Foundation of Artificial Intelligence

Revision slides
L7 Machine Learning
L8 Type of Machine Learning
L9 Supervised Learning
L10 UnSupervised Learning
L11 Reinforcement Learning
L12 Deep Learning
L13 AI System Development

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Foundation of Artificial Intelligence

Revision

Week 15
Thurs

L7 Machine Learning

L8 Type of Machine Learning

L9 Supervised Learning

L10 UnSupervised Learning

Week16
Thurs

L11 Reinforcement Learning

L12 Deep Learning

L13 AI System Development

Consultation

Week 16
Wed(Tutorial/Practical)

Week16
Friday(Tutorial/Practical)

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Foundation of Artificial Intelligence

Revision
L7 Machine Learning

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Learning Objectives

Understand concept of Machine Learning

Describe how the machine learning application is formed

Copyright © Nanyang Polytechnic.
All Rights Reserved.

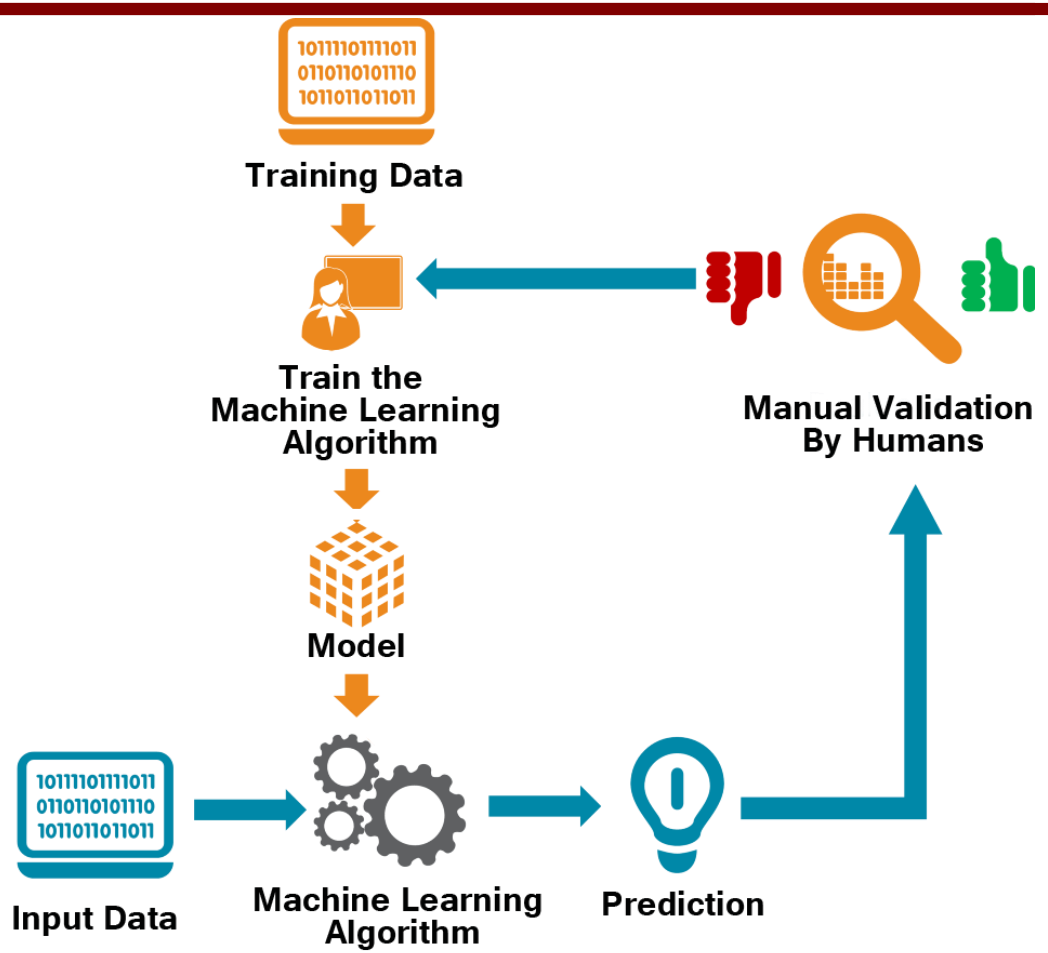
Machine Learning Approach

- Machine Learning is an application of artificial intelligence (AI) that provides systems the ability to **automatically learn** and i**mprove from experience without being explicitly programmed**. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Machine Learning Approach

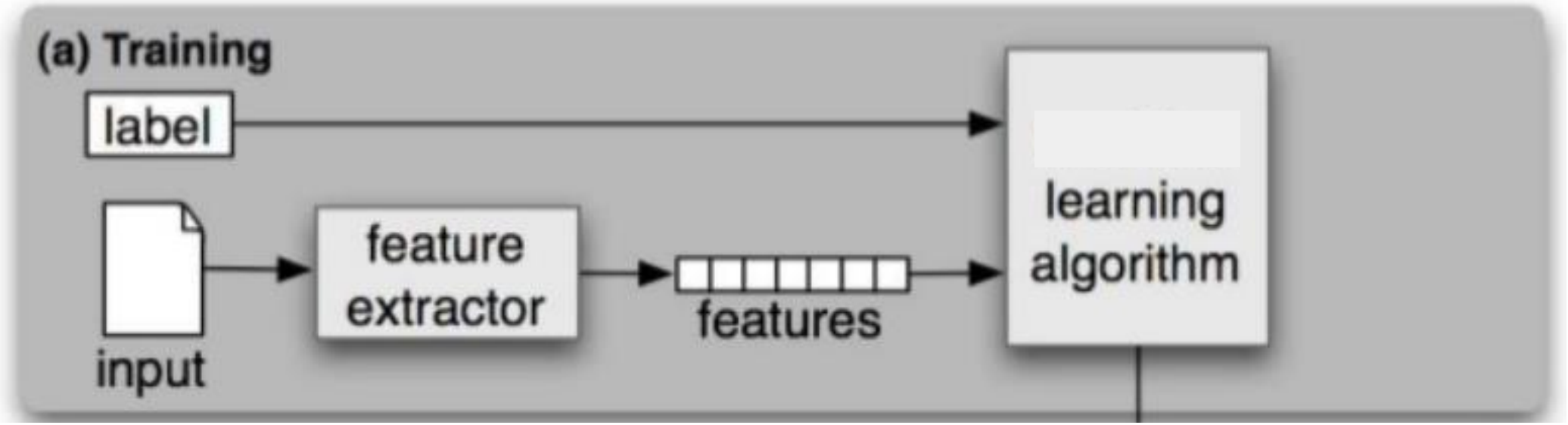
- **direct experience, or instruction**, in order to look for **patterns in data** and **make better decisions** in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.



Copyright © Nanyang Polytechnic.
All Rights Reserved.

Learning Approach

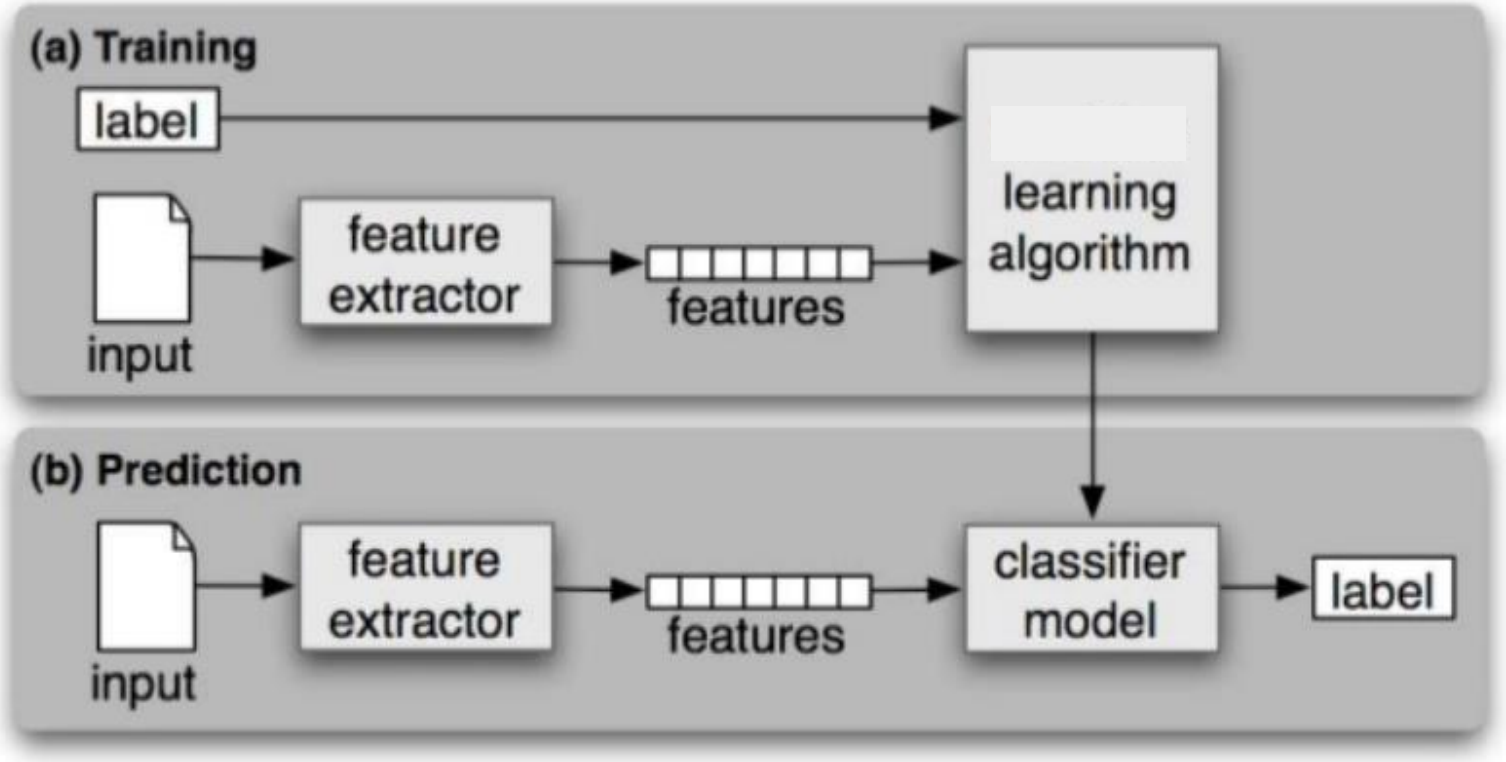
- **Learning- Model learn from data**



Copyright © Nanyang Polytechnic.
All Rights Reserved.

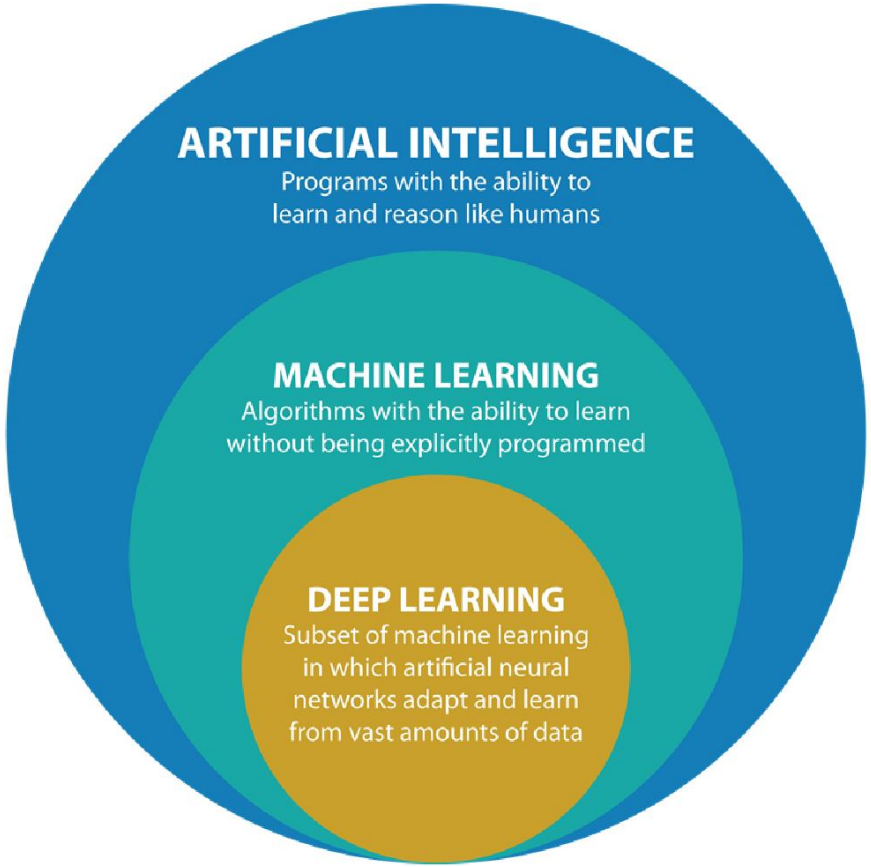
Machine Learning Approach

- **Machine Learning- Inferencing data for prediction/classification**



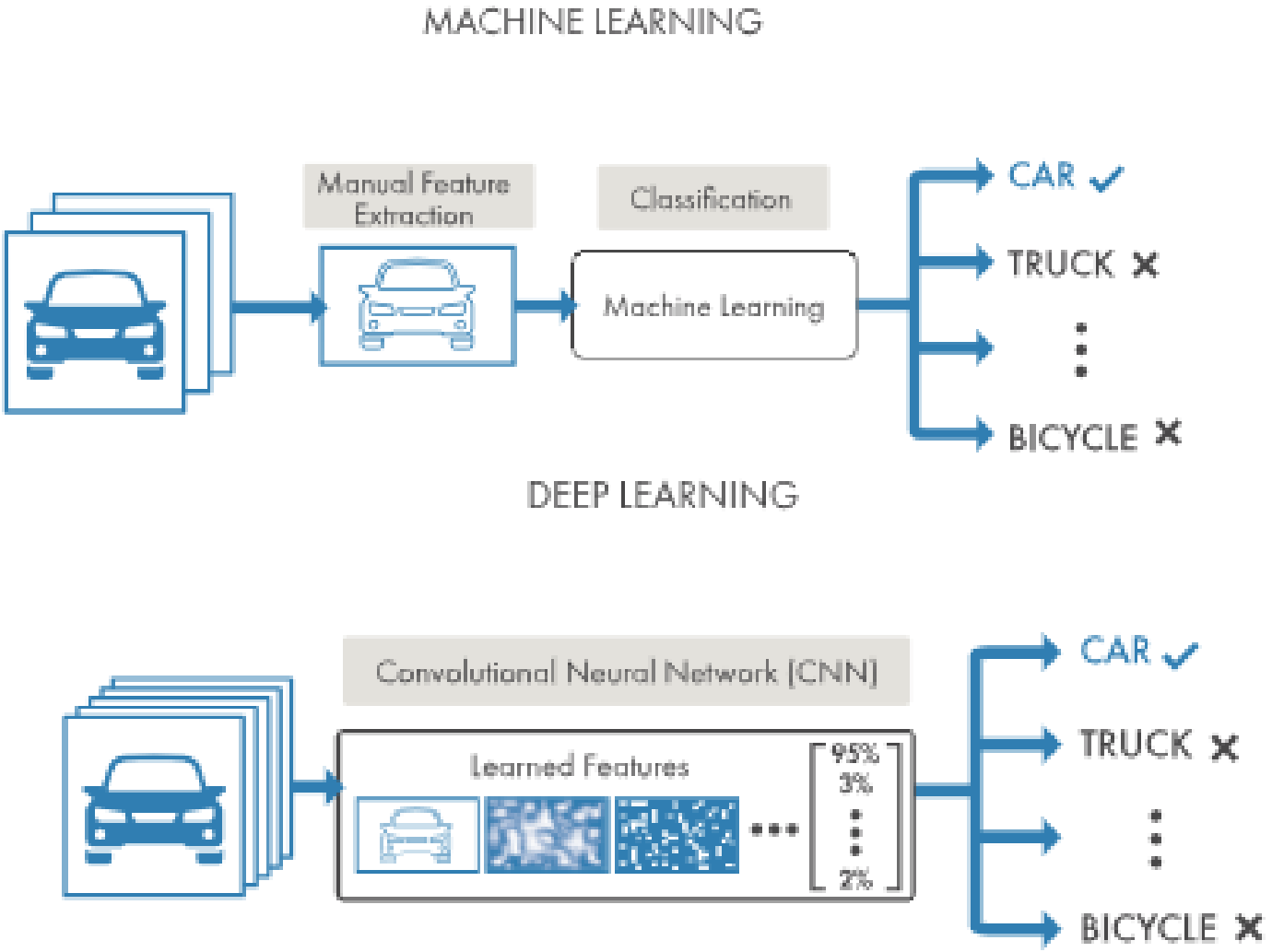
Copyright © Nanyang Polytechnic.
All Rights Reserved.

Differences between AI, ML and DL



Copyright © Nanyang Polytechnic.
All Rights Reserved.

Difference Between Machine Learning & Deep Learning

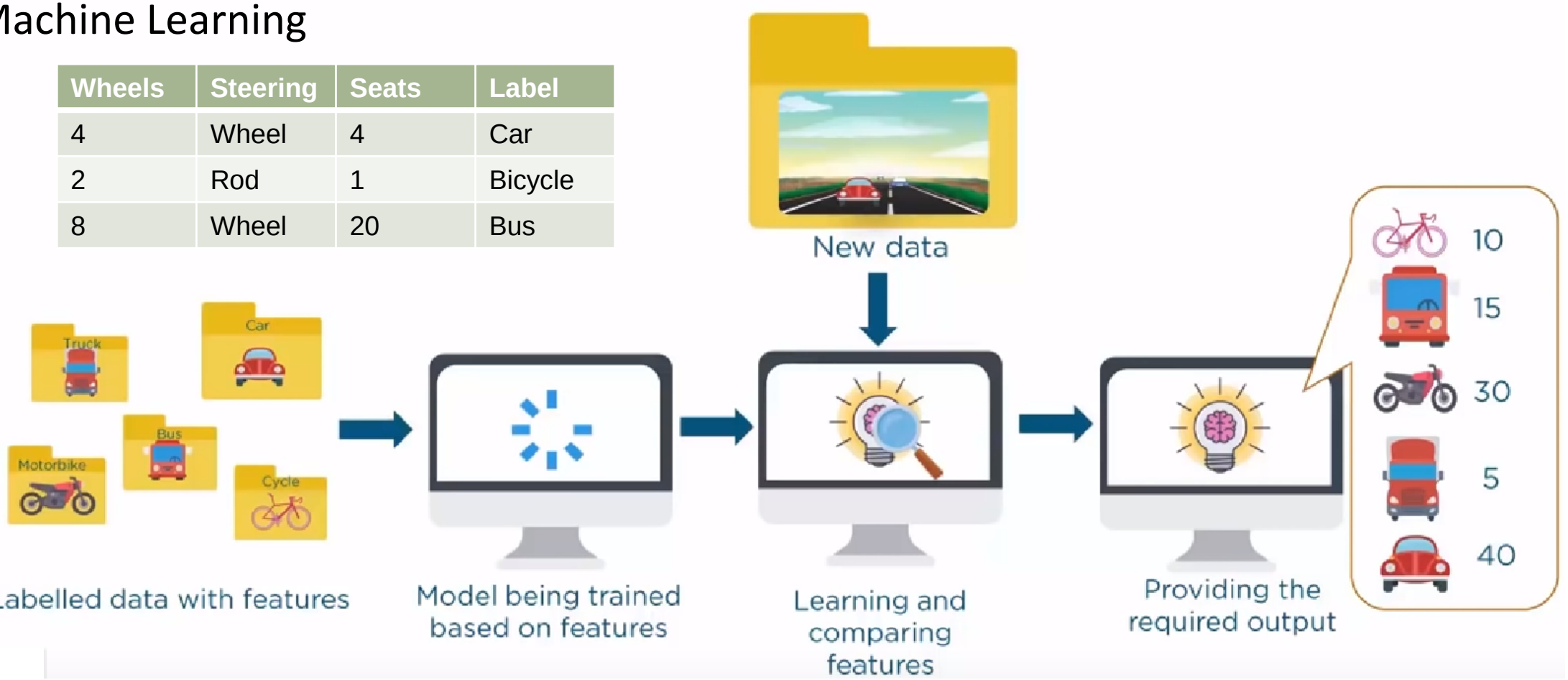


Copyright © Nanyang Polytechnic.
All Rights Reserved.

Difference Between Machine Learning & Deep Learning

Machine Learning

Wheels	Steering	Seats	Label
4	Wheel	4	Car
2	Rod	1	Bicycle
8	Wheel	20	Bus



Copyright © Nanyang Polytechnic.
All Rights Reserved.

Machine Learning vs Deep Learning

When this problem is solved through machine learning:

To help the ML algorithm categorize the images in the collection according to the two categories of dogs and cats, you will need to present to it these images collectively. But how does the algorithm know which one is which?

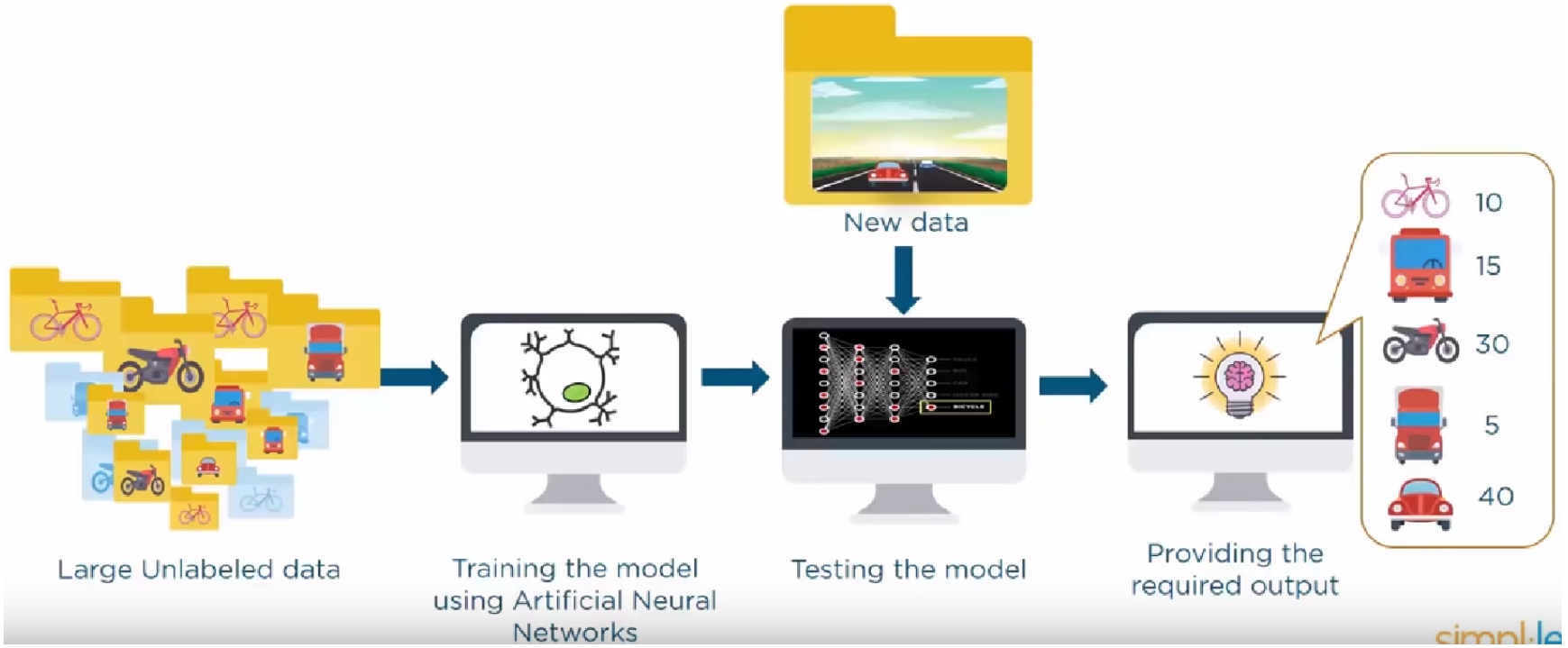
The answer to this question, as in the above definition of machine learning for dummies, is **structured data**. You simply label the pictures of car ,bicycle and motorcycle in a way which will define specific features of the vehicles. This data will be enough for the machine learning algorithm to learn, and then it will continue working based on the labels that it understood, and classify millions of other pictures of vehicles as per the features it learned through the said labels.

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Difference Between Machine Learning & Deep Learning

Deep Learning

Photos	Labels
Car001.jpg	Car
Bicycle001.jpg	Bicycle
Car002.jpg	Car
……..	…..



Copyright © Nanyang Polytechnic.
All Rights Reserved.

Foundation of Artificial Intelligence

Revision
L8 Type of Machine Learning

Copyright © Nanyang Polytechnic.
All Rights Reserved.

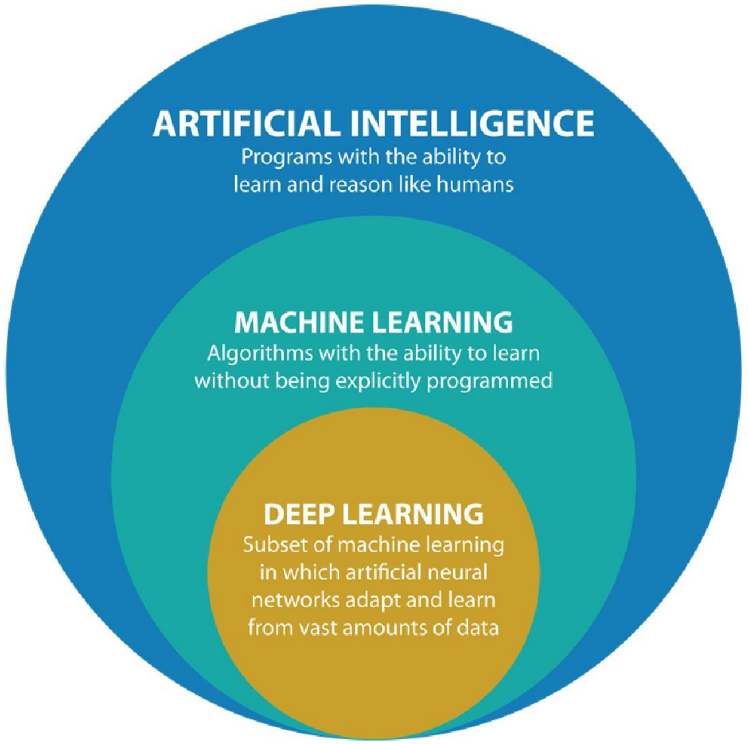
Learning Outcomes

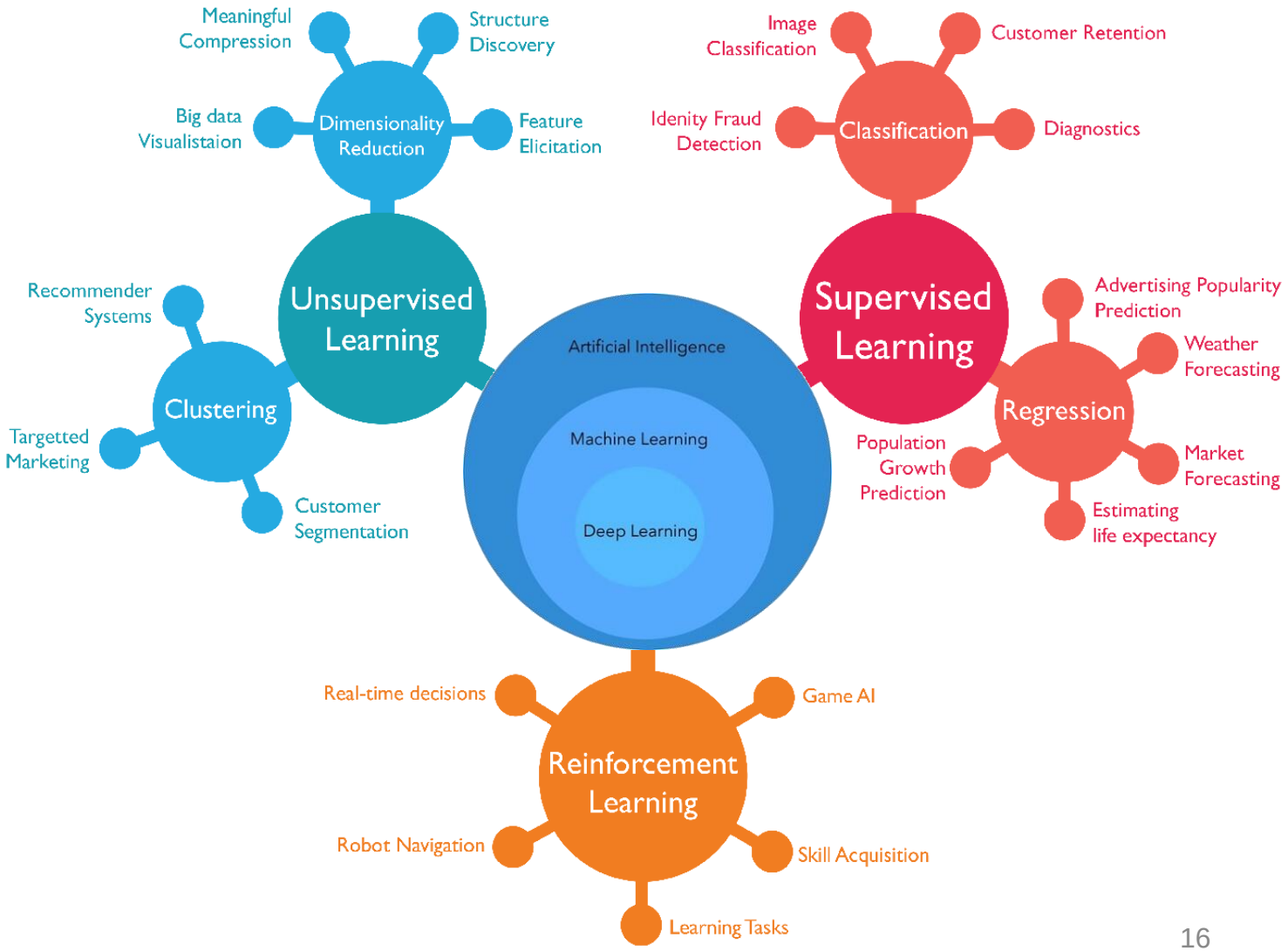
Describe the type of Machine Learning

Apply the type of Machine learning for different purposes

Copyright © Nanyang Polytechnic.
All Rights Reserved.

AI Technique Overview



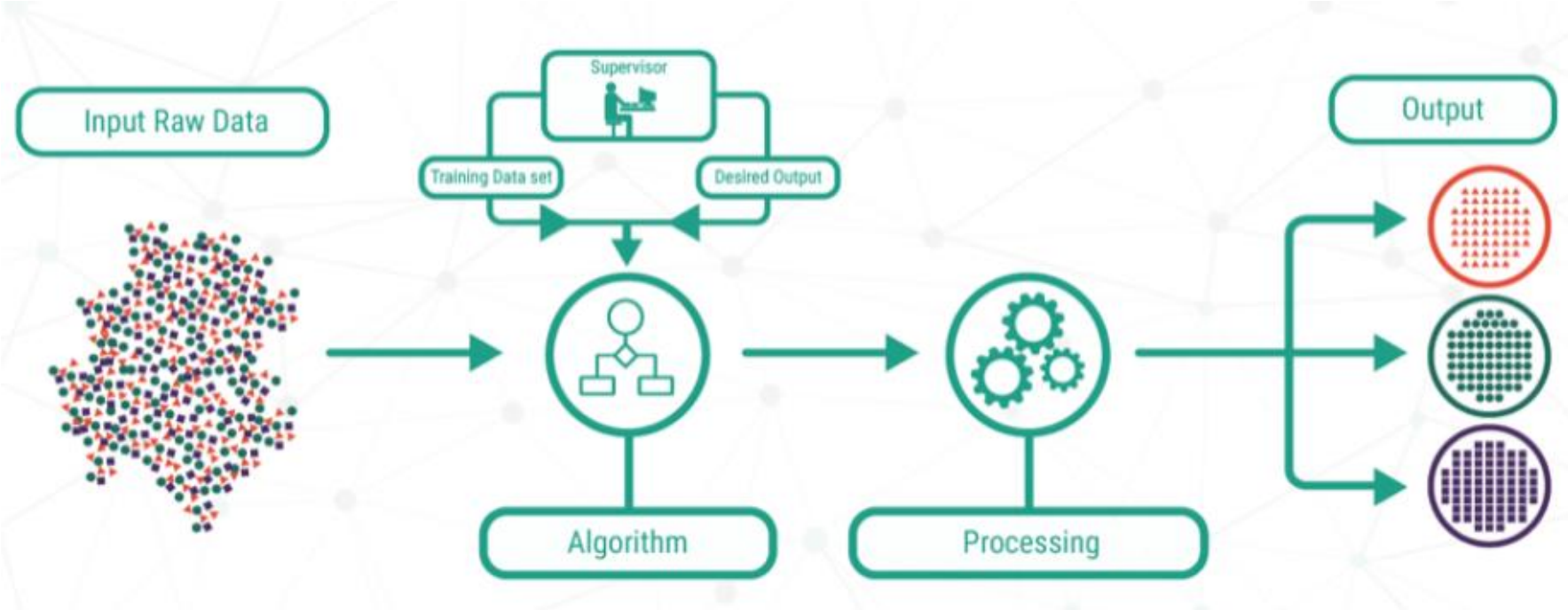


Copyright © Nanyang Polytechnic.
All Rights Reserved.

AI Technique Overview

Supervised Learning

A supervised learning algorithm takes a known set of input data and known responses to the data (output) and trains a model to generate reasonable predictions for the response to new data.



Copyright © Nanyang Polytechnic.
All Rights Reserved.

AI Technique Overview

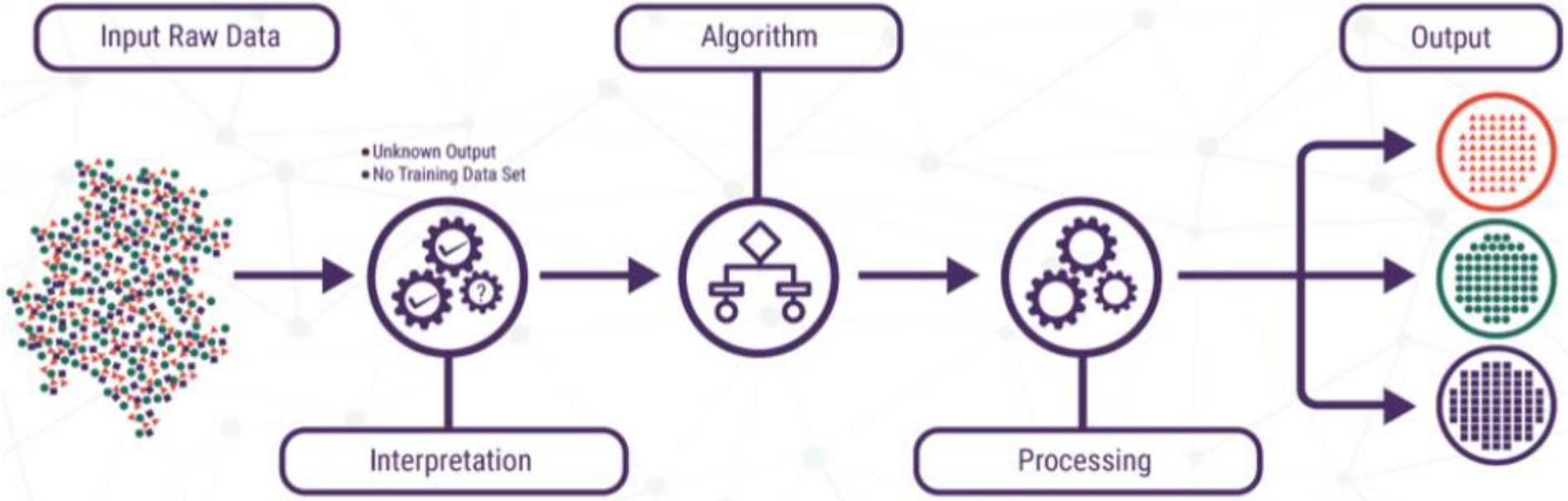
- **Supervised Learning Algorithms**
  - **Support Vector Machines**
  - **Linear regression**
  - **Logistic regression**
  - **Naive Bayes**
  - **Linear discriminant analysis**
  - **Decision trees**
  - **K-Nearest Neighbour algorithm (KNN)**
  - **Neural Networks (Multilayer perceptron)**

Copyright © Nanyang Polytechnic.
All Rights Reserved.

AI Technique Overview

Unsupervised Learning

Unsupervised learning finds hidden patterns or intrinsic structures in data. It is used to draw inferences from dataset consisting of input data without labelled responses.

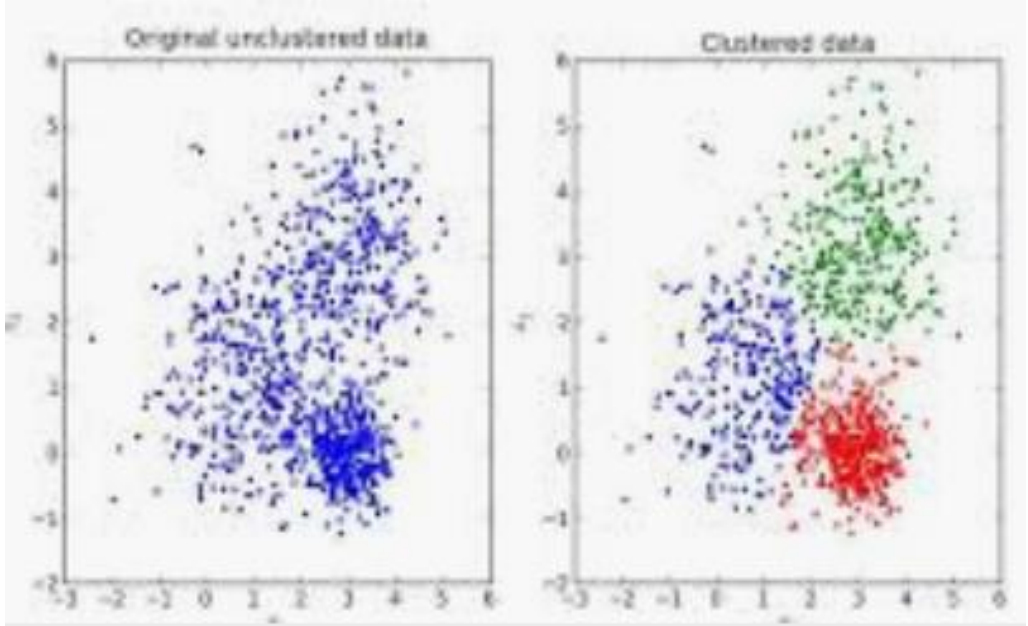


Copyright © Nanyang Polytechnic.
All Rights Reserved.

AI Technique Overview

Unsupervised Learning

Clustering is the most common unsupervised learning technique. It is used for exploratory data analysis to find hidden patterns or groupings in data. Applications for cluster analysis include gene sequence analysis, market research, and object recognition.



Copyright © Nanyang Polytechnic.
All Rights Reserved.

AI Technique Overview

- **Unsupervised Learning Algorithms**
  - **Clustering**
  - **Anomaly detection**
  - **Dimensionality Reduction**

Copyright © Nanyang Polytechnic.
All Rights Reserved.

AI Technique Overview

Reinforcement Learning

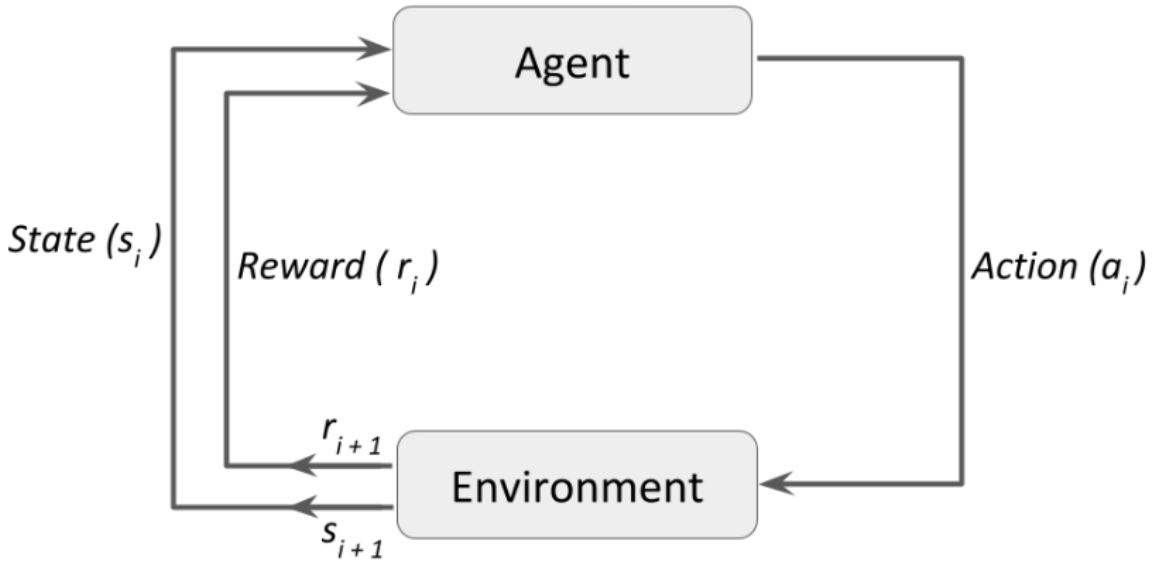
Reinforcement learning in formal terms is a method of machine learning wherein the software agent learns to perform certain actions in an environment which lead it to maximum reward. It does so by exploration and exploitation of knowledge it learns by repeated trials of maximizing the reward.

Copyright © Nanyang Polytechnic.
All Rights Reserved.

AI Technique Overview

Basic Concept and Terminology

In the previous example, you are the **agent** who is trying to walk across the field, which is the **environment**. Walking is the **action** the **agent** performs on the **environment**. The distance the **agent** walks acts as the **reward**. The **agent** tries to perform the **action** in such a way that the **reward** maximizes. This is how Reinforcement Learning works in a nutshell. And in the proper technical terms, and generalizing to fit more examples into it, the diagram becomes -



Copyright © Nanyang Polytechnic.
All Rights Reserved.

AI Technique Overview

Q Learning

Q-learning is one of the most popular Reinforcement learning algorithms. Q Learning is an algorithm to learn(1) ***a policy*** (2) that will tell us how to interact(3) with an environment(4) under different circumstances(5) in such a way to maximize rewards(6).

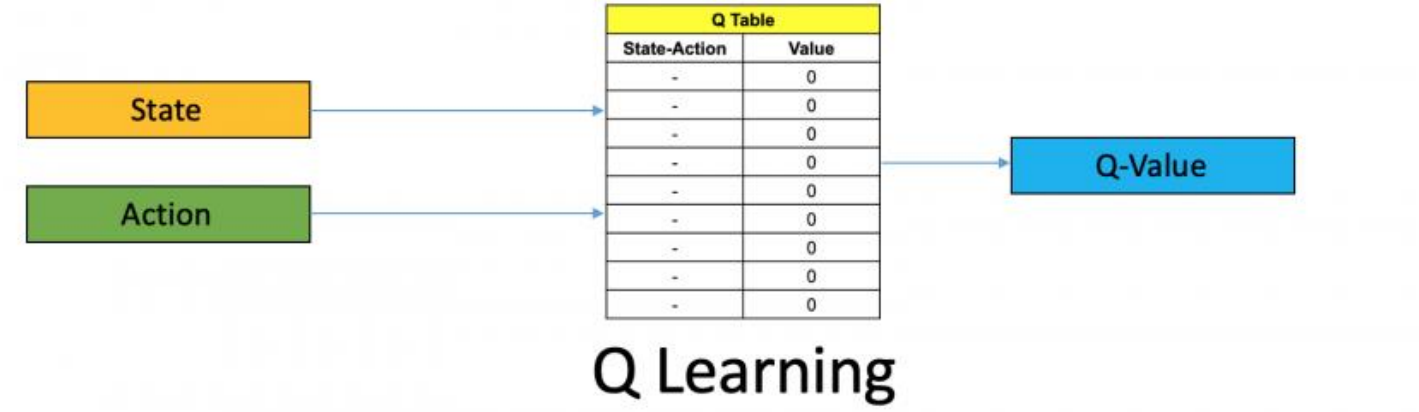
Deep Q-Networks

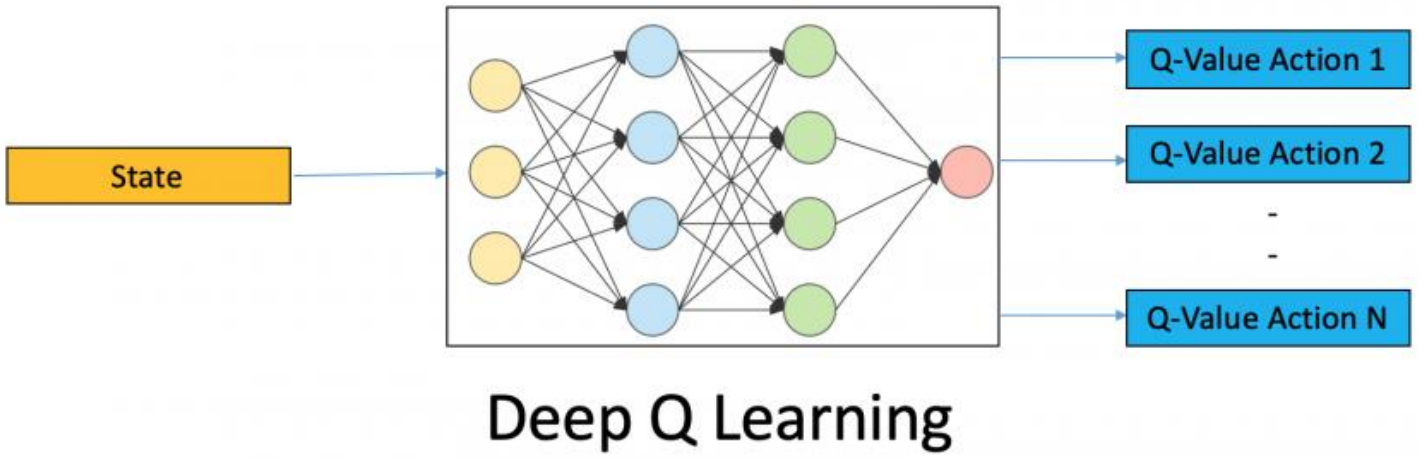
In deep Q-learning, we use a neural network to approximate the Q-value function. The state is given as the input and the Q-value of all possible actions is generated as the output

Copyright © Nanyang Polytechnic.
All Rights Reserved.

AI Technique Overview

The comparison between Q-learning & deep Q-learning is wonderfully illustrated below:





Copyright © Nanyang Polytechnic.
All Rights Reserved.

AI Learning-Data

Different types of data sources

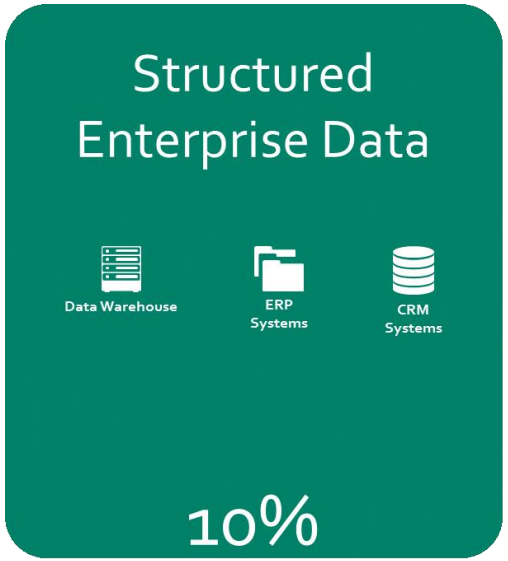
Data manifests itself in many different shapes. Each shape of data may hold much value to the business. In some shapes this is easier to extract than others. Different shapes of data require different storage solutions and should therefore be dealt with in different ways. We can distinguish between three shapes of data:

- Structured data
- UnStructured data
- Semi-Structured data

Copyright © Nanyang Polytechnic.
All Rights Reserved.

AI Learning-Data

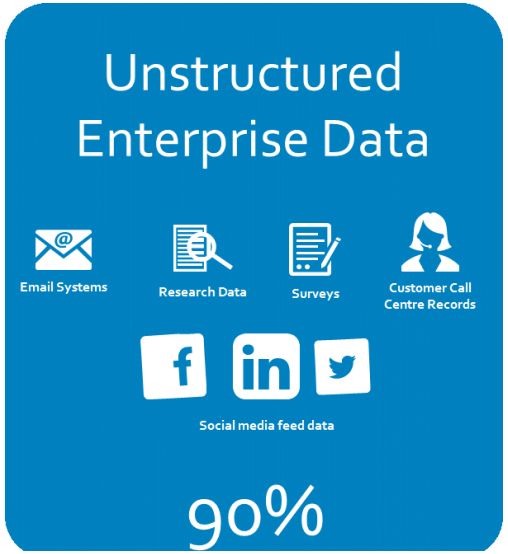
Structured data is tabular data (rows and columns) which are very well defined. Meaning that we know which columns there are and what kind of data they contain. Often such data is stored in ***databases***. In databases we can use the power of the language SQL to answer queries about the data and easily create data sets to use in our data science solutions.



Copyright © Nanyang Polytechnic.
All Rights Reserved.

AI Learning-Data

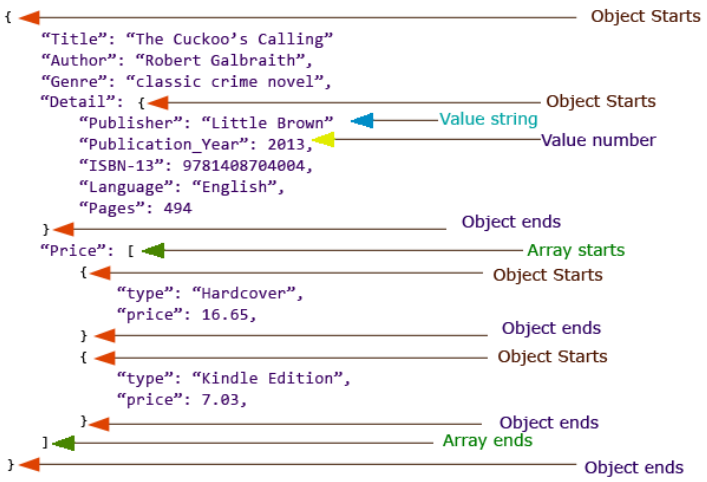
Unstructured data is the rawest form of data. It can be any type of file, for example; ***texts, pictures, sounds or videos***. This data is often stored in a repository of files. Extracting value out of this shape of data is often the hardest. You first need to extract structured features from the data that describe then abstract meaning from it. For example, to use text you might want to extract the topics and whether the text is positive or negative about them.

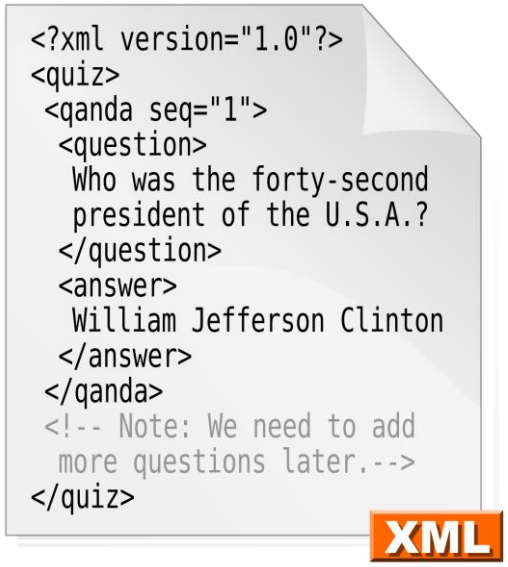


Copyright © Nanyang Polytechnic.
All Rights Reserved.

AI Learning-Data

Semi-structured data is anywhere between unstructured and structured data. A consistent format is defined however the structure is not very strict, like it is not necessarily tabular and parts of the data may be incomplete or differing types. Semi-structured data are often stored as files. However, some kinds of semi-structured data (like ***JSON or XML***) can be stored in document oriented-databases. Such databases allow you to query the semi-structured data.





Copyright © Nanyang Polytechnic.
All Rights Reserved.

AI Learning-Data

What are the data types?
Most data fall into one of two groups: numerical or categorical.

- **Numerical data.** These ***data have meaning as a measurement***, such as a person's height, weight, IQ, or blood pressure; or they're a count, such as the number of stock shares a person owns, how many teeth a dog has, or how many pages you can read of your favorite book before you fall asleep. (Statisticians also call numerical data ***quantitative data***.)
- **Categorical data:** Categorical ***data represent characteristics*** such as a person's gender, marital status, hometown, or the types of movies they like. Categorical data can take on numerical values (such as "1" indicating male and "2" indicating female), but those numbers ***don't have mathematical meaning***. You couldn't add them together, for example. (Other names for categorical data are ***qualitative data***, or *Yes/No data.*)

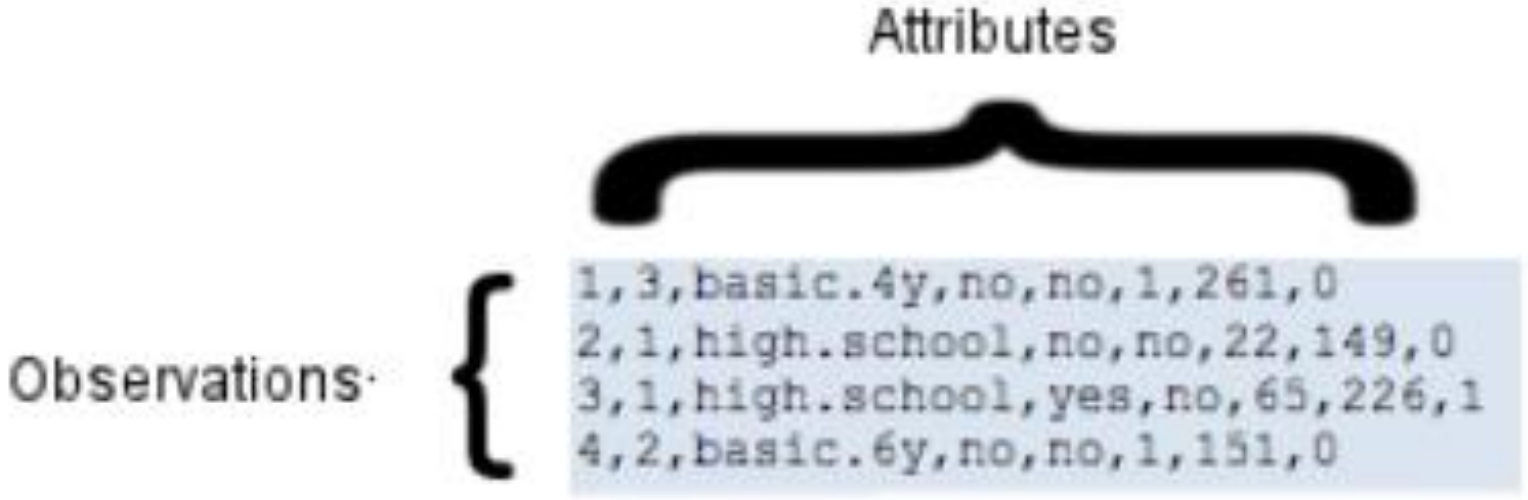
Copyright © Nanyang Polytechnic.
All Rights Reserved.

AI Learning-Data

Format data into dataset for learning

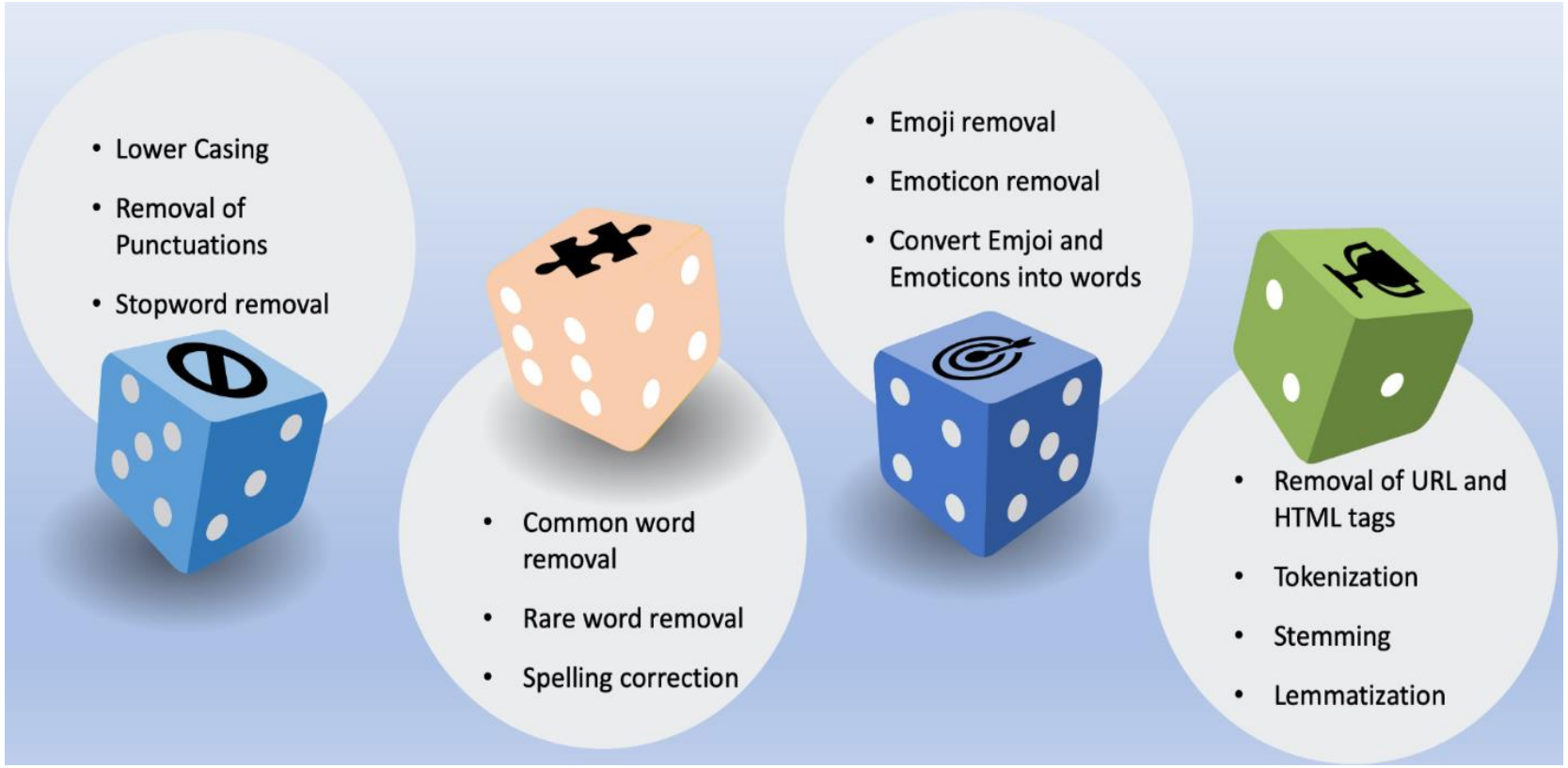
- You must save your input data in the comma-separated values (.csv) format. Each row in the .csv file is a single data record or observation. Each column in the .csv file contains an attribute of the observation.

customerId,jobId,education,housing,loan,campaign,duration,willRespondTo Campaign.



Copyright © Nanyang Polytechnic.
All Rights Reserved.

Text pre-process and clustering



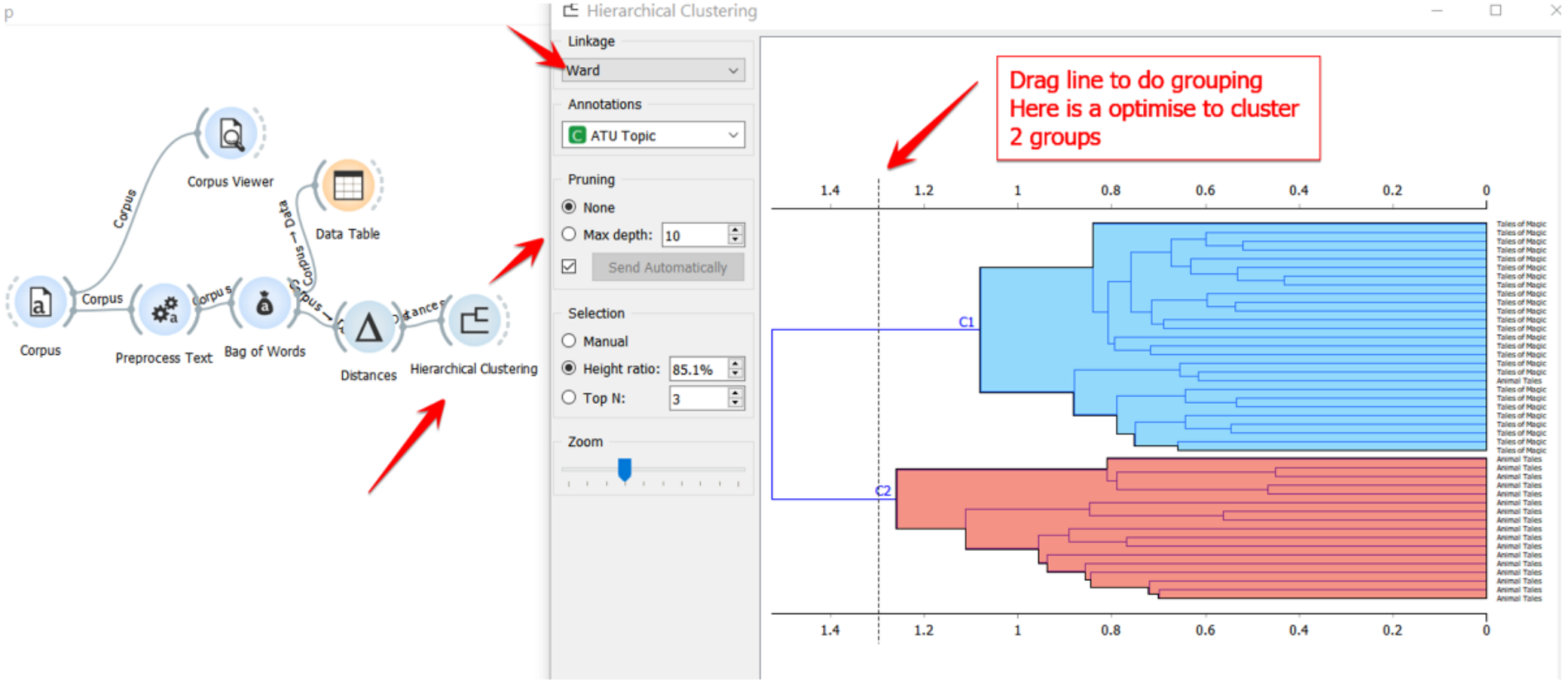
Text preprocessing
-lower case
-removal of punctuations
-stopword removal
-Tokenization
-Stemming
-Lemmatization

Convert to dataset
-Bag of word

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Text pre-process and clustering

Analysis the cluster using Hierarchical clustering



Copyright © Nanyang Polytechnic.
All Rights Reserved.

Foundation of Artificial Intelligence

Revision
L9 Supervised Learning

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Learning Outcomes

Understand the Supervised Learning Regression approach

Understand the Supervised Learning Classification approach

Conduct the Regression Model Training

Conduct the Classification Model Training

Apply the Regression Model to do inference

Apply the Classification Model to do inference

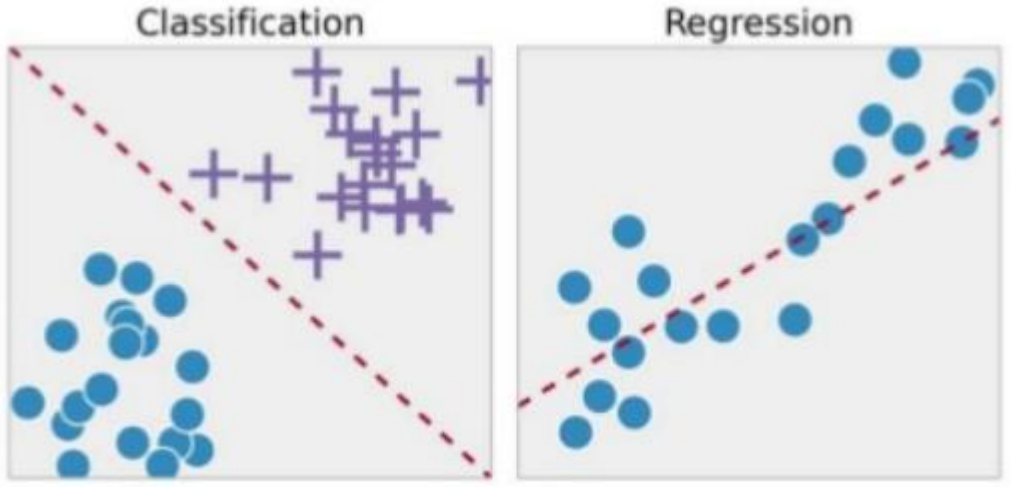
Copyright © Nanyang Polytechnic.
All Rights Reserved.

Supervised learning

In supervised learning, we are given a data set and already know what our correct output should look like, having the idea that there is a relationship between the input and output.

Two types of Supervised Learning :-

1. Regression — Estimate continuous values (Real valued output)
2. Classification — Identify a unique class (Discrete values, Boolean or Categories)



Copyright © Nanyang Polytechnic.
All Rights Reserved.

Supervised learning

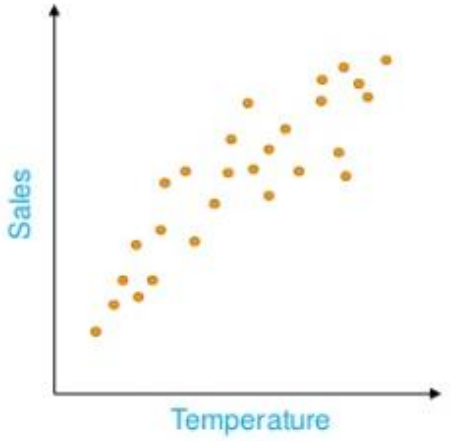
- Algorithms
  - Linear regression
  - Logistic regression
  - Decision trees
  - K-Nearest Neighbour algorithm (KNN)
  - Support Vector Machines
  - Naive Bayes
  - Linear discriminant analysis
  - Neural Networks (Multilayer perceptron)

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Supervised learning-Linear Regression

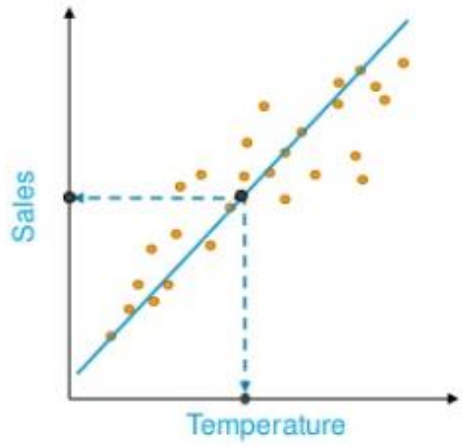
Linear Regression is a linear modelling algorithm to find relationship between one or more independent variables(predictors) denoted as X and dependent variable (target) denoted as Y

Example: Regression Prediction of sales of Ice Cream based on temperature



Plotting the sales of ice cream based on temperature

$$y=wx + b$$



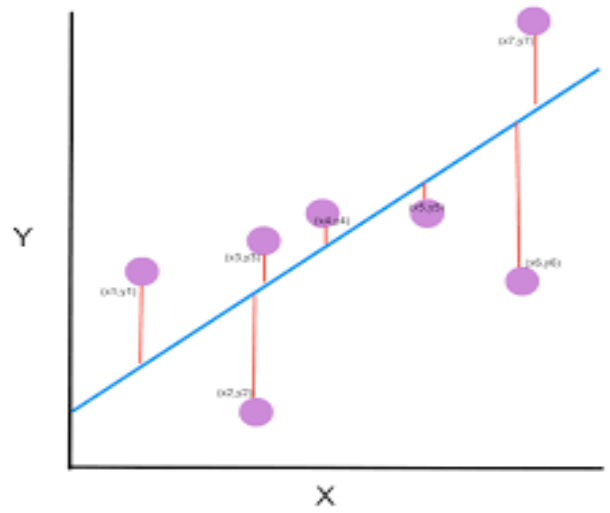
Regression line to predict the sales

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Supervised learning-Linear Regression

Regression performance measurement

Mean Absolute Error.



Mean Absolute Error

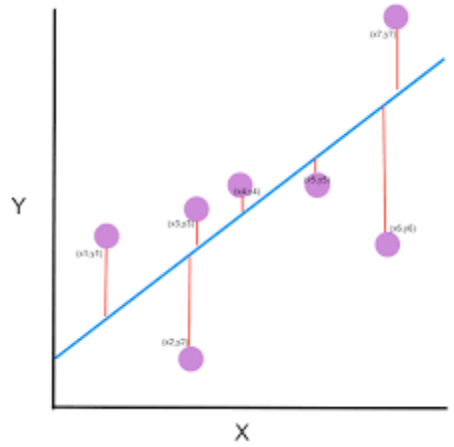
$$\mathrm{MAE} = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j|$$

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Supervised learning-Linear Regression

Regression performance measurement

Mean Square Error.



Mean Square Error

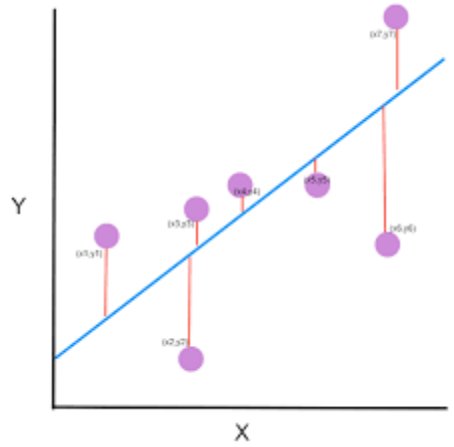
$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}$$

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Supervised learning-Linear Regression

Regression performance measurement

Root Mean Square Error



Root Mean Square Error

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

Copyright © Nanyang Polytechnic.
All Rights Reserved.

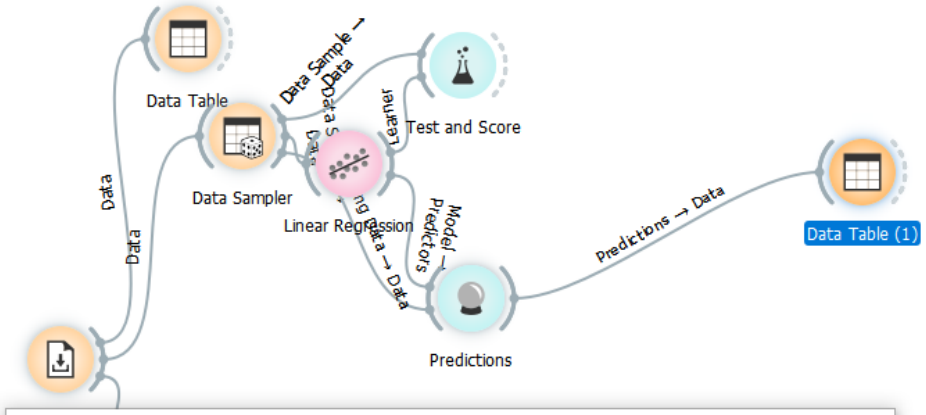
Supervised learning-Linear Regression

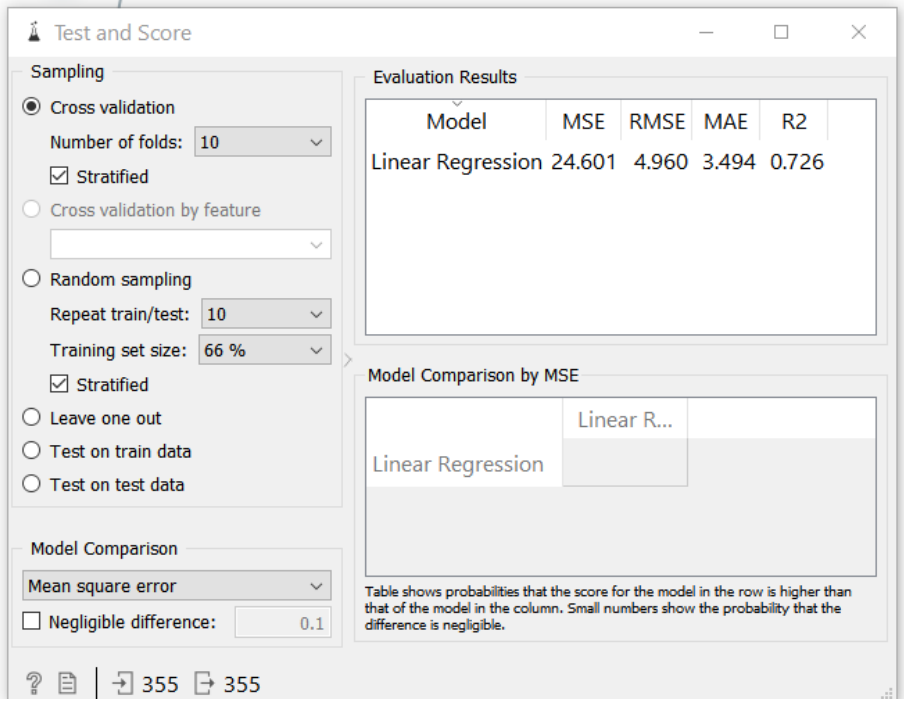
	X	Y	Est(Yhat)	error(Y-Yhat)	error sq
y=2.7x+1.5	0	2	1.5	0.5	0.25
	1	3.7	4.2	-0.5	0.25
	2	5.1	6.9	-1.8	3.24
	3	7.4	9.6	-2.2	4.84
	5	13.4	15	-1.6	2.56
				mse	2.228

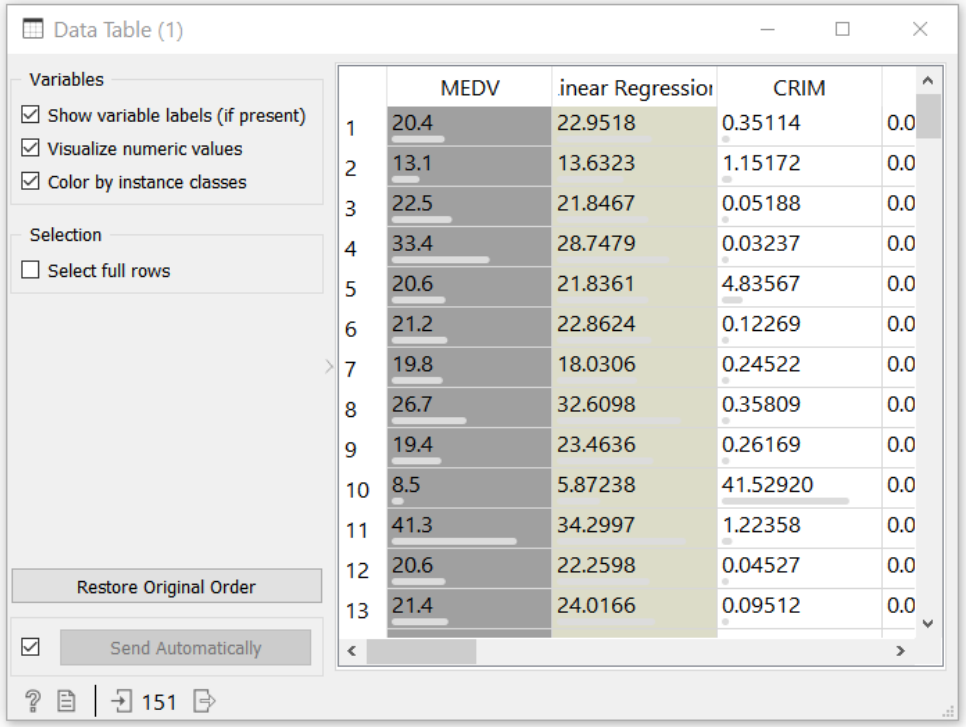
	X	Y	Est(Yhat)	error(Y-Yhat)	error sq
y=1.2x+2.3	0	2	2.3	-0.3	0.09
	1	3.7	3.5	0.2	0.04
	2	5.1	4.7	0.4	0.16
	3	7.4	5.9	1.5	2.25
	5	13.4	8.3	5.1	26.01
				mse	5.71

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Supervised learning-Linear Regression





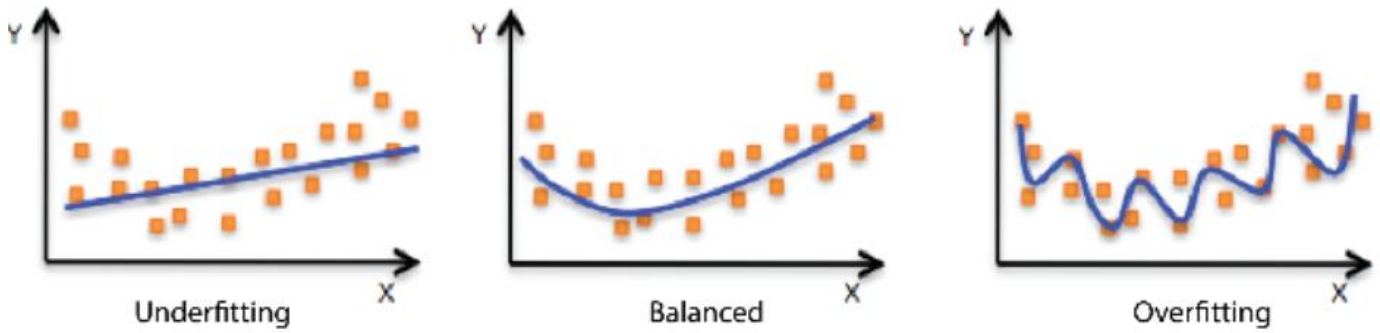


Copyright © Nanyang Polytec
All Rights Reserved.

Supervised learning-Linear Regression

Model Fit: Underfitting vs. Overfitting

Understanding model fit is important for understanding the root cause for poor model accuracy. This understanding will guide you to take corrective steps. We can determine whether a predictive model is underfitting or overfitting the training data by looking at the prediction error on the training data and the evaluation data.



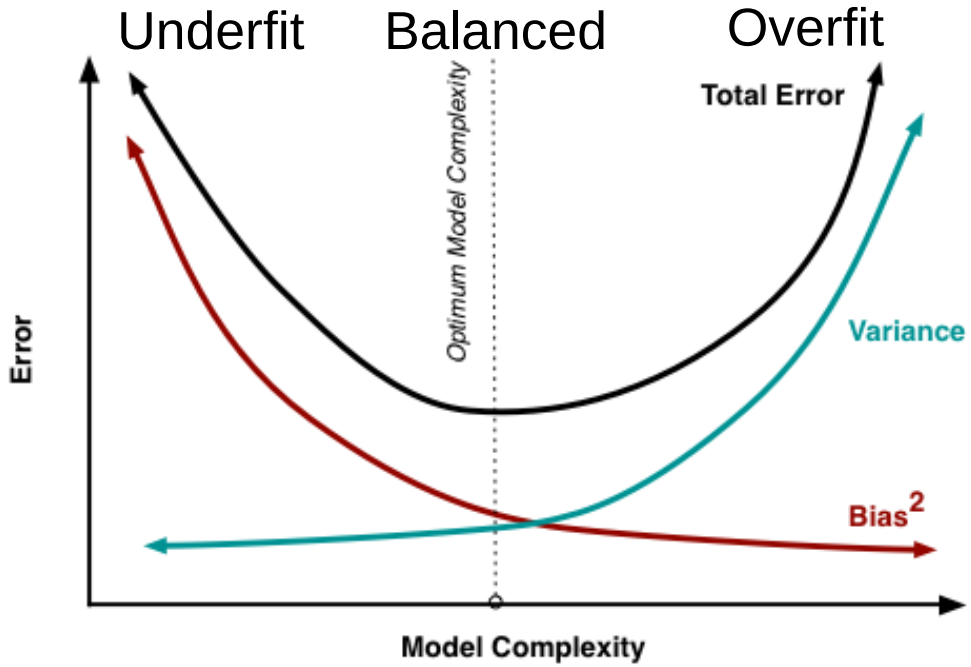
Copyright © Nanyang Polytechnic.
All Rights Reserved.

Supervised learning-Linear Regression

Bias and Variance

- Bias occurs when algo has limited flexibility to learn the true signal form a dataset
- Variance refers to an algo's sensitivity to specific sets of training data

Bias is the same as the mean square error (MSE). **Variance** shows how subject the model is to outliers, meaning those values that are far away from the **mean**.



Copyright © Nanyang Polytechnic.
All Rights Reserved.

Supervised learning-Logistic Regression

Logistic Regression is Classification algorithm used to predict discrete/categorical values(label)

Who will default on their credit card payment?

Credit card users

Make credit card transaction

Plot of monthly credit card balance and annual income

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Supervised learning-Logistic Regression

Classification performance measurement

Classification Accuracy

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

It tell how good the algorithm can predict but it does not give information where is the error that we can improve on

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Supervised learning-Logistic Regression

Confusion Matrix

Not a metric but look like one

Help to gain insight into the type of errors a model is making

Helps to understand some other metrics

	Predicted: 0	Predicted: 1
True: 0	126	13
True: 1	24	60

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Supervised learning-Logistic Regression

Precision, Recall, F1 score

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

	Predicted: 0	Predicted: 1
True: 0	TN = 126	FP = 13
True: 1	FN = 24	TP = 60

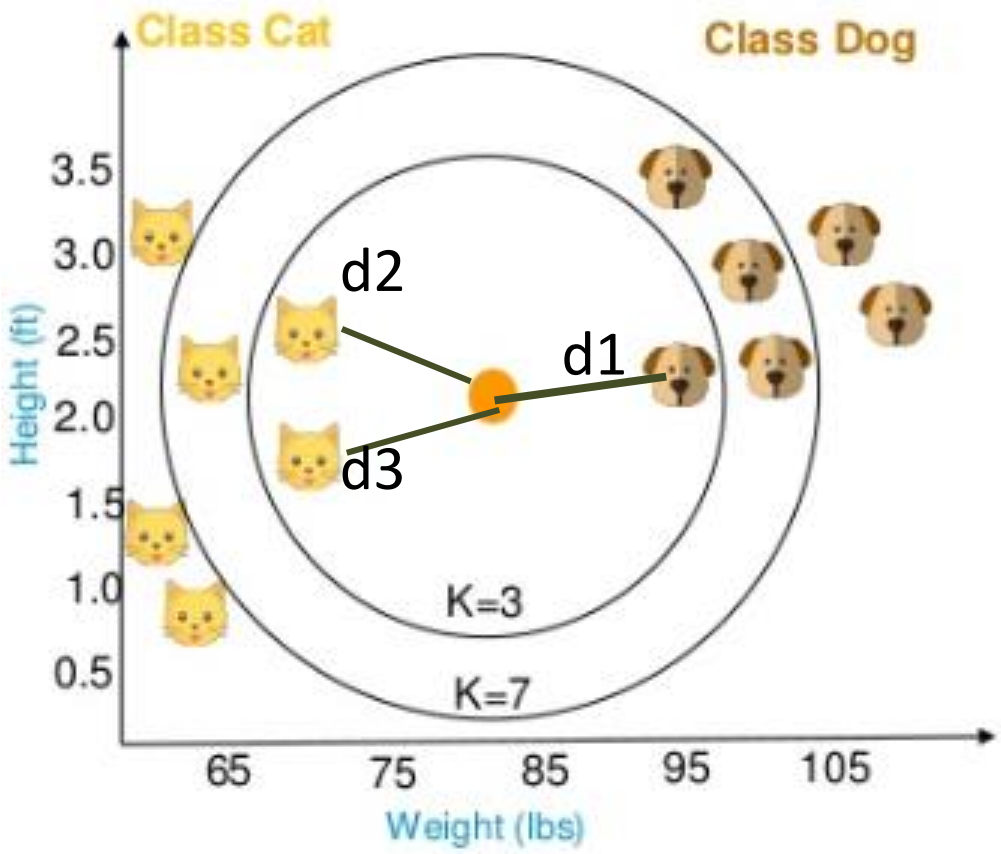
$$\text{F1 score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 * TP}{2 * TP + FP + FN}$$

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Supervised learning-K Nearest Neighbors(KNN)

K Nearest Neighbors(KNN) Approach

Choosing a K will define what class a new datapoint is assigned to:



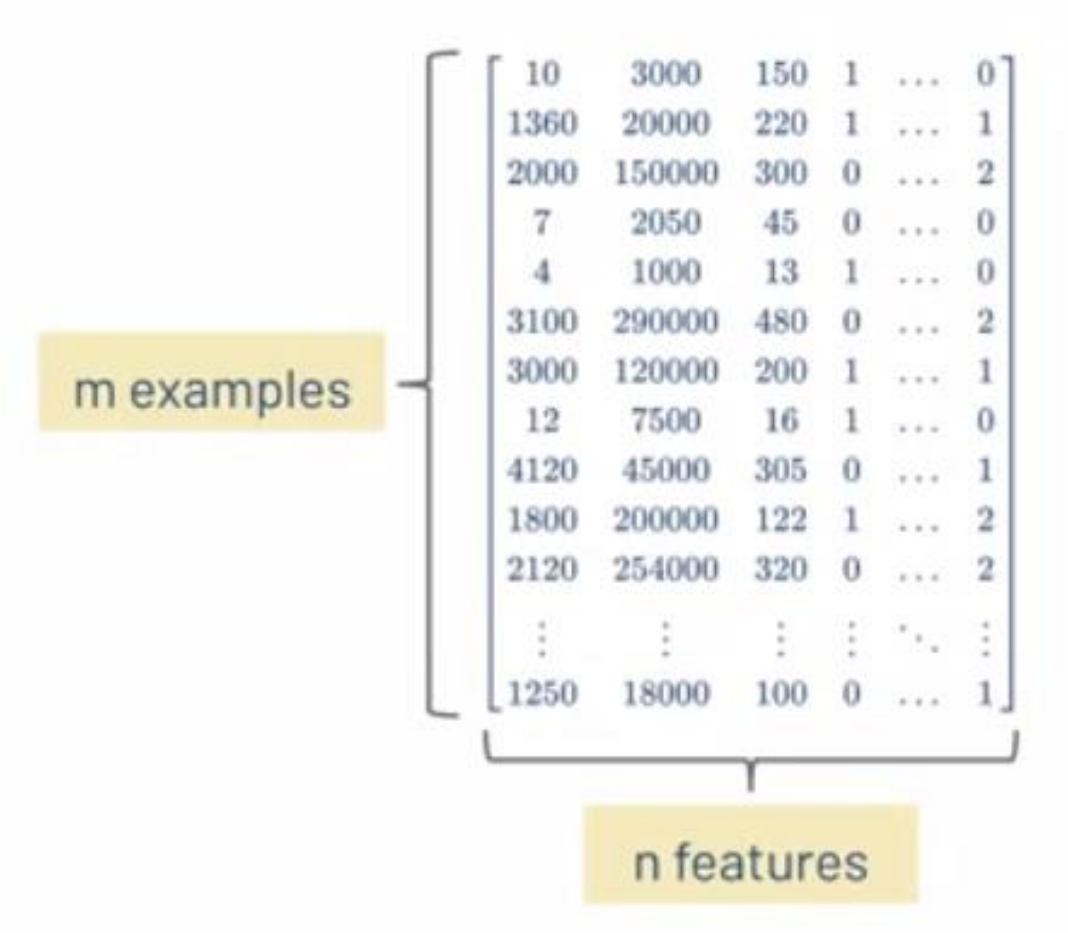
If K=3, the new data point belongs to class Cat

If K=7, the new data point belongs to class Dog

$$dist(d) = \sqrt{(x - a)^2 + (y - b)^2}$$

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Supervised learning-K Nearest Neighbors(KNN)



1. Calculate distances

```
for i in [1,m]:
    distance(x_new, x_i)
```

2. Pick k closest neighbours

$x_6 = 2$
$x_{11} = 2$
$x_{14} = 1$
$\vdots$
$x_{17} = 2$

k

3. Majority wins!

$$x_{new} = 2$$

4. Assign the cluster with the greatest number to be the unknown label

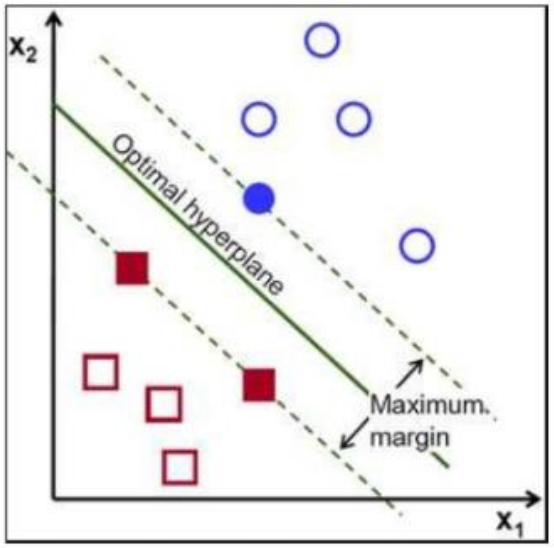
Copyright © Nanyang Polytechnic.
All Rights Reserved.

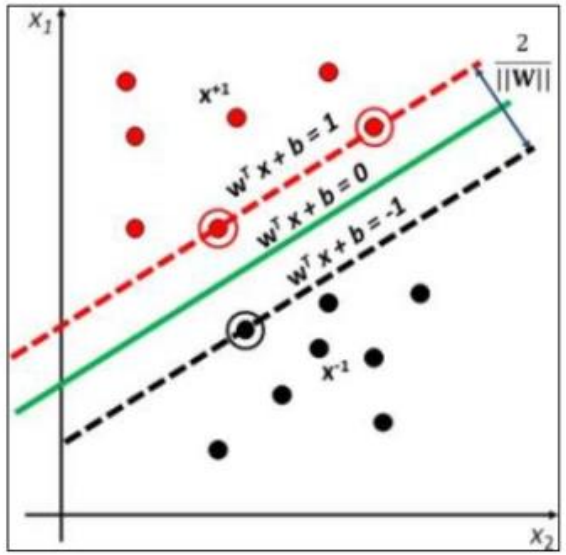
Supervised learning-Support Vector Machine

Support Vector Machine model associate learning algorithm that analyse data and recognize pattern

An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible

A main advantage of SVM is that it can be used for both classification and regression problems.

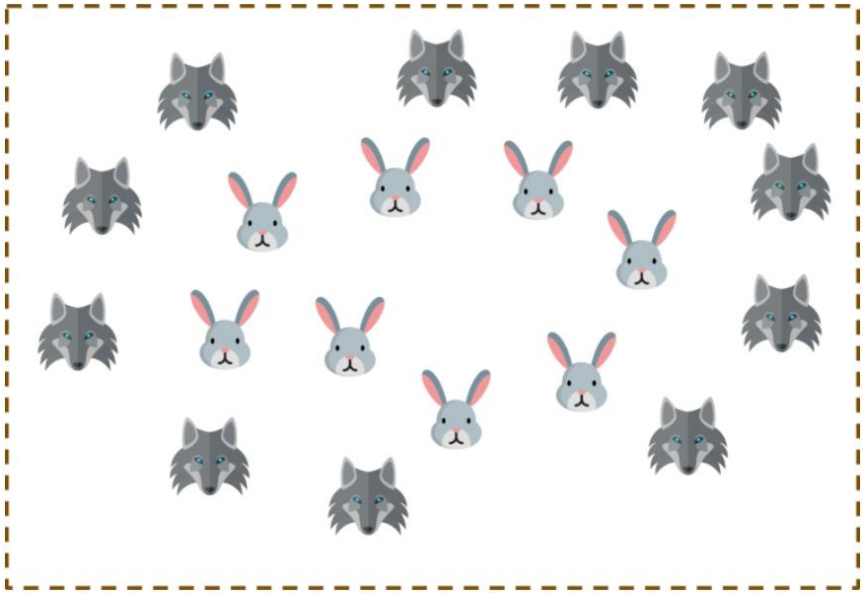




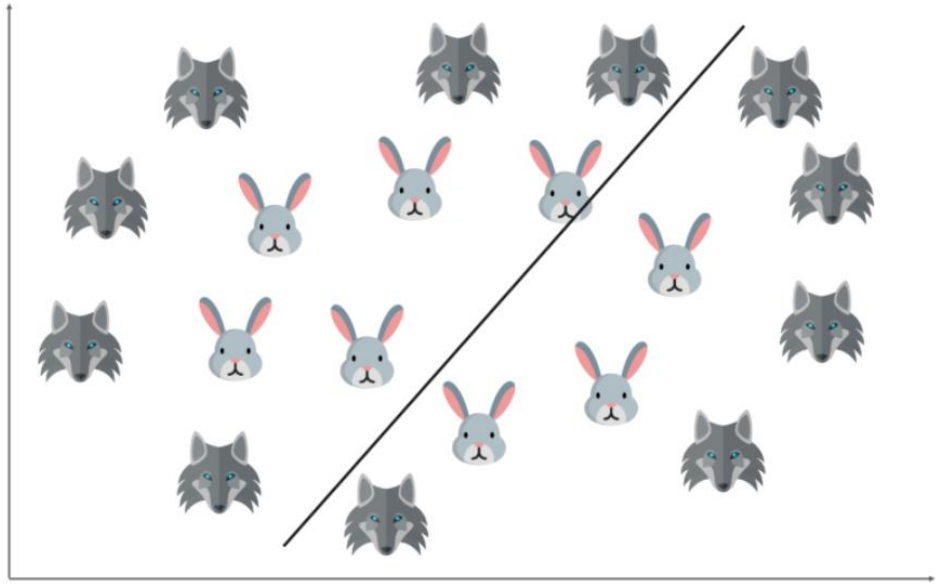
Copyright © Nanyang Polytechnic.
All Rights Reserved.

Supervised learning-Support Vector Machine

But what will you do if the data set is like this?



You possibly can't draw a hyperplane like this! It doesn't separate the two classes.

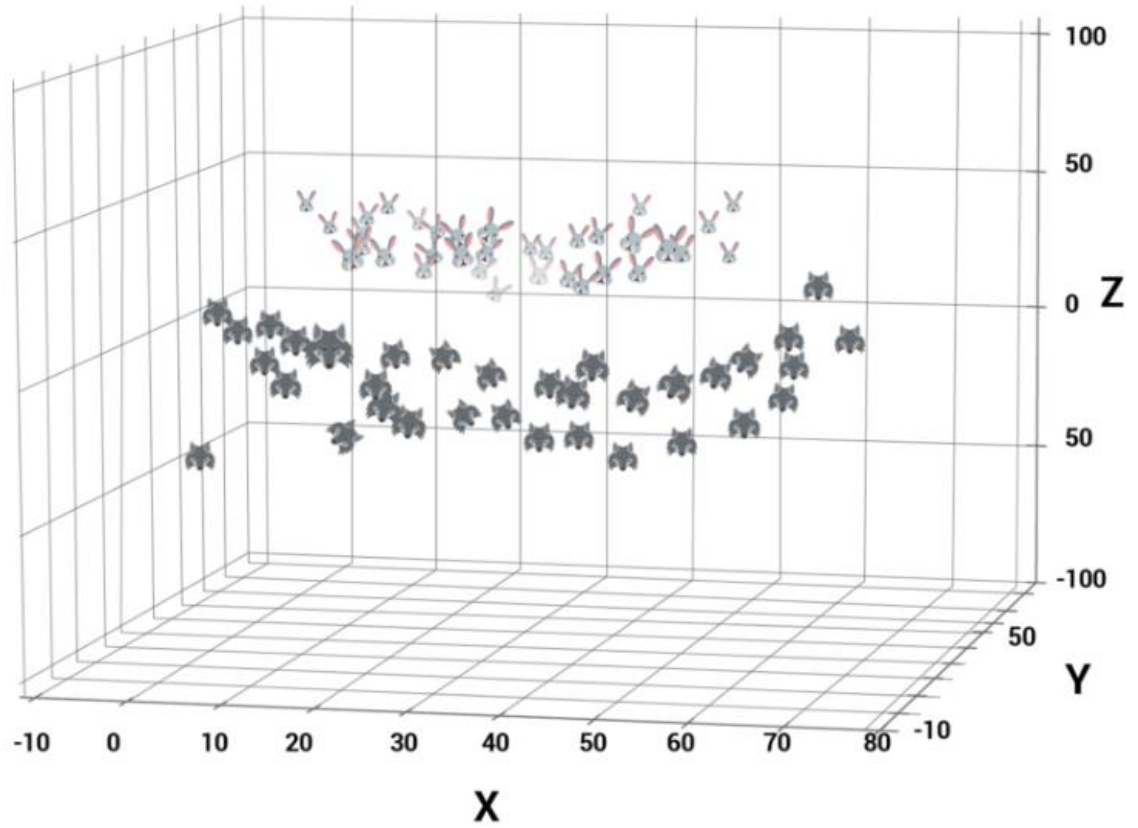


Copyright © Nanyang Polytechnic.
All Rights Reserved.

Supervised learning-Support Vector Machine

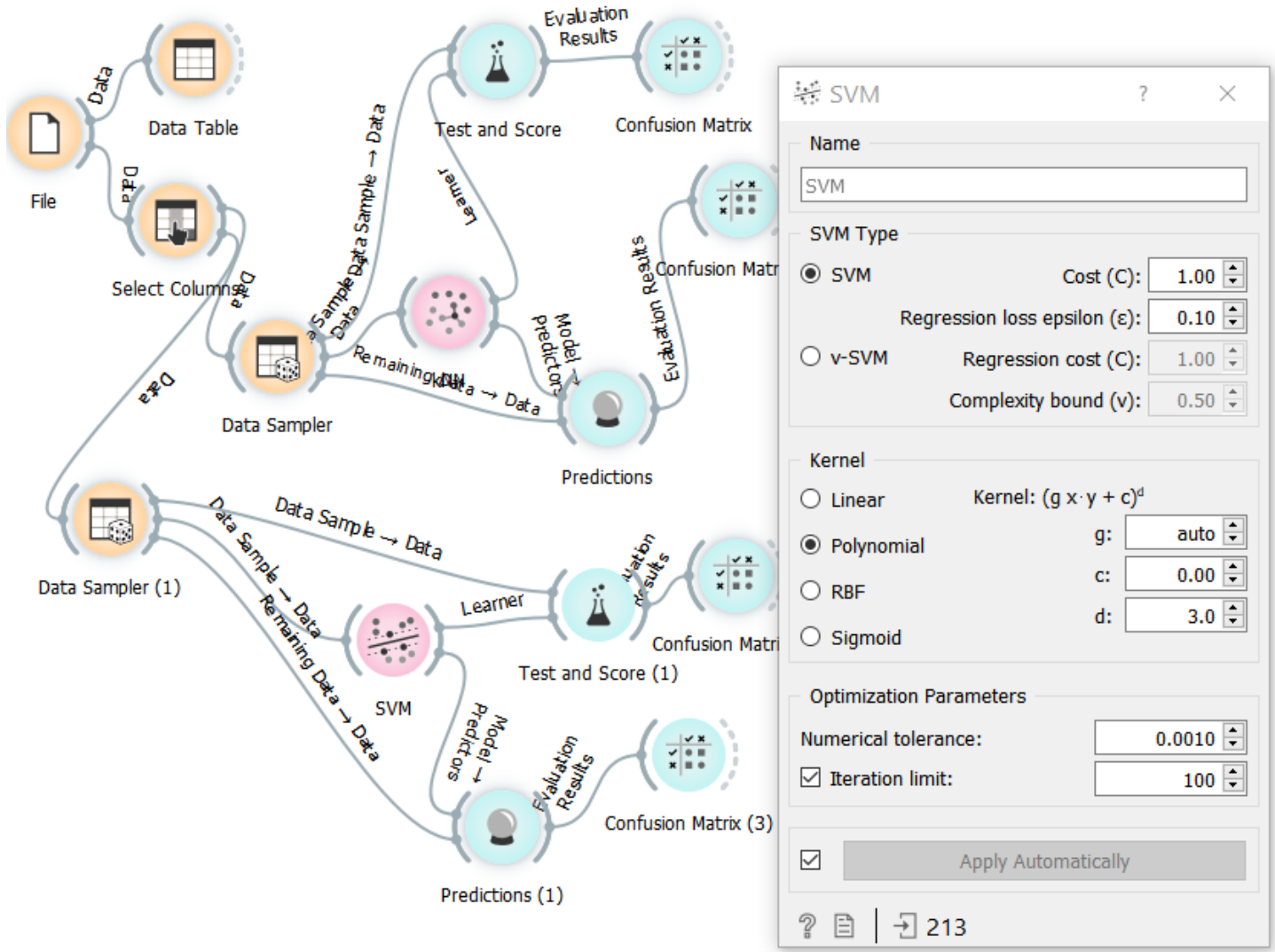
Nonlinear Support Vector Machine

a kernel can be used to transform data into another dimension that has a clear dividing margin between classes of data.



Copyright © Nanyang Polytechnic.
All Rights Reserved.

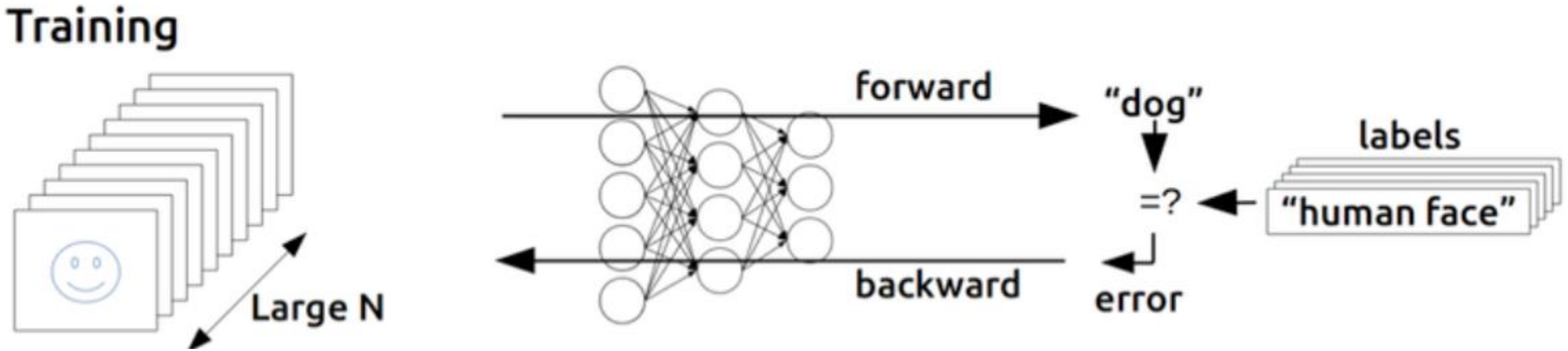
Supervised learning-Support Vector Machine



Copyright © Nanyang Polytechnic.
All Rights Reserved.

Supervised learning-Deep learning Neural Network

Our goal in using a neural net is to arrive at the point of least error as fast as possible. Each step for a neural network involves a guess, an error measurement and a slight update in its weights, an incremental adjustment to the coefficients, as it slowly learns to pay attention to the most important features.



Copyright © Nanyang Polytechnic.
All Rights Reserved.

Foundation of Artificial Intelligence

L 10 Unsupervised Learning

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Learning Outcomes

Understand what is unsupervised learning

Understand the algorithms used in the clustering

Apply the clustering algorithms

Apply dimensionality reduction

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Unsupervised learning

Application Examples

Clustering: A clustering problem is where you want to discover the inherent groupings in the data

- such as grouping customers by purchasing behavior

D**imensionality reduction**: technique that maximally informative **dimensions**, which finds a lower-**dimensional** representation of a dataset such that as much information as possible about the original data is preserved.

- Dimensionality reduction is an important method in Fingerprint recognition that often faces high-dimensional data

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Unsupervised learning-Clustering

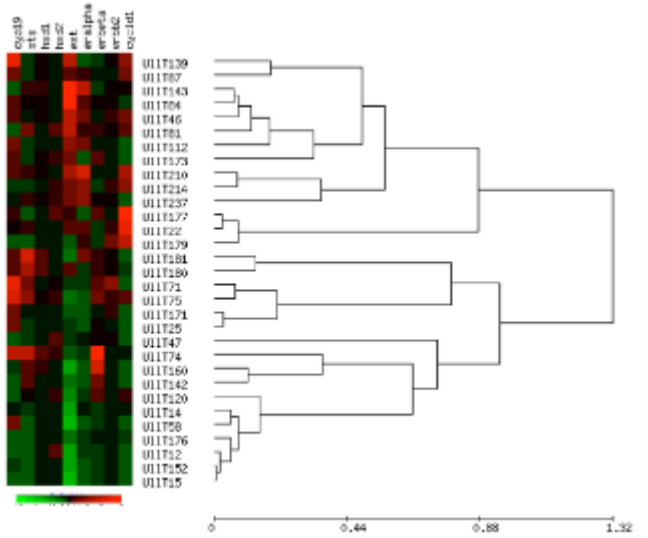
Hierarchical Clustering

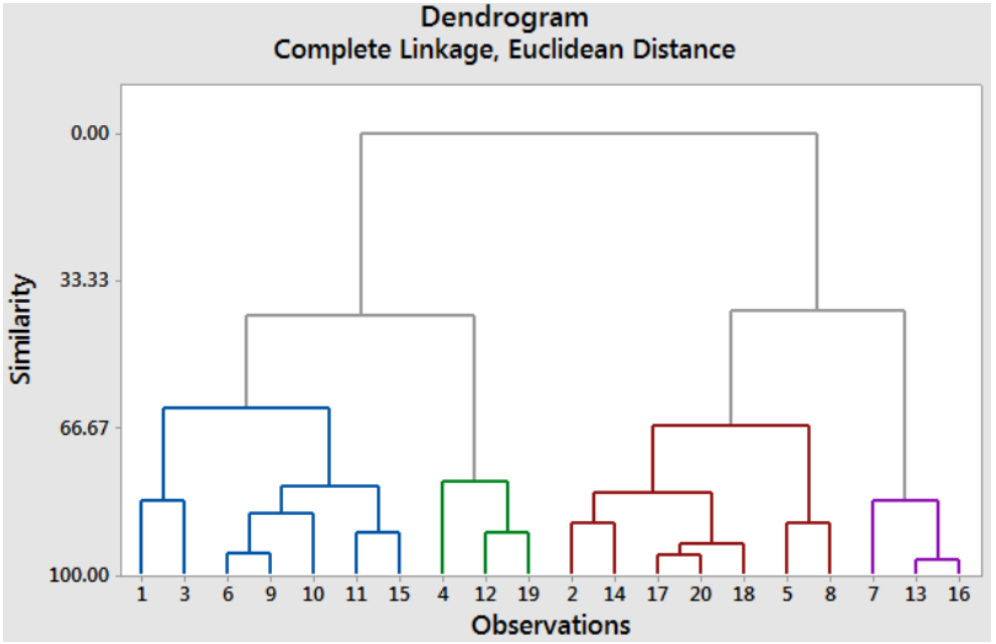
The basic notion behind this type of clustering is to create a hierarchy of clusters. As opposed to Partitioning Clustering, it does not require pre-definition of clusters upon which the model is to be built. There are two ways to perform Hierarchical Clustering. The first approach is a bottom-up approach, also known as Agglomerative Approach and the second approach is the Divisive Approach which moves hierarchy of clusters in a top-down approach. As a result of this type of clustering, we obtain a tree-like representation known as a dendogram.

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Unsupervised learning-Hierarchical Clustering

Hierarchical clustering is an alternative to prototype-based clustering algorithms. The main advantage of Hierarchical clustering is that we do not need to specify the number of clusters, it will find it by itself. In addition, it enables the plotting of dendrograms. Dendrograms are visualizations of a binary hierarchical clustering.





Copyright © Nanyang Polytechnic.
All Rights Reserved.

Unsupervised learning-Hierarchical Clustering

Hierarchical clustering, also known as **hierarchical cluster analysis**, is an **algorithm** that groups similar objects into groups called **clusters**. The endpoint is a set of **clusters**, where each **cluster** is distinct from each other **cluster**, and the objects within each **cluster** are broadly similar to each other.

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Unsupervised learning-Hierarchical Clustering

Distance

- Determine the similarity between two clusters .

Cosine Similarity $\frac{\sum x_i y_i}{\sqrt{\sum x_i^2}\sqrt{\sum y_i^2}}$

Manhattan Distance $\sum |x_i - y_i|$

Euclidean Distance $\sqrt{\sum |x_i - y_i|^2}$

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Unsupervised learning-Hierarchical Clustering

Advantages of Hierarchical Clustering

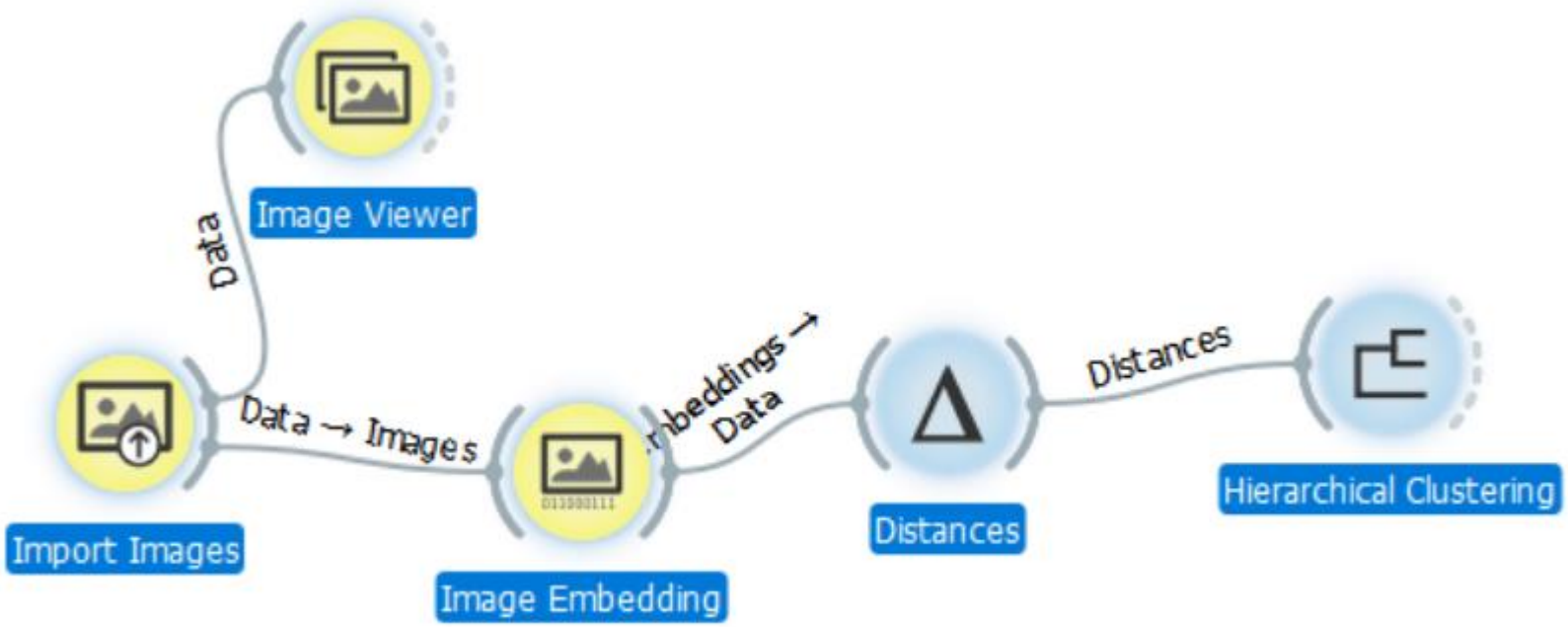
- The resulting hierarchical representations can be very informative.
- Dendrograms provide an interesting and informative way of visualization.
- They are specially powerful when the dataset contains real hierarchical relationships.

Disadvantages of Hierarchical Clustering

- They are very sensitive to outliers and, in their presence, the model performance decreases significantly.
- They are very expensive, computationally speaking.

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Unsupervised learning-Hierarchical Clustering



Copyright © Nanyang Polytechnic.
All Rights Reserved.

Unsupervised learning-Clustering

Partitioning Clustering

Partitional clustering (or partitioning clustering) are clustering methods used to classify observations, within a data set, into multiple groups based on their similarity. The algorithms require the analyst to specify the number of clusters to be generated. *K-Means Clustering* is the most popular type of partitioning clustering method.

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Unsupervised learning- K-Means Clustering

K-Means

Given k, the k-means algorithm works as follows:

1. Choose k (random) data points (seeds) to be the initial centroids, cluster centres
2. Assign each data point to the closest centroid
3. Re-compute the centroids using the current cluster memberships
4. If a convergence criterion is not met, repeat steps 2 and 3

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Unsupervised learning-K-Means Clustering

Advantages of K-means

- It is easy to implement k-means and identify unknown groups of data from complex data sets.
- The results are easy to interpret. It generates cluster descriptions in a form minimized to ease understanding of the data.
- Compared to using other clustering methods, a k-means clustering technique is fast and efficient in terms of its computational cost
- K-means is suitable for a large number of datasets and it’s computed much faster . It can also produce higher clusters.

Copyright © Nanyang Polytechnic.
All Rights Reserved.

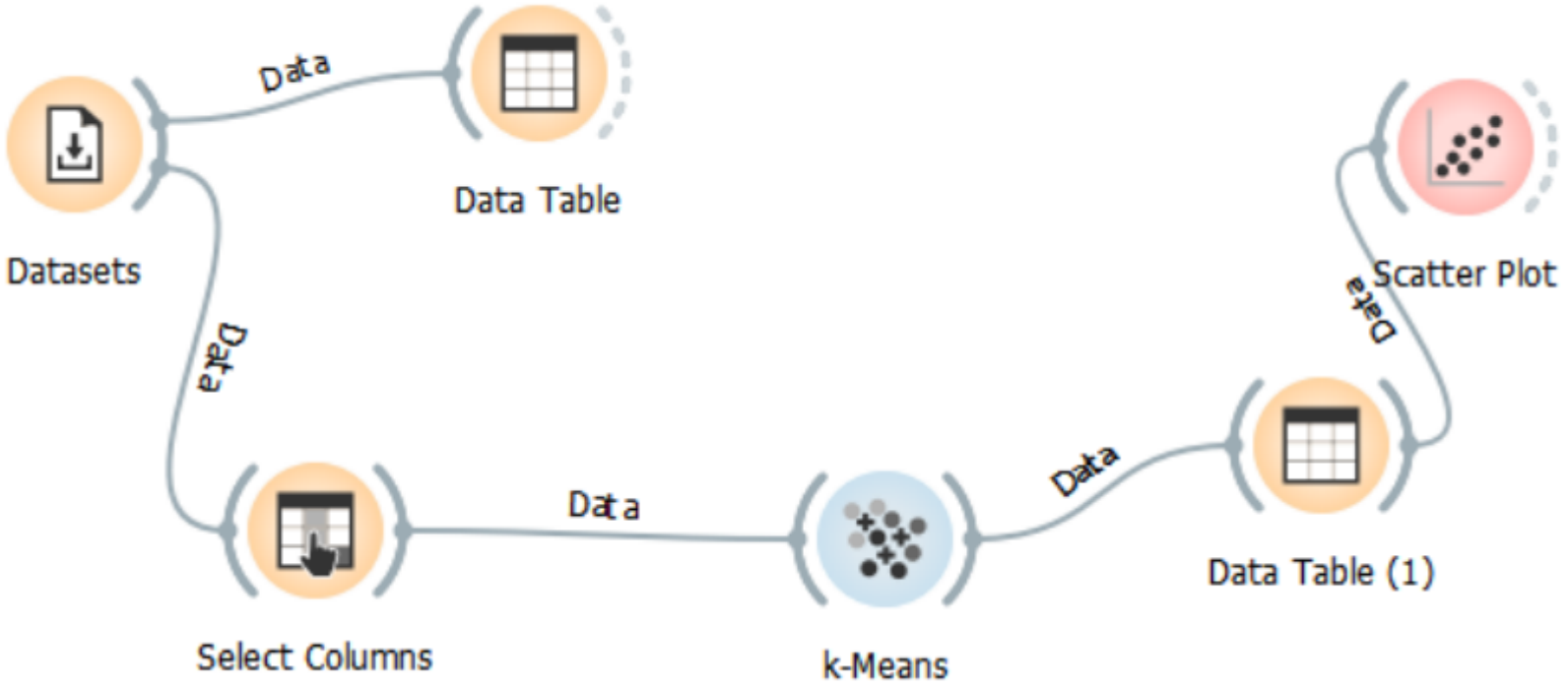
Unsupervised learning-K-Means Clustering

Disadvantages of K-means

- K-means doesn’t allow development of an optimal set of clusters and for effective results, you should decide on the clusters before.
- K-means clustering gives varying results on different runs of an algorithm. A random choice of cluster patterns yields different clustering results resulting in inconsistency.
- It is difficult to predict the k-values or the number of clusters. It is also difficult to compare the quality of the produced clusters.

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Unsupervised learning-K-Means Clustering



Copyright © Nanyang Polytechnic.
All Rights Reserved.

Unsupervised learning-Dimensionality Reduction

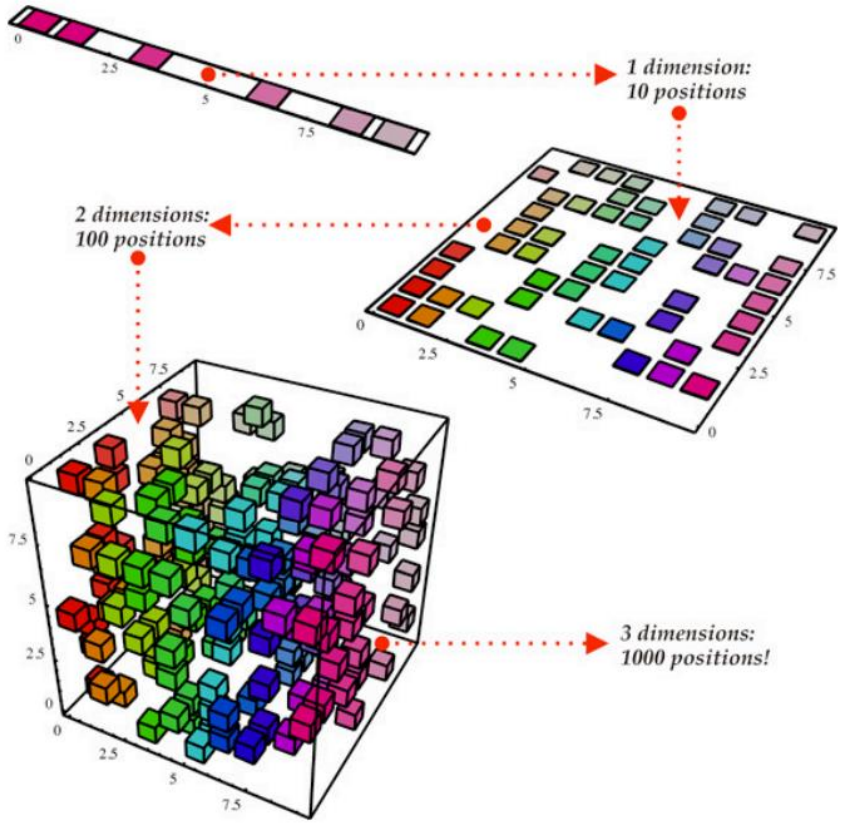
What is Dimensionality Reduction?

In machine learning classification problems, there are often too many factors on the basis of which the final classification is done.

The higher the number of factors(features), the harder it gets to visualize the training set and then work on it.

Sometimes, most of these features are correlated, and hence redundant. This is where dimensionality reduction algorithms come into play.

Dimensionality reduction is the process of reducing the number of random variables under consideration, by obtaining a set of principal variables.



Copyright © Nanyang Polytechnic.
All Rights Reserved.

Unsupervised learning-Dimensionality Reduction

There are two components of dimensionality reduction:

- Feature selection: In this, we try to find a subset of the original set of variables, or features, to get a smaller subset which can be used to model the problem. It usually involves ways:

  Variance Thresholds
  Correlation Thresholds

- Feature extraction: This reduces the data in a high dimensional space to a lower dimension space, i.e. a space with lesser no. of dimensions.

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Unsupervised learning-Dimensionality Reduction

- Feature extraction

Principal Component Analysis (PCA)

Principal component analysis (PCA) is an unsupervised algorithm that creates linear combinations of the original features. The new features are orthogonal, which means that they are uncorrelated. Furthermore, they are ranked in order of their "explained variance." The *first principal component* (PC1) explains the most variance in your dataset, PC2 explains the second-most variance, and so on.

Therefore, you can reduce dimensionality by limiting the number of principal components to keep based on cumulative explained variance. For example, you might decide to keep only as many principal components as needed to reach a cumulative explained variance of 90%.

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Unsupervised learning-Dimensionality Reduction

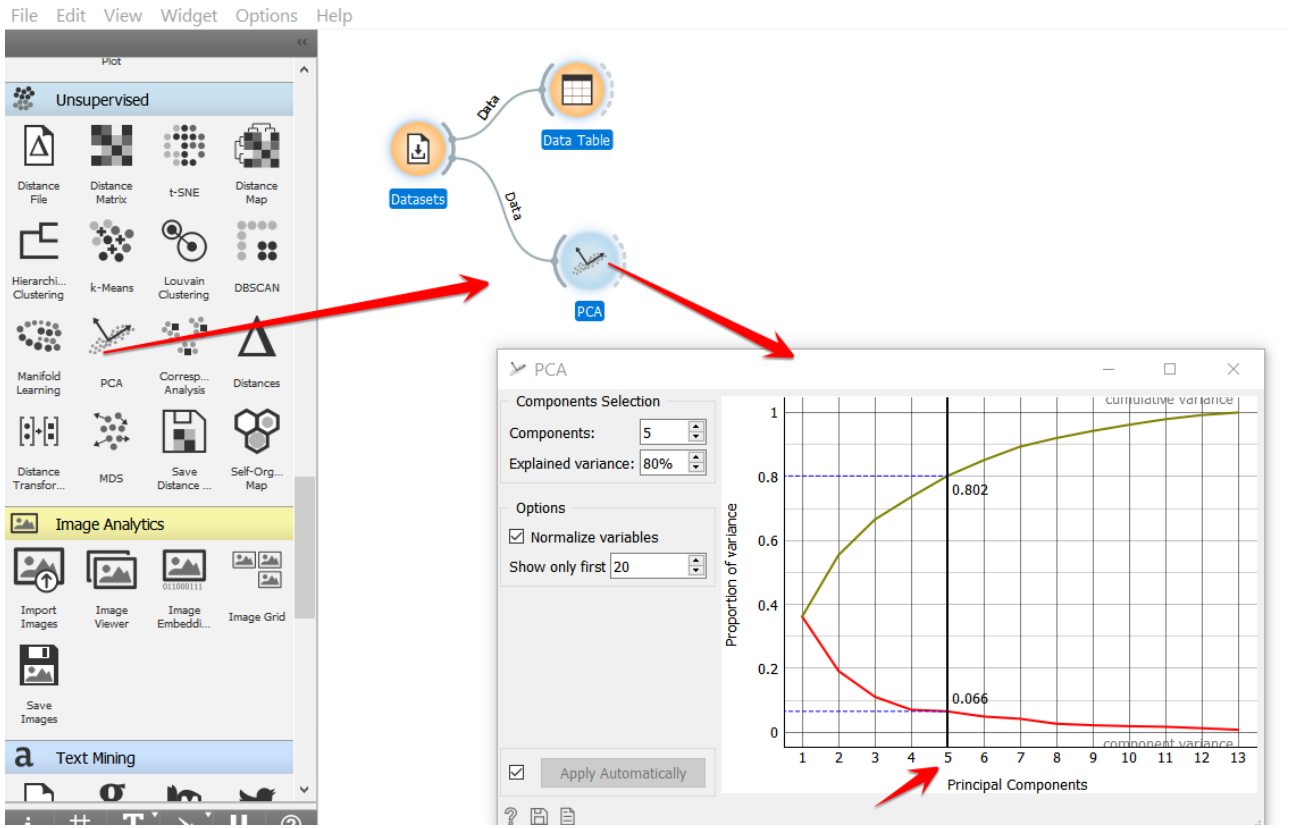
- Feature extraction

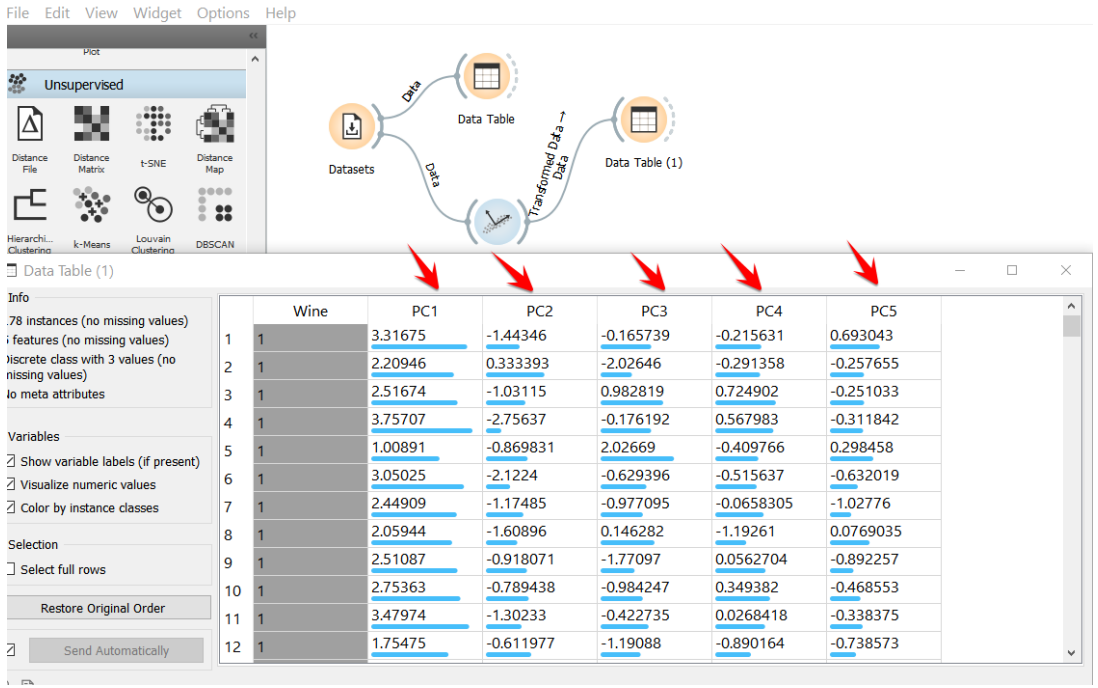
Principal Component Analysis (PCA)

- Strengths: PCA is a versatile technique that works well in practice. It's fast and simple to implement, which means you can easily test algorithms with and without PCA to compare performance. In addition, PCA offers several variations and extensions (i.e. kernel PCA, sparse PCA, etc.) to tackle specific problems.
- Weaknesses: The new principal components are not interpretable, which may be a deal-breaker in some settings. In addition, you must still manually set or tune a threshold for cumulative explained variance.

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Unsupervised learning-Dimensionality Reduction





Copyright © Nanyang Polytechnic.
All Rights Reserved.

Foundation of Artificial Intelligence

L11 Reinforcement Learning

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Learning Outcomes

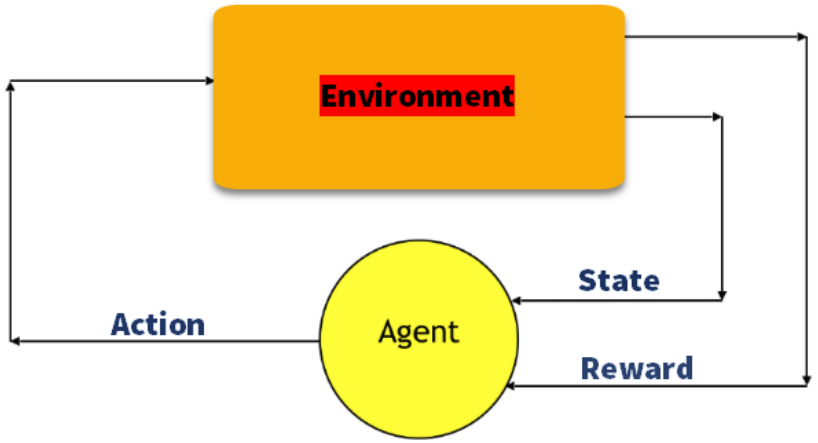
- Understand what is a reinforcement learning
- Understand how to apply reinforcement learning

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Reinforcement Learning

Markov Decision Process

The following parameters are used to get a solution:



- Set of actions- A
- Set of states -S
- Reward- R
- Policy- n
- Value- V

The mathematical approach for mapping a solution in reinforcement Learning is recon as a Markov Decision Process or (MDP).

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Reinforcement Learning

Q-Learning

Q learning is a value-based method of supplying information to inform which action an agent should take.

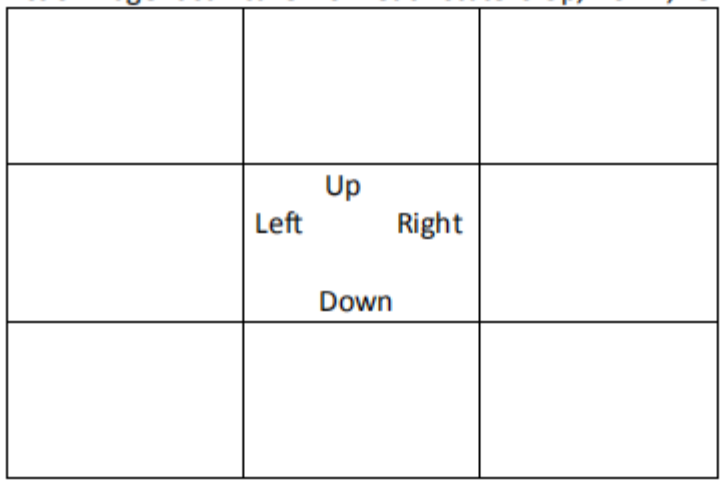
Copyright © Nanyang Polytechnic.
All Rights Reserved.

Reinforcement Learning

1 	2	3	4	5
10	9	8 	7	6
11	12 	13 	14	15
20	19	18	17	16
21	22	23	24	25

states

Action - agent can take from each state is Up, Down, Left or Right



Rewards are
-If agent encounter Green triangle -1
-If agent encounter white space 0
-if agent encounter Red circle 100

Rewards Table

	action			
State	Up	Down	Left	Right
1	-1	0	-1	0
2				
3				

9	0	-1	0	-1

14	0	0	100	0

25				

Q Table

	action			
State	Up	Down	Left	Right
1				
2				
3				
....				
25				

Q[state, action] = Q[state, action] + lr * (reward + gamma * np.max(Q[new_state, :]) — Q[state, action])

Copyright © Nanyang Polytechnic.
All Rights Reserved.

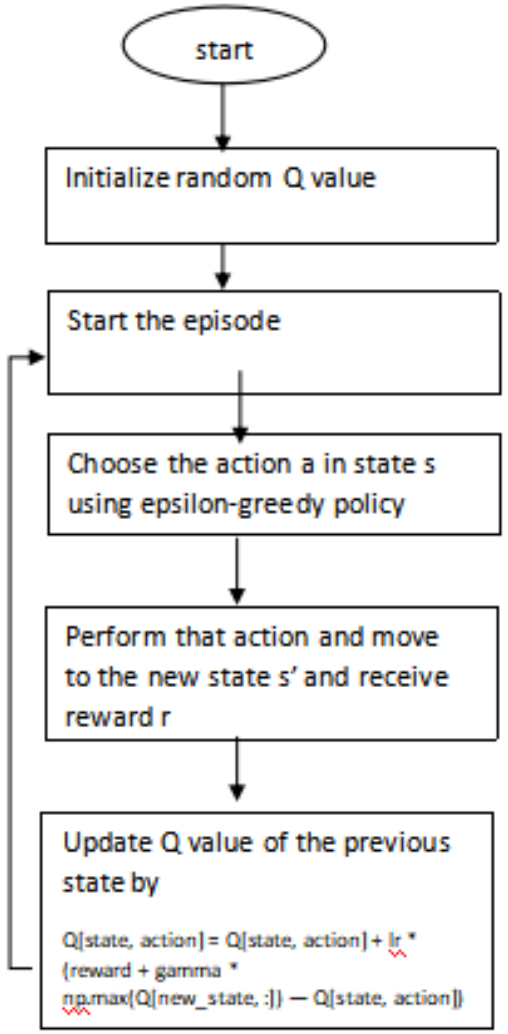
Reinforcement Learning

Q[state, action] = Q[state, action] + lr * (reward + gamma * np.max(Q[new_state, :]) — Q[state, action])

Learning Rate: lr or learning rate, often referred to as *alpha* or α, can simply be defined as how much you accept the new value vs the old value.

Gamma: gamma or *γ* is a discount factor. It's used to balance immediate and future reward..

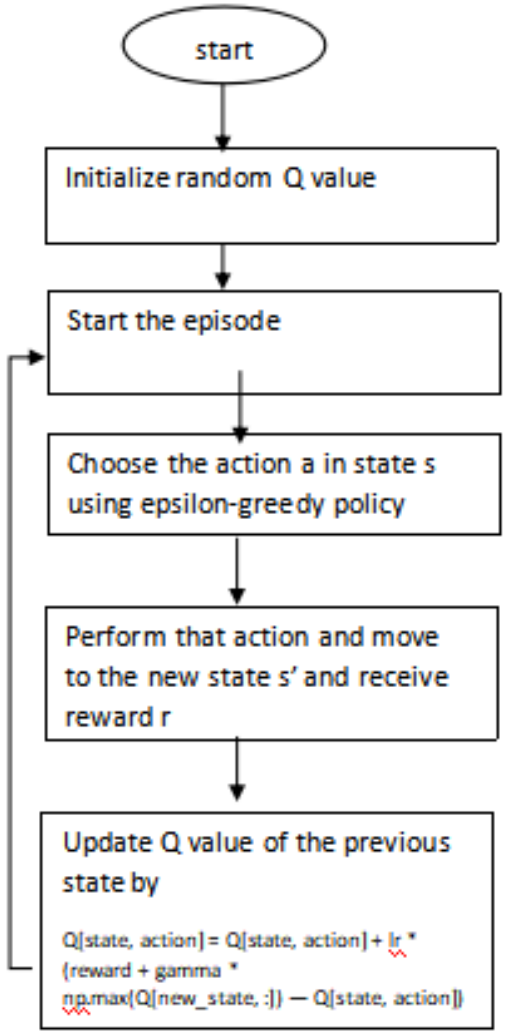
Reward: reward is the value received after completing a certain action at a given state.



Copyright © Nanyang Polytechnic.
All Rights Reserved.

Reinforcement Learning

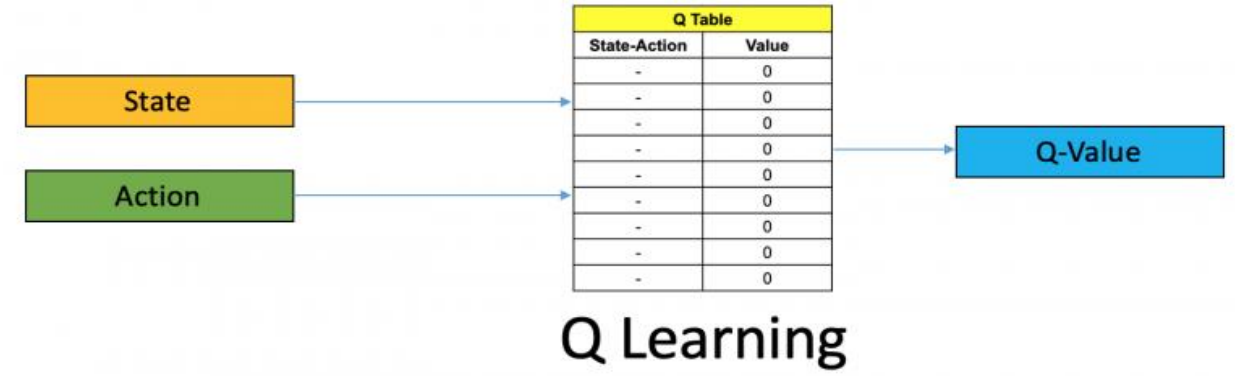
Reinforcement learning is a type of machine learning wherein the software agent learns to perform certain actions from it state in an environment which lead it to maximum reward. It does so by exploration and exploitation of knowledge it learns by repeated trials of maximizing the reward.

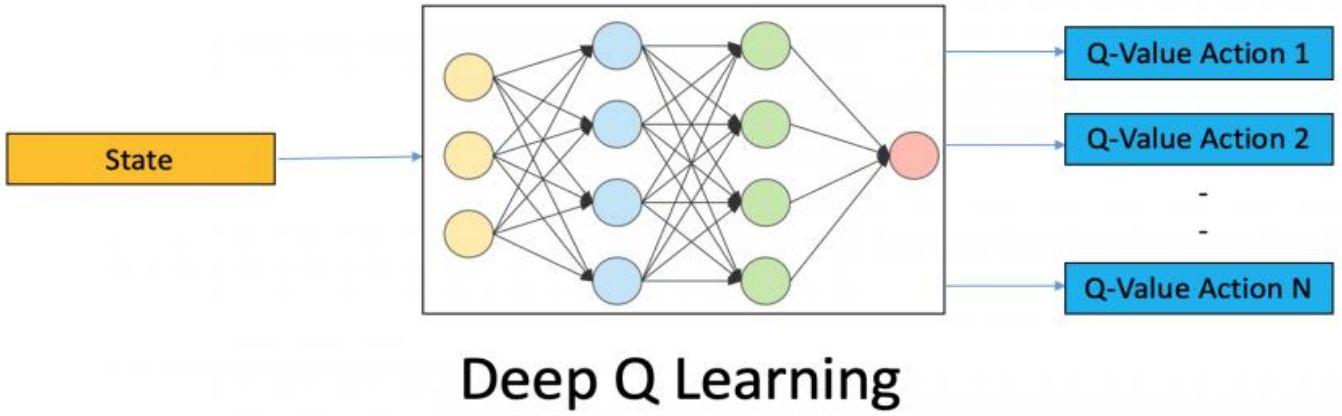


Copyright © Nanyang Polytechnic.
All Rights Reserved.

Reinforcement Learning

Q learning vs Deep Q learning





Copyright © Nanyang Polytechnic.
All Rights Reserved.

Reinforcement Learning vs. Supervised Learning

Parameters	Reinforcement Learning	Supervised Learning
Decision style	reinforcement learning helps you to take your decisions sequentially.	In this method, a decision is made on the input given at the beginning.
Works on	Works on interacting with the environment.	Works on examples or given sample data.
Dependency on decision	In RL method learning decision is dependent. Therefore, you should give labels to all the dependent decisions.	Supervised learning the decisions which are independent of each other, so labels are given for every decision.
Best suited	Supports and work better in AI, where human interaction is prevalent.	It is mostly operated with an interactive software system or applications.
Example	Chess game	Object recognition

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Foundation of Artificial Intelligence

L 12 Deep Learning

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Learning Outcomes

Understand what is Deep learning

Understand the algorithms used in the deep learning

Apply the Deep Learning algorithms

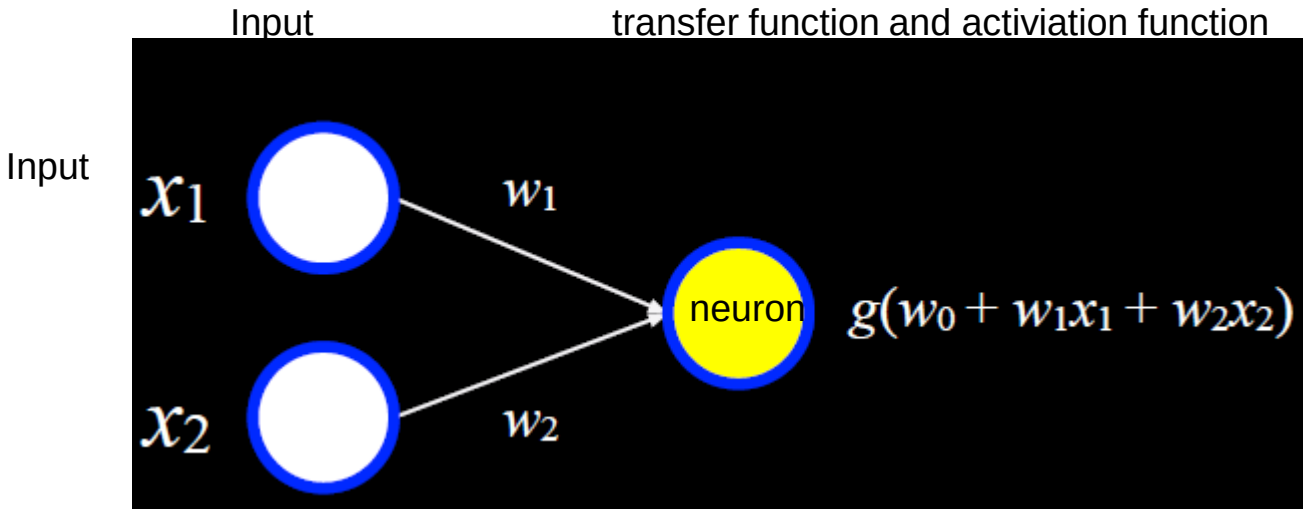
Copyright © Nanyang Polytechnic.
All Rights Reserved.

Deep Learning-Artificial neural network

What is artificial neural network?

Model mathematical function from inputs to outputs based on the structure and parameters of the network.

Allows for learning the network's parameters based on data



Copyright © Nanyang Polytechnic.
All Rights Reserved.

Deep Learning-Artificial neural network

Activation function

It ensure non-linearity in the model

It takes the "weighted sum input plus the bias" as the input to the function and decides whether it should be fired or not.

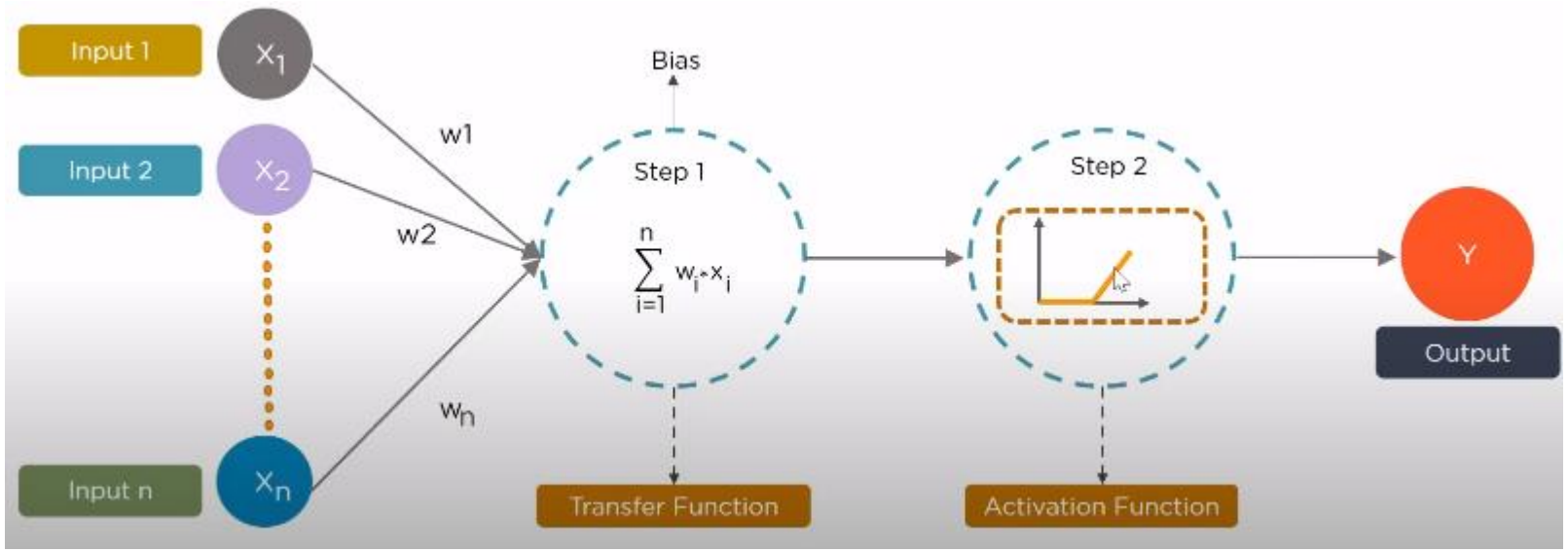


Image source: simplelearn

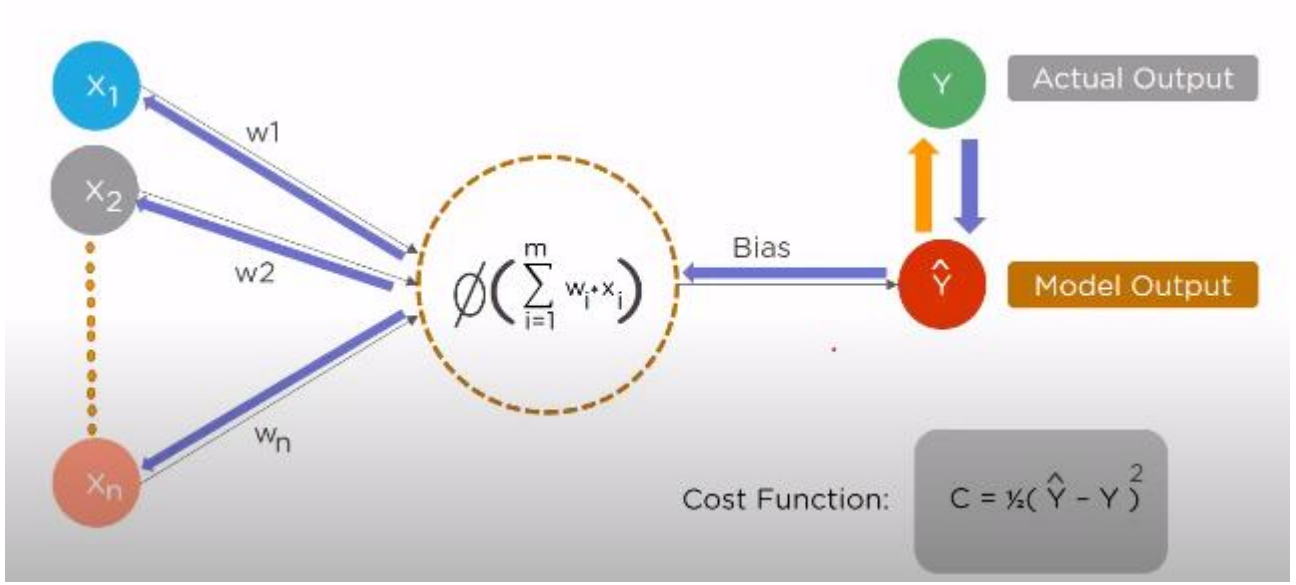
Sigmoid
ReLU
Tanh

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Deep Learning-Example

Deep Learning – Artificial neural network example

How ANN train the model- Use cost function to understand the training result. Also known as error rate.

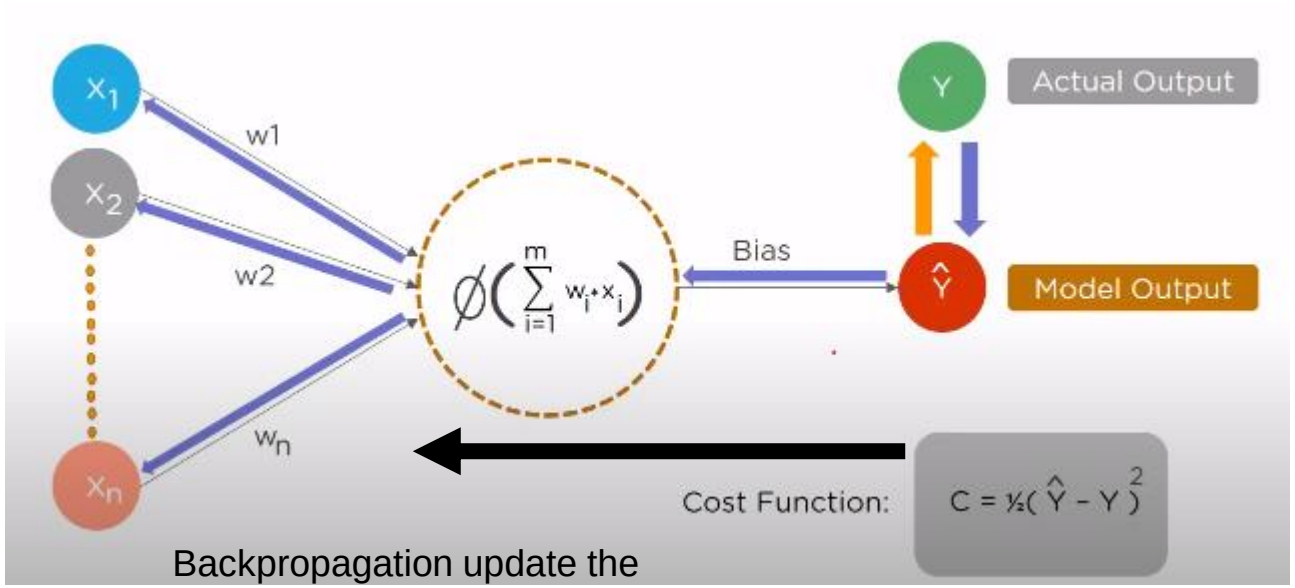


Copyright © Nanyang Polytechnic.
All Rights Reserved.

Deep Learning-Example

Deep Learning – Artificial neural network example

After training, it use the error rate to do the Backpropagation to improve the performance of the network



Backpropagation update the Weights and Bias

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Deep Learning-Example

Deep Learning – Artificial neural network example

How to optimize the cost function-to achieve smallest error?

The amount that the weights are updated during **training** is referred to as the step size or the "**learning rate**." Specifically, the **learning rate** is a configurable hyperparameter used in the **training** of **neural networks** that has a small positive value, often in the range between 0.0 and 1.0.

The bigger the learning rate the faster will be the training.

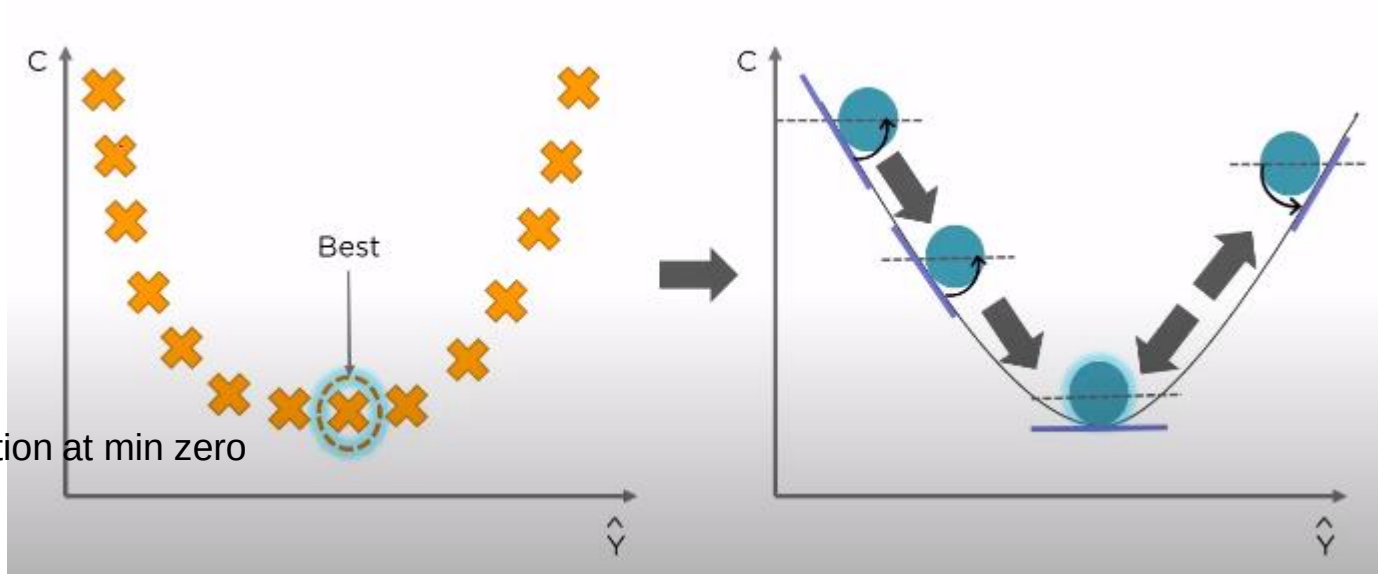
Copyright © Nanyang Polytechnic.
All Rights Reserved.

Deep Learning-Example

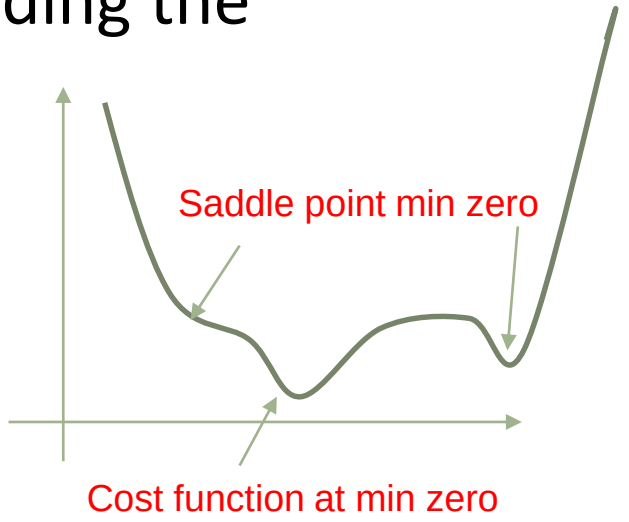
Deep Learning – Artificial neural network example

How to optimize the cost function-to achieve smallest error?

Gradient Descent is an **optimization algorithms** for finding the minimum value of the cost function(error rate).



Cost function at min zero



Copyright © Nanyang Polytechnic.
All Rights Reserved.

Deep Learning-Advantages

- it robust enough to understand and use novel data, but most data scientists have learned to control the learning to focus on what's important to them. Deep learning takes advantage of this by allowing you to control the learning, but not the statistical modeling
- it has a high dimensionality. This means that we can create more learning models by adding more layers to our neural network.
- it can adapt automatically to all data, but it makes for a nice alternative to traditional machine learning that relies on human expertise for feature extraction
- it can be used in datasets that are too large, complex and repetitive for traditional computer systems.

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Deep Learning- Disadvantages

- it tends to learn on its own, and it's also hard to see the evolution of a system in time.
- it requires huge data sets in order to train.
- it is computationally very expensive, requiring a large amount of memory and computational resources.
- it requires much larger datasets with many more features. As a result, it takes longer to train *the algorithm and it takes* more memory for it to work with the data.
- it requires very advanced optimization techniques, and these should have been incorporated to obtain good results.

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Foundation of Artificial Intelligence

L13 AI System Development

Copyright © Nanyang Polytechnic.
All Rights Reserved.

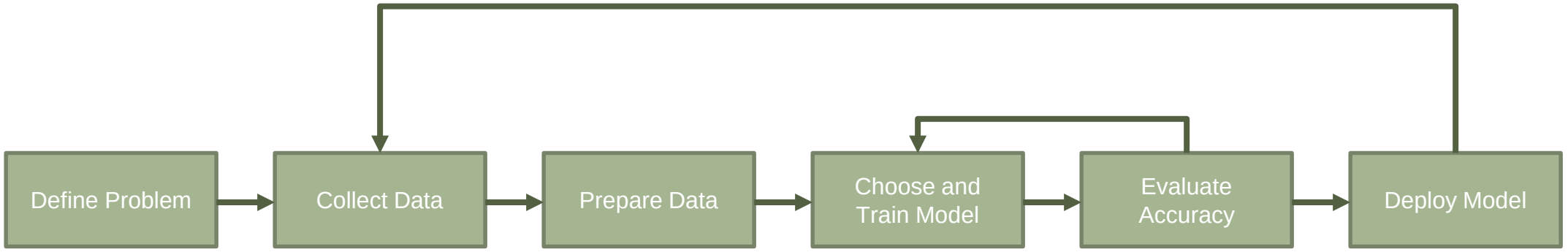
Learning Outcomes

- Understand AI project workflow
- Able to apply it in the project development

Copyright © Nanyang Polytechnic.
All Rights Reserved.

AI Project Workflow

- As with any project, Artificial Intelligence/Machine Learning/Deep Learning project undergoes a few stages:



Copyright © Nanyang Polytechnic.
All Rights Reserved.

Define Problem Supervised Learning - Classification

Supervised learning-Classification/Regression
Unsupervised learning-Clustering/ Dimension Reduction
Reinforcement Learning-Q Learning/Deep Q Learning
Deep Learning- Classification/Regression

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Data Collection-Data Sources

- Existing Digital Data:
  - Enterprise Software Systems (all historical transactions, financial, operational, etc)
  - User Generated Data (Social Media, etc)
  - Internet of Things (sensor readings, videos, etc)
  - Open Source Data (ImageNet, Kaggle)
  - Internet Sources (news, facts, unstructured data)

- Other Data Sources:
  - Surveys (need transcription)
  - Hardcopy data (scanned and require transcription / digitization)

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Why the Need to Prepare?

- Format:
  - AI Models Require Data in Specific Format
  - AI Models are generally mathematical computations
    data needs to be converted into numbers

- Skewing:
  - Some data are skewed in the raw form
    For example, salary is in the positive scale between 0 to 5,000,000, age is always between in the 0 – 100 scale, the scale of different data can cause inaccuracies)

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Why the Need to Prepare?

- Text:
  - Tokenization aka Word Splitting
    especially for Chinese/Japanese/Korean where there's no natural word break like in English
  - Normalization
    convert everything to lowercase
  - Lemmatization
    language-specific conjugations that do not carry important meaning, eg. Run, ran, running may not be meaning to your application
  - Stop-Word Removal
    Remove words like the, a, an that do not carry important meaning
  - Handling Numbers
    How to treat the word "hundred" and "100"

Copyright © Nanyang Polytechnic.
All Rights Reserved.

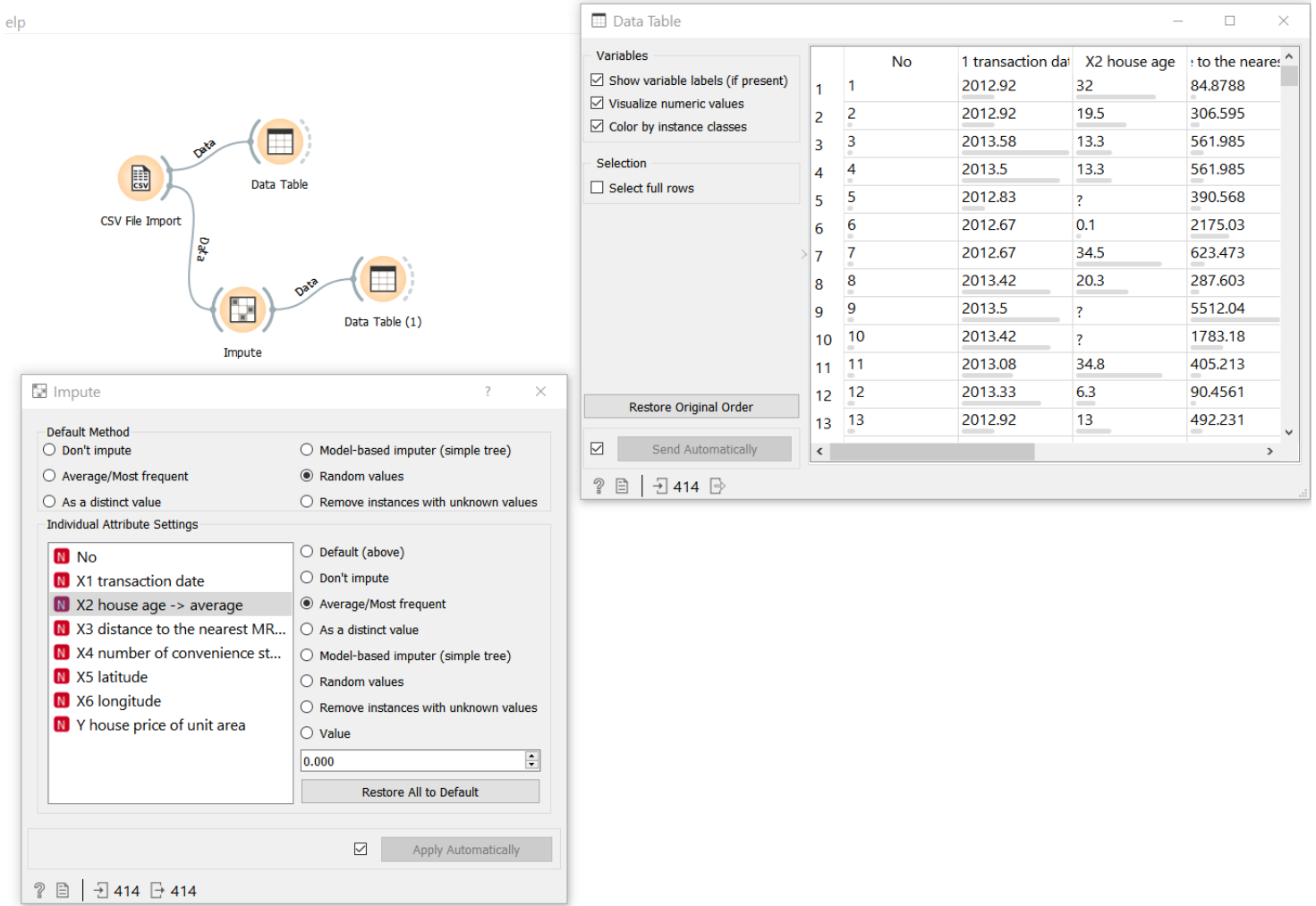
Why the Need to Prepare?

- Images:
  - Converting JPG/PNGs into R,G,B Bitmaps
    JPG/PNGs are compressed representation of images, and has to expanded to the RGB bitmap forms
  - Resolution
    Most AI models work on smaller resolution eg. 224x224, so image has to be scaled down
  - RGB Color Scaling
    Bitmaps representing the RGB components of the image in intensity values of 0 (darkest) to 255 (brightest). Scaling helps to set their intensity values to -1.0 (darkest) to 1.0 (brightest).

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Why the Need to Prepare?

- Data-miss/duplication/corrupted



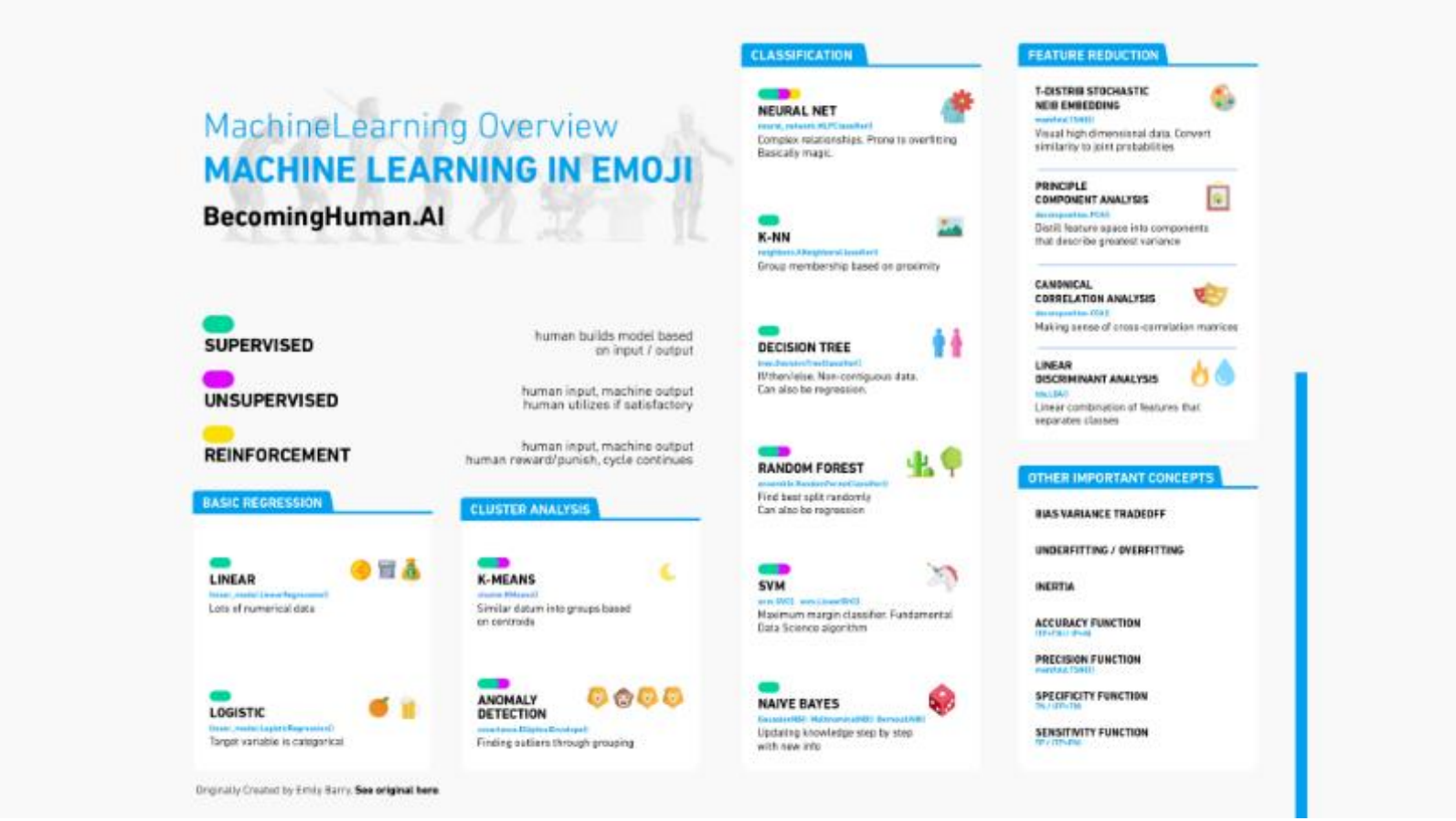
Copyright © Nanyang Polytechnic.
All Rights Reserved.

Why the Need to Prepare?

- Training Dataset and Test Dataset:
  - Training dataset
    - usually a large set of data
    - used to train the AI model
  - Test dataset
    - a smaller set
    - used to test the accuracy of the AI model.
    - It must not be seen in the training dataset
  - Typical ratio of **training : test** about **80% : 20%**

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Machine Learning Algorithms



https://becominghuman.ai/cheat-sheets-for-ai-neural-networks-machine-learning-deep-learning-big-data-678c51b4b463

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Accuracy Measurements

Classification -Confusion Matrix

- **Precision:**

  (TP)/(TP+FP) = (100)/(100+10) = 0.91

- **Recall:**

  (TP)/(TP+FN) = (100)/(100+5)= 0.95

- **F1 Score:**

  (2x TP)/(2xTP+FP+FN))=(2x100)/(2x100+10+5))=0.93

n=165	**Predicted: NO**	**Predicted: YES**	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Accuracy Measurements

Regression

Mean Absolute Error

$$\mathrm{MAE} = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j|$$

Mean Square Error

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}$$

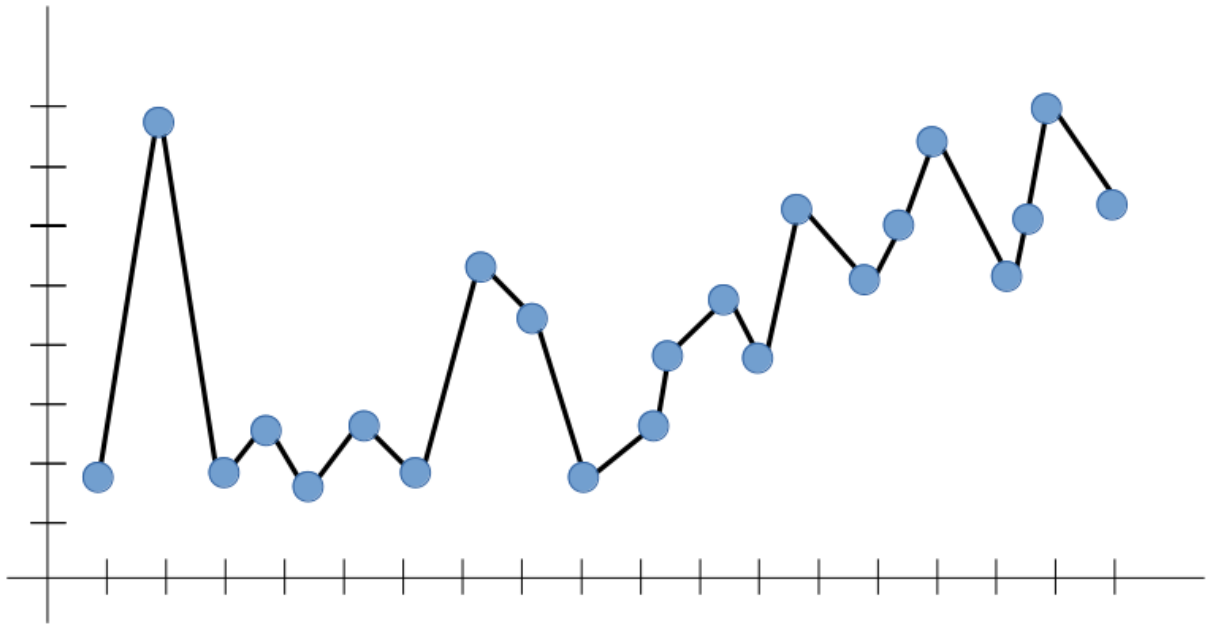
Root Mean Square Error

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Interpreting the test result

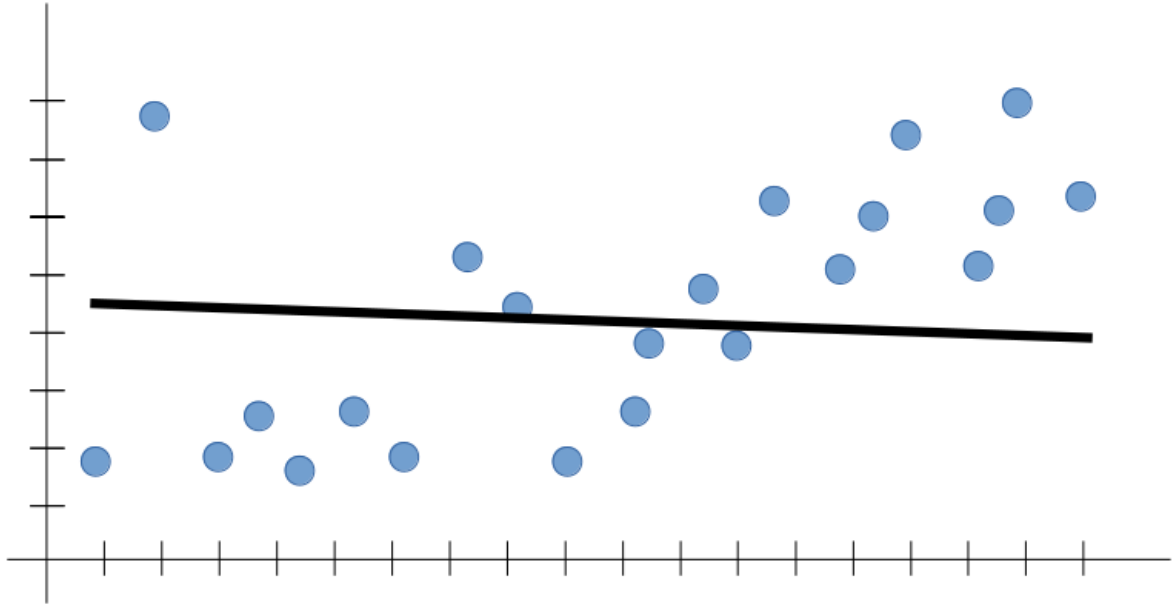
Overfitting is the case where the overall cost is really small, but the generalization of the model is unreliable. This is due to the model learning “too much” from the training data set.



Copyright © Nanyang Polytechnic.
All Rights Reserved.

Interpreting the test result

Underfitting is the case where the model has " not learned enough" from the training data, resulting in low generalization and unreliable predictions.



Copyright © Nanyang Polytechnic.
All Rights Reserved.

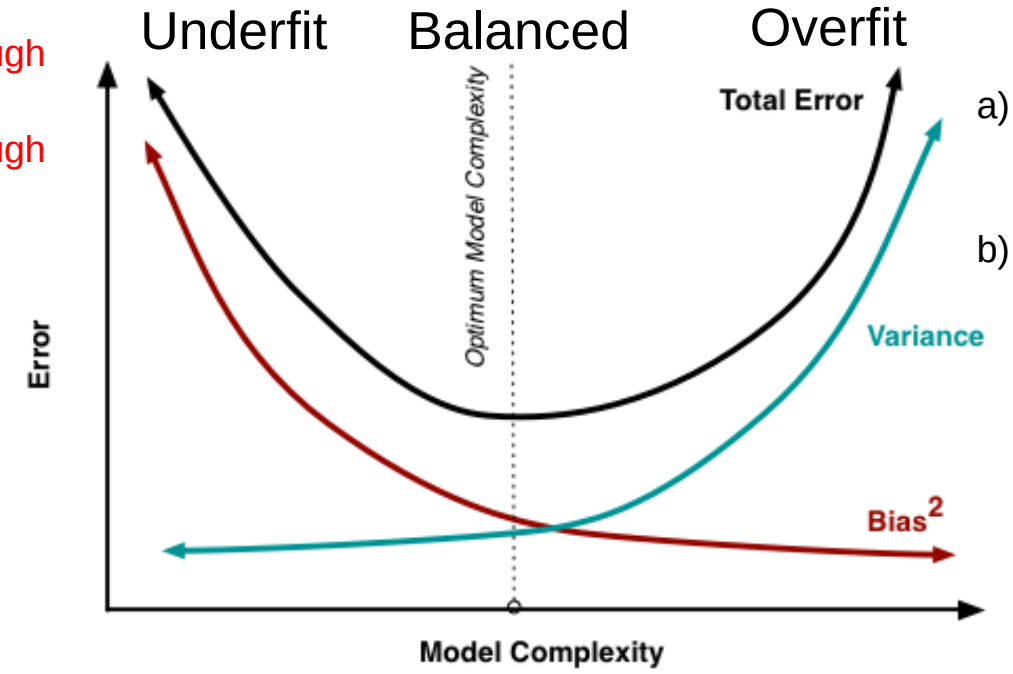
Bias and Variance

Bias and Variance

- Bias occurs when algo has limited flexibility to learn the true signal form a dataset
- Variance refers to an algo's sensitivity to specific sets of training data

Bias is the same as the mean square error (MSE). **Variance** shows how subject the model is to outliers, meaning those values that are far away from the **mean**.

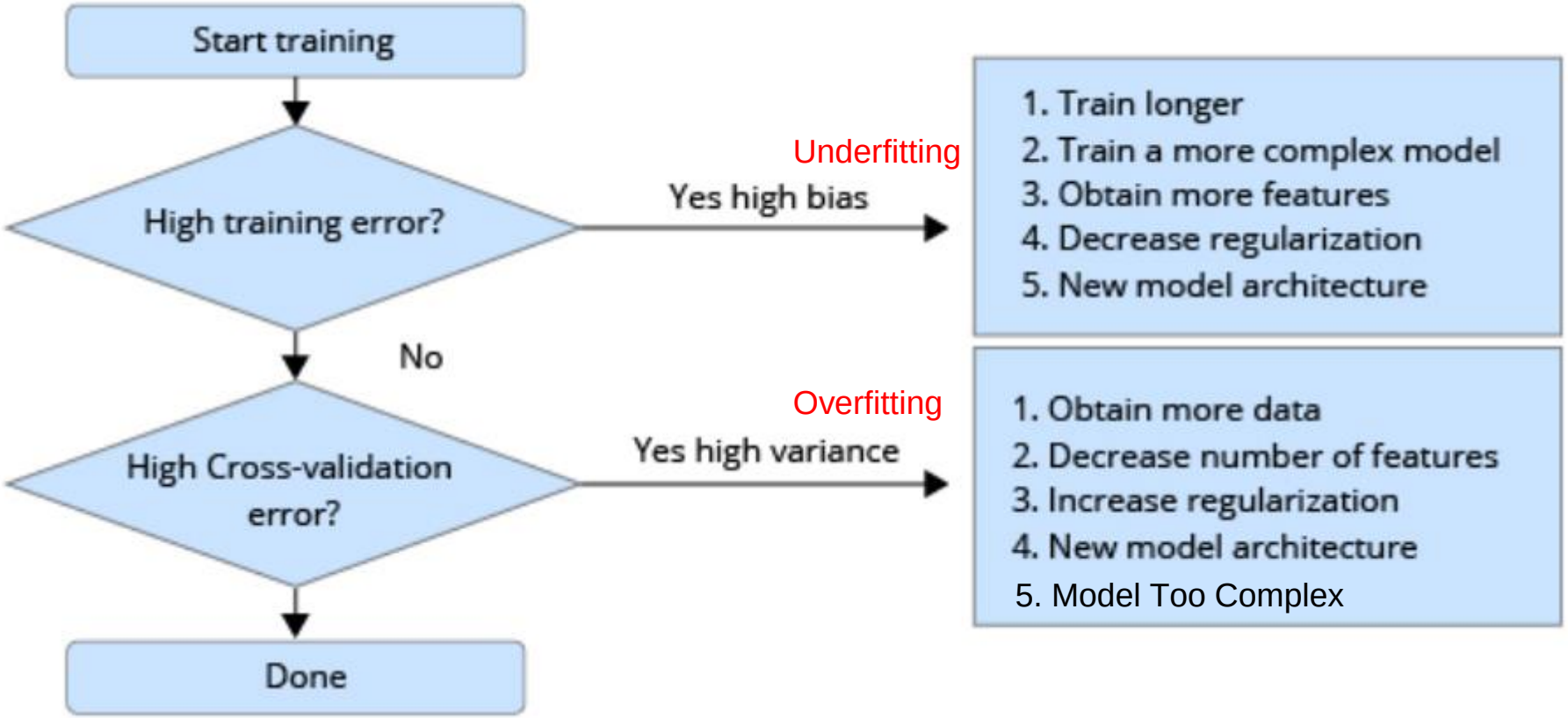
a) Not enough data
b) Not enough fractures



a) Too many of features in the dataset
b) Outlier in dataset

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Bias and Variance



Source:A systematic approach to solve a Bias-Variance Problem by Dr. Andrew Ng

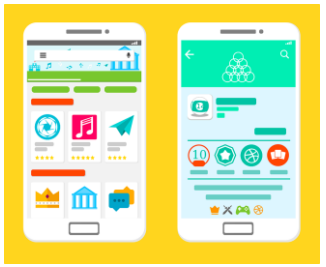
Copyright © Nanyang Polytechnic.
All Rights Reserved.

Where Is Your Trained Model Deployed?

Small AI Models Within Website

Web Sites

Small AI Models In Phones



Mobile Apps

Small AI Models in PCs

Enterprise Apps

Server

Small / Large AI Models At Server

Copyright © Nanyang Polytechnic.
All Rights Reserved.

Deployment Challeneges

Managing Data Science Languages

As you may know, ML applications often comprise of elements written in different **programming languages...** that don't always interact well with each other.

Compute Power and GPU's

Neural nets are often very deep, which means that training and using them for inference takes up a lot of compute power.

Portability.

Another interesting challenge of model deployment is the lack of portability. I noticed that it is often a problem with legacy analytics systems. Lacking the capability to easily migrate a software component to another host environment and run it there, organizations can become locked into a particular platform. This can create barriers for data scientists when creating models and deploying them.

Scalability.

Scalability is a real issue for many AI projects. Indeed, you need to make sure that your models will be able to scale and meet increases in performance and application demand in production. At the beginning of a project, we usually rely on relatively static data on a manageable scale. As the model moves forward to production, it is typically exposed to larger volumes of data and data transport modes. Your team will need several tools to both monitor and solve for the performance and scalability challenges that will show up over time.

Copyright © Nanyang Polytechnic.
All Rights Reserved.