



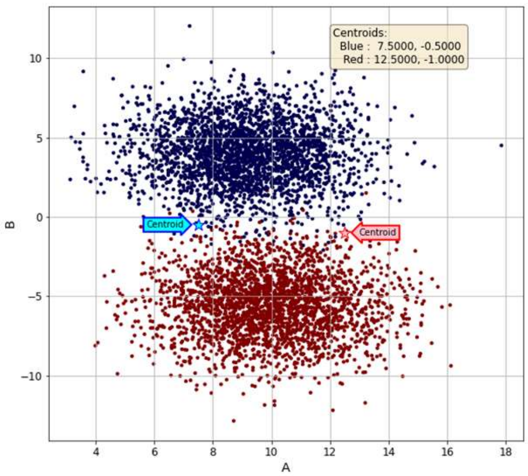
4/6/2024

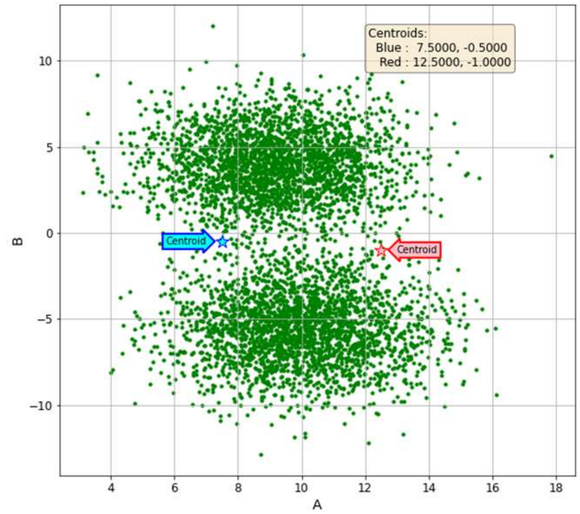
pra-sâmi

69

Iterations

- Another example with not so obvious separation



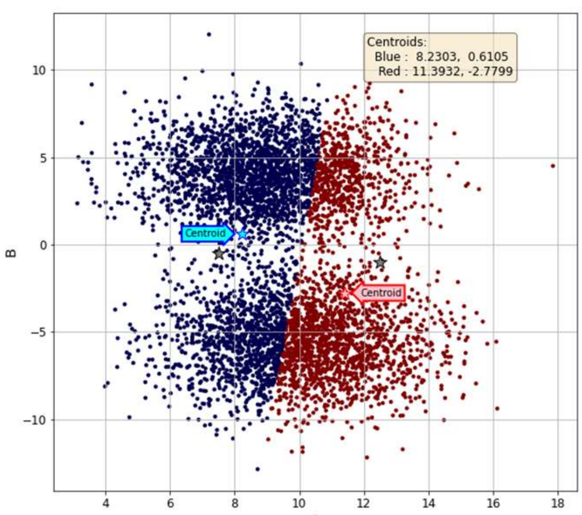


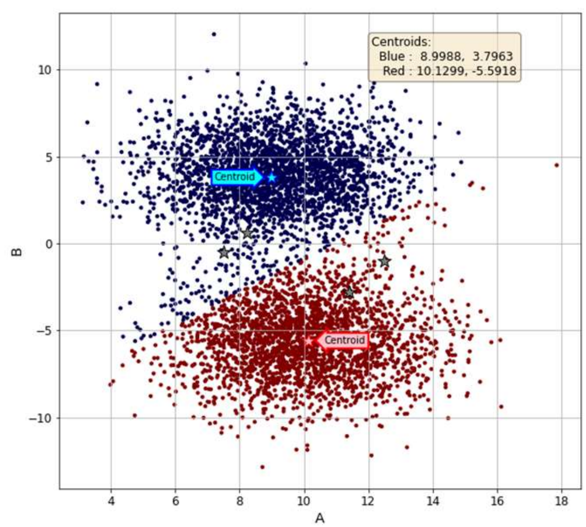
4/6/2024

pra-sâmi

70

Iteration 1 and 2



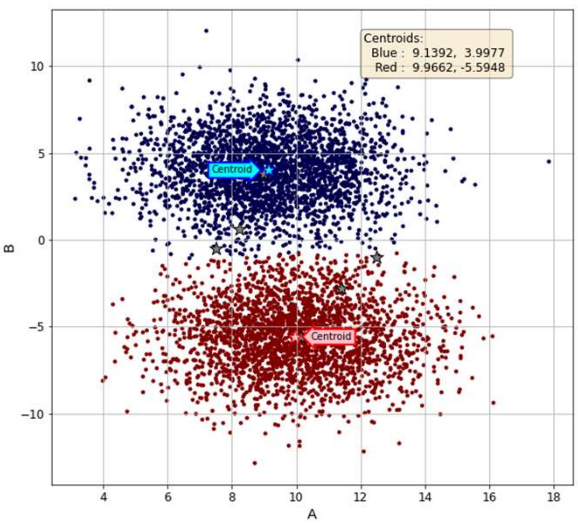


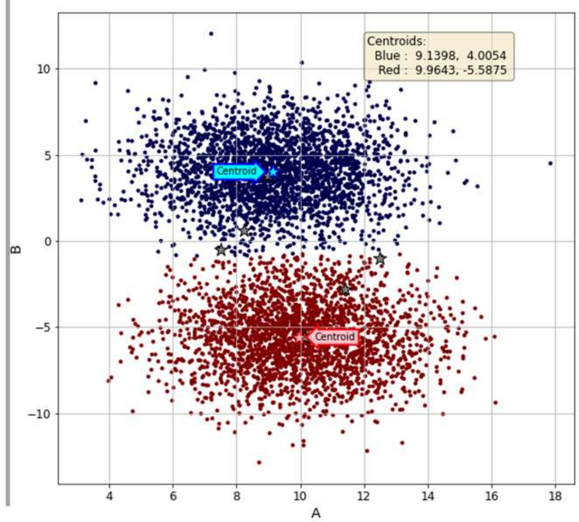
4/6/2024

pra-sâmi

71

Iteration 3 and 4



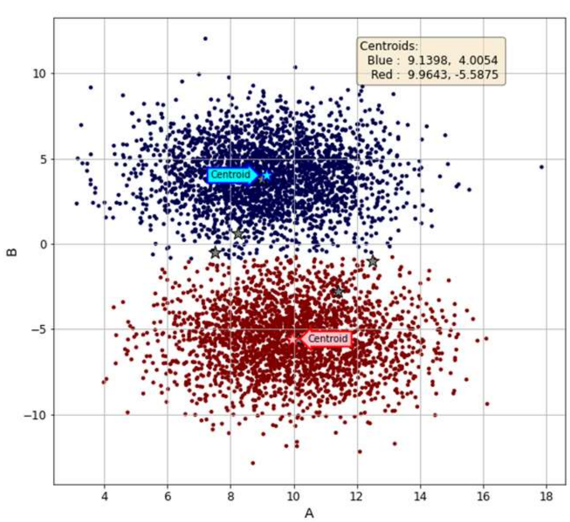


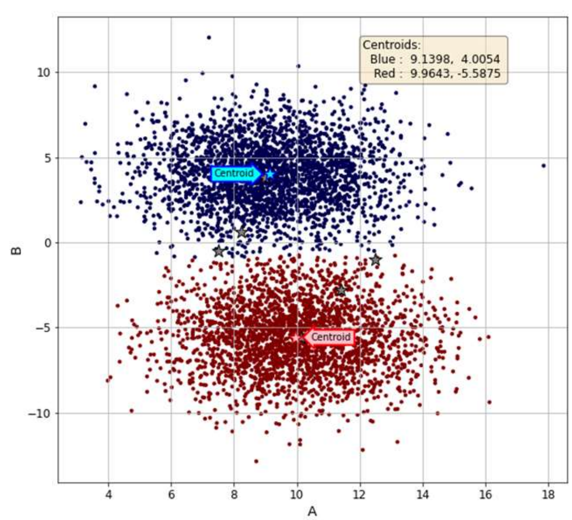
4/6/2024

pra-sâmi

72

Iteration s 4 and 5





4/6/2024

pra-sâmi

Clustering Distance Measures

Dissimilarity or Distance matrix

- Classification of observations into groups requires
  - Some methods for computing the **distance** or
  - The (dis)**similarity** between each pair of observations
- Distance measures is a critical step in clustering.
- Wide varieties methods available
- Method of calculations of the similarity of two elements (x, y) influences the shape of the clusters

75

Methods for Distance Measures

- The classical methods for distance measures :
  - Euclidean Distances
  - Manhattan Distances
  - Minkowski Distances
  - Hamming Distance
- Others:
  - Pearson correlation distance
  - Eisen cosine correlation distance (Eisen et al., 1998)
  - Spearman correlation distance
  - Kendall correlation distance
- Pearson correlation analysis is the most commonly used method.
  - Also known as a parametric correlation which depends on the distribution of the data
- Kendall and Spearman correlations are non-parametric
  - Used to perform rank-based correlation analysis

4/6/2024

pra-sâmi

76

Euclidean distance

- Most common method of distance measurement
- The distance between two real-valued vectors : x, y ∈ $\mathbb{R}$
- Used for calculating the distance between numerical values – Floats or integers.

$$d_{euc}(\, x, y\,) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

- If columns have values with differing scales:
  - normalize or standardize the numerical values to prevent one column prevail over other

4/6/2024

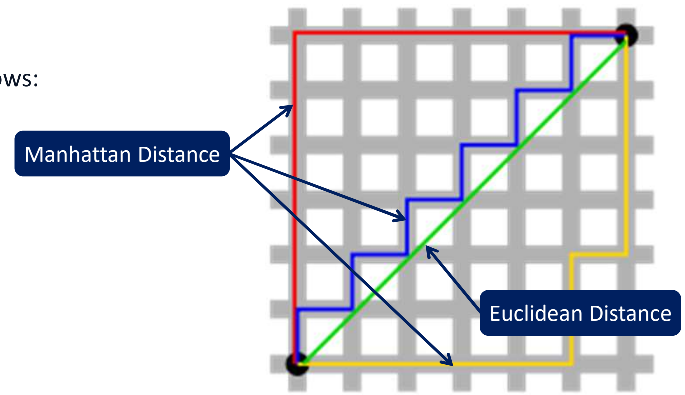
pra-sâmi

77

Manhattan Distance

- Names allude to the grid layout of most streets on the island of Manhattan
- Causes the shortest path a car could take between two intersections to have length equal to the intersections' distance
- Manhattan Distance is calculated as follows:

$$d_m(\text{x, y}) = \sum_{i=1}^{n} |x_i - y_i|$$



pra-sâmi

78

Minkowski Distances

- Both Euclidean and Manhattan Distances are special case of Minkowski Distance
- The formula for Minkowski Distance is given as:

$$\mathrm{d_m}(\text{x, y}) = (\sum_{i=1}^{n} (x_i - y_i)^p)^{1/p}$$

  - For p = 1 and p = 2, it becomes Manhattan and Euclidean Distances

pra-sâmi

79

Hamming Distance

- Measures the similarity between two strings of the same length.
- The number of positions at which the corresponding characters are different

P	R	A	M	O	D	–	
P	R	O	M	O	T	E	
0	0	1	0	0	1	1	3
							3/7 = 0.43

4/6/2024

pra-sâmi

80

Pearson correlation distance

- A correlation-based distances
- Widely used for gene expression data analyses
- Correlation-based distance is defined by subtracting the correlation coefficient from 1
- Pearson correlation measures the degree of a linear relationship between two profiles
- It is calculated as follows:

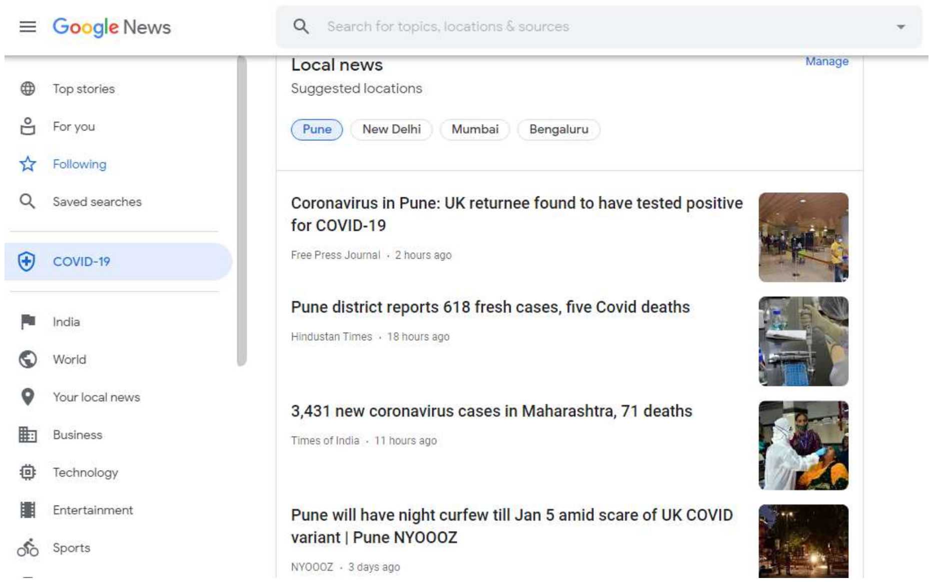
$$d_{cor}(x, y) = 1 - \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 . \sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

4/6/2024

pra-sâmi

81

News Clustering



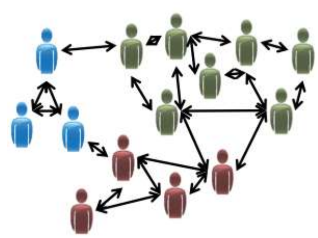
4/6/2024

pra-sâmi

83

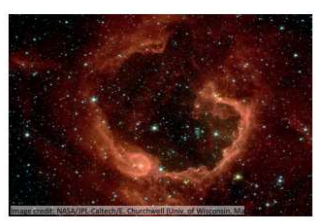
Unsupervised Learning

Organize computing clusters



Social network analysis

Market segmentation



Astronomical data analysis

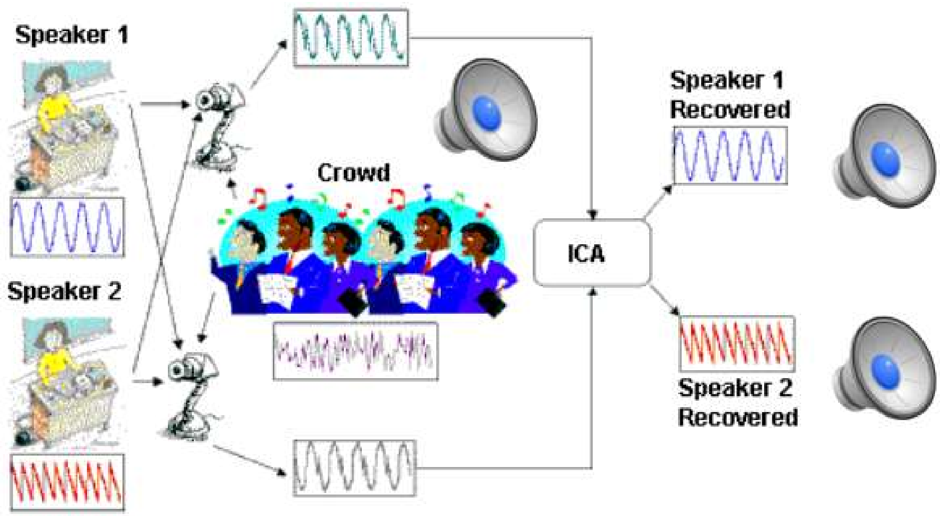
4/6/2024

pra-sâmi

84

Unsupervised Learning

- Independent component analysis – separate a combined signal into its original sources



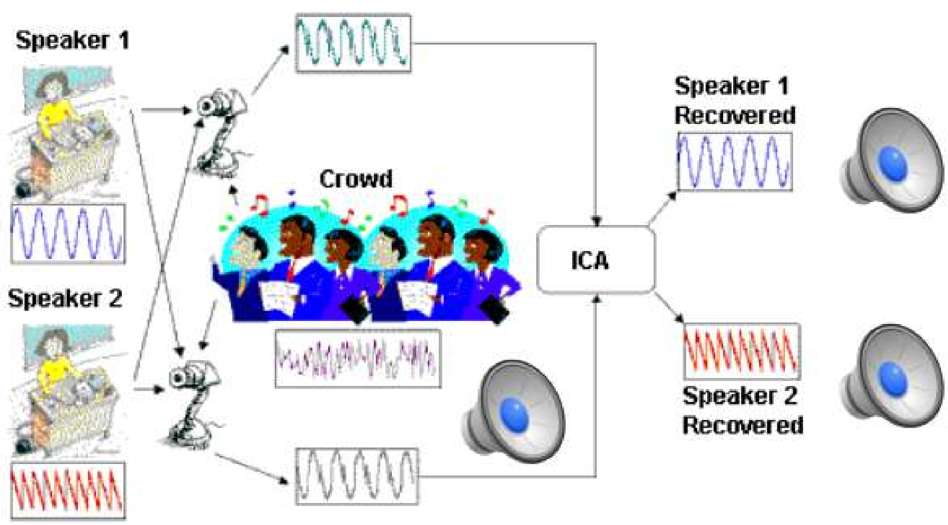
4/6/2024

pra-sâmi

85

Unsupervised Learning

- Independent component analysis – separate a combined signal into its original sources



4/6/2024

pra-sâmi

86

Dimensionality of the dataset…

- What is dimensionality of following dataset??

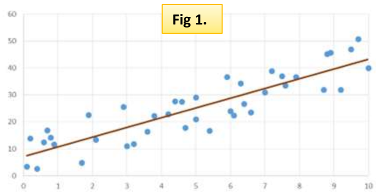


Fig 1.

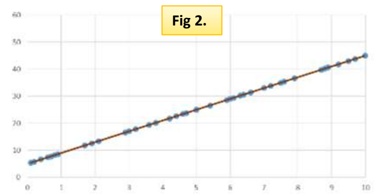


Fig 2.

- Data in Fig 1. has two dimensions
- Same is true for second figure as well…
- Can data in second fig be converted to one dimension???
- How about Fig. 3.

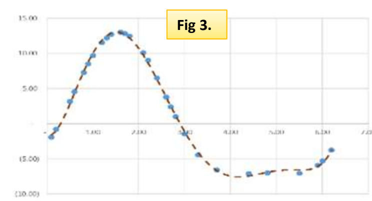


Fig 3.

87

Self-Supervised Learning

88

Self-Supervised Learning

- A specific type of supervised learning,
- Labels generated from the input data, typically using a heuristic algorithm
- Examples:
  - Autoencoders
  - Next frame in a video, given past frames,
  - Next word in a text, given previous words,

4/6/2024

pra-sâmi

89

Semi-Supervised Learning

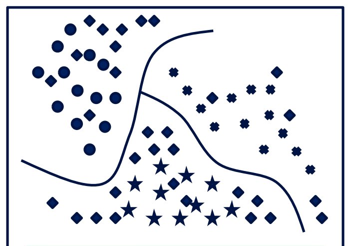
4/6/2024

pra-sâmi

90

Semi-supervised learning

- A combination of supervised and unsupervised
- Uses a small amount of labeled data and a large amount of unlabeled data
  - E.g. text document classifier



Semi-supervised learning

pra-sâmi

91

How Semi Supervised Learning Works

- Train the model with small amount of training data similar to supervised learning
  - The labels for this small training data are known
- Use the trained model with training set to predict the output
  - This output may be far-far away from the ground truth
- Link the labels from labeled training data with the predicted labels created above
- Link the input data in the labeled training set and that in unlabeled training set
- Train the model like supervised model using this pseudo labeled data and reduce the error

pra-sâmi

92

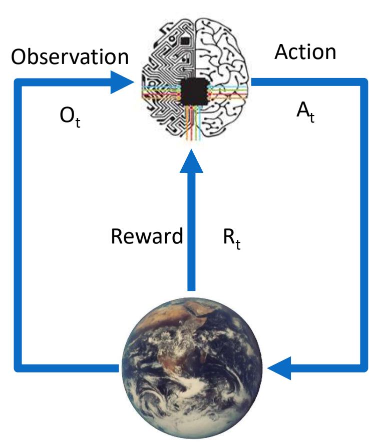
Reinforcement Learning

pra-sâmi

93

Reinforcement Learning

- Model learns by experience
  - Every action is measured in terms of “Reward”
  - Model tries to maximize the reward
  - Major application in gaming industry
- Given a sequence of states and actions with (delayed) rewards, output a policy
  - Policy is a mapping from states $\Rightarrow$ actions
    - Tells you what to do in a given state
- Examples:
  - Self-driving cars, Robotics, Resource management, Education, Credit assignment problem, Game playing, Robot in a maze, Balance a pole on your hand

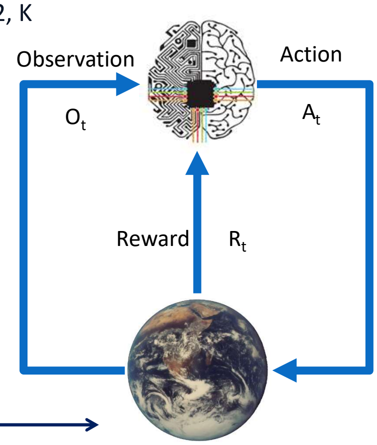


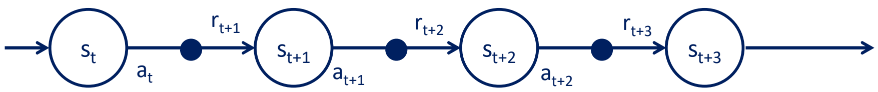
pra-sâmi

94

The Agent-Environment Interface

- Agent and environment interact at discrete time steps : t = 0, 1, 2, K
  - Agent observes state at step t: $s_t \in S$
  - Produces action at step t : $a_t \in A(s_t)$
  - Gets resulting reward : $r_{t+1} \in \Re$
  - And resulting next state : s_{t+1}





95

Q-Learning

- Q-learning is a popular algorithm used in reinforcement learning
- It is based on the Bellman equation
- The agent tries to learn the policies that can provide the best actions to perform for “maximizing” the rewards under particular circumstances
- The agent learns these optimal policies from past experiences
- In Q-learning, the Q is used to represent the quality of the actions at each state, and the goal of the agent is to maximize the value of Q

96

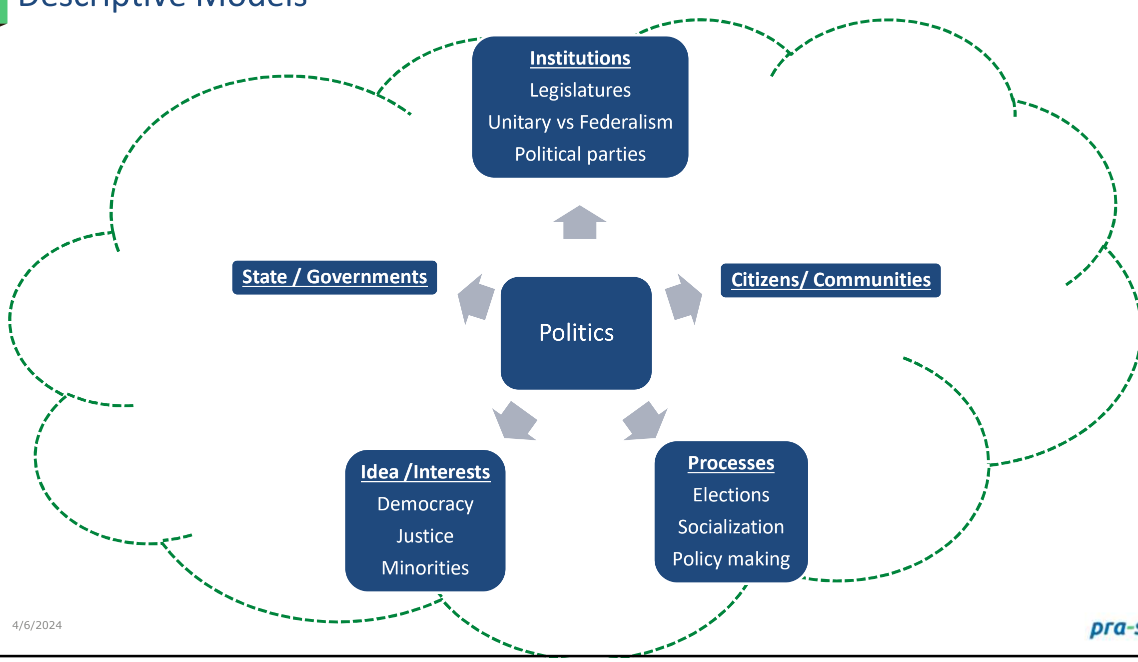
Descriptive Modelling

4/6/2024

pra-sâmi

97

Descriptive Models



4/6/2024

pra-sâmi

98

Descriptive Models

- Descriptive models are more flexible compared to predictive models
- A descriptive model describes the data in a form that allows for future action strategies
  - But it is not a precise event yet
  - It is a perspective looking at the data, so business can make sense of the data
- In Social Network Analysis
  - Between two person, if their views are closely associated one set of messaging is needed
  - Otherwise you may need different messaging!
- Personas are descriptive models of users that embody important and realistic user characteristics
  - Can be helpful in situations where iterative, rapid development is necessary,
  - Obtaining sufficient participants is not feasible, or
  - When user testing may require challenging situations for the target participant group(s)

4/6/2024

pra-sâmi

99

Reflect…

- How can you handle missing or corrupted data in a dataset?
  - Drop missing rows or columns
  - Assign a unique category to missing values
  - Replace missing values with mean/median/mode
  - All of the above
- Machine learning algorithms build a model based on ____data
  - Training Data
  - Transfer Data
  - Data Training
  - None of the above
- A Machine Learning technique that helps in detecting the outliers in data.
  - Clustering
  - Classification
  - Anamoly Detection
  - All of the above
- Real-Time decisions, Game AI, Learning Tasks, Skill acquisition, and Robot Navigation are applications of ....
  - Reinforcement Learning
  - Supervised Learning: Classification
  - Unsupervised Learning: Regression
  - None of the above

4/6/2024

pra-sâmi

Reflect

- Which of the following is not a supervised learning?
  - Naive Bayesian
  - PCA
  - Linear Regression
  - Decision Tree

- Which of the following is not type of learning?
  - Unsupervised Learning
  - Supervised Learning
  - Semi-unsupervised Learning
  - Reinforcement Learning

- Targeted marketing, Recommended Systems, and Customer Segmentation are applications in which of the following
  - Supervised Learning: Classification
  - Unsupervised Learning: Clustering
  - Unsupervised Learning: Regression
  - Reinforcement Learning

- Fraud Detection, Image Classification, Diagnostic, and Customer Retention are applications in which of the following
  - Unsupervised Learning: Regression
  - Supervised Learning: Classification
  - Unsupervised Learning: Clustering
  - Reinforcement Learning

ADDITIONAL MATERIAL

102

Examples of Reinforcement Learning

- Fly stunt maneuvers in a helicopter
- Defeat the world champion at Backgammon
- Manage an investment portfolio
- Control a power station
- Make a humanoid robot walk
- Play many different Atari games better than humans

103

Fly stunt maneuvers in a helicopter

- Stanford Computer Scientists have developed an artificial intelligence system that enables robotic helicopters to teach themselves to fly difficult stunts by watching other helicopters perform the same maneuvers.
- The technique is known as "apprenticeship learning."
- The result is an autonomous helicopter than can fly dazzling stunts on its own.

https://youtu.be/M-QUkgk3HyE

104

Game Over: Kasparov and the Machine (trailer)

The documentary "Game Over: Kasparov and the Machine", about the 1997 chess match between Garry Kasparov and Deep Blue, the IBM computer. It also features Yasser Seirawan, Anatoly Karpov and others.

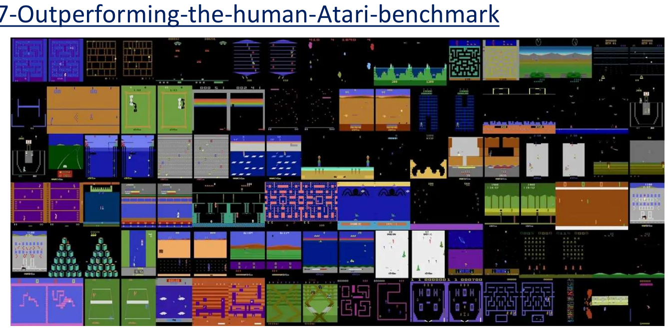
https://youtu.be/y9UMt-8gfW8

4/6/2024 pra-sâmi

105

DeepMind Made A Superhuman AI For 57 Atari Games!

- The Atari57 suite of games is a long-standing benchmark to gauge agent performance across a wide range of tasks. We've developed Agent57, the first deep reinforcement learning agent to obtain a score that is above the human baseline on all 57 Atari 2600 games. Agent57 combines an algorithm for efficient exploration with a meta-controller that adapts the exploration and long vs. short-term behaviour of the agent.
- https://deepmind.com/blog/article/Agent57-Outperforming-the-human-Atari-benchmark



- https://youtu.be/dJ4rWhpAGFI

4/6/2024 pra-sâmi

106

Automated Cargo Wharf Yangshan, Shanghai

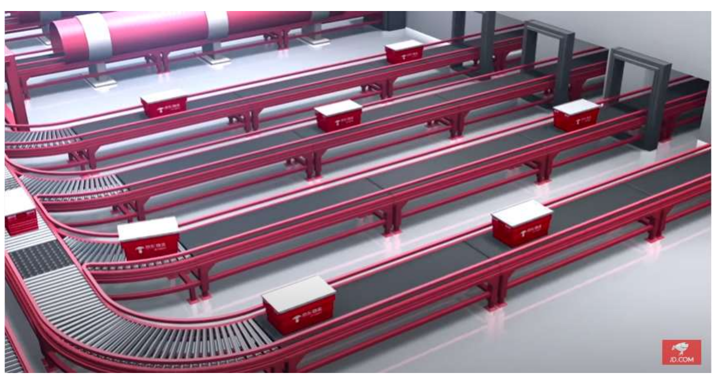
- World's biggest automated cargo wharf, the fourth phase of the Yangshan deep-water port started operation
- The core technology of the robotic port was developed independently by China
- The forth phrase of Yangshan port takes up an area of 2.23 million square meters
- Coastline stretches as long as 2,350 meters
- It consists of two 70,000 dead-weight tonnage (DWT) berths and five 50,000 DWT berths

https://youtu.be/IzOeAGAu60k

107

Urban Logistics in Smart Cities

- JD.com (Nasdaq: JD) With over 300 million customers, have a vast network of warehouses and delivery stations, and deliver most orders in less than a day

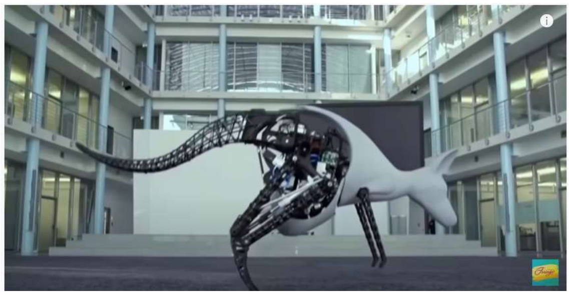


https://www.youtube.com/watch?v=XGSl9DCkxvo

108

Advanced & Futuristic Animal Robots That Exist In Real World

- We all have seen robots
  - in our lives, in films and other times in real world,
- We are surrounded by robots at all walks of our life today.
- See animal robots that look 100 percent similar to their real counterparts ?
- These futuristic animal robots can make our life simple!



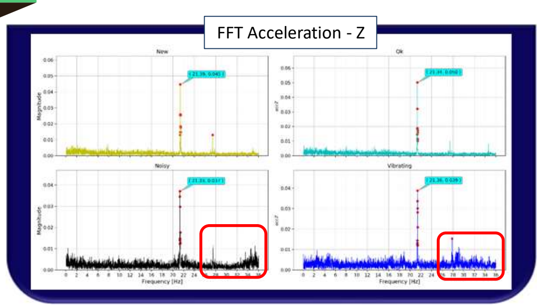
https://www.youtube.com/watch?v=OKse21PN30A

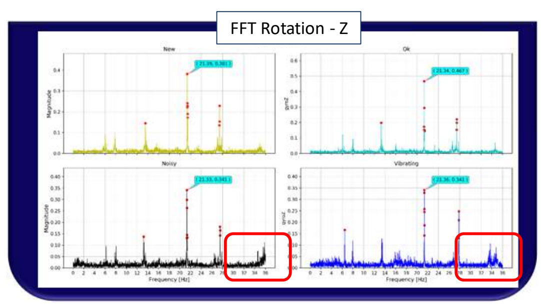
4/6/2024

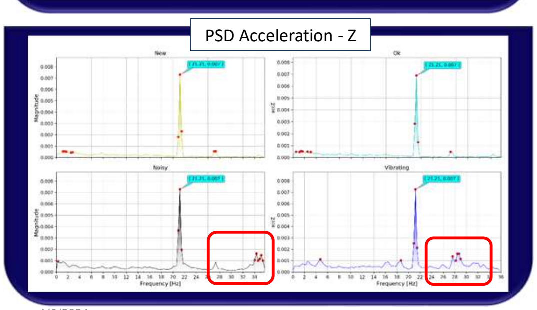
pra-sâmi

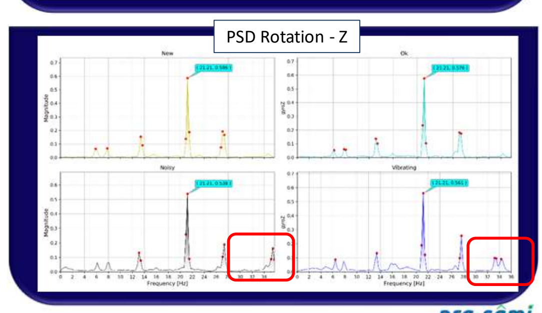
109

FFT and PSD analysis









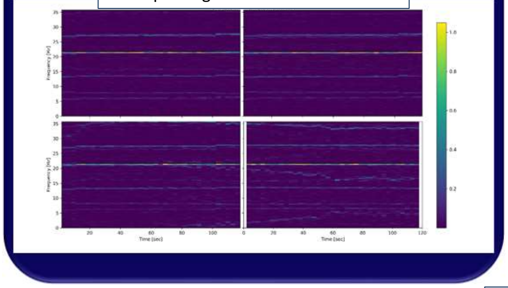
4/6/2024

pra-sâmi

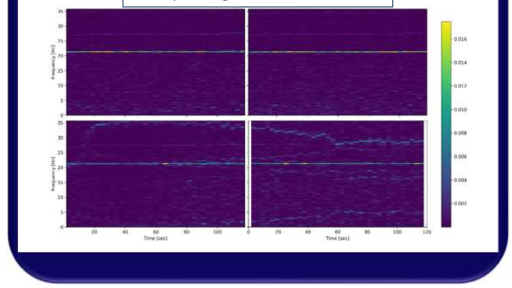
110

Spectrogram & P-F Curve

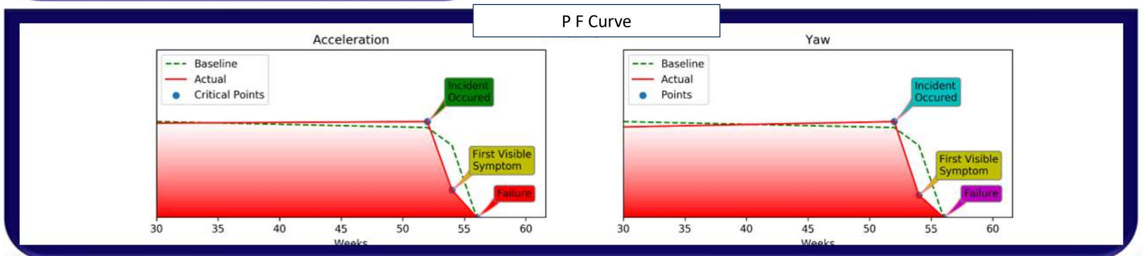
Spectrogram Acceleration - Z



Spectrogram Rotation - Z



P F Curve



pra-sâmi