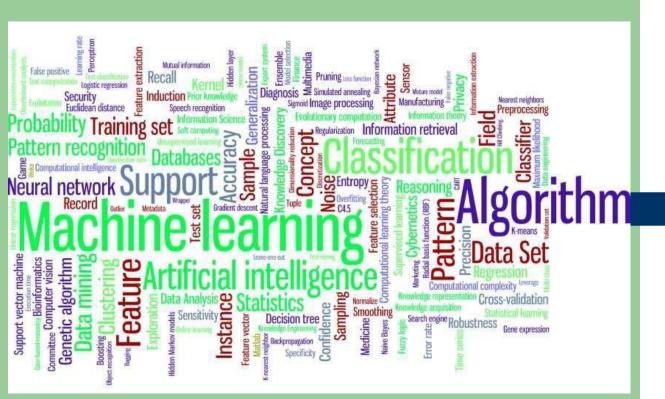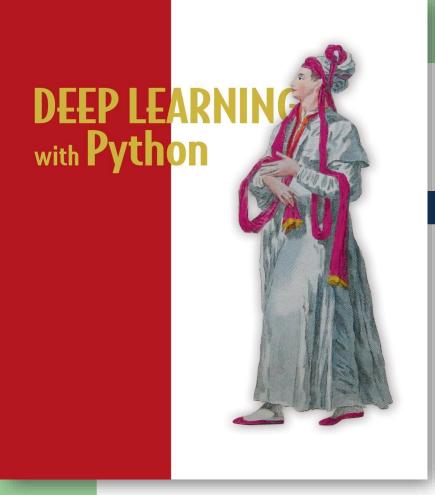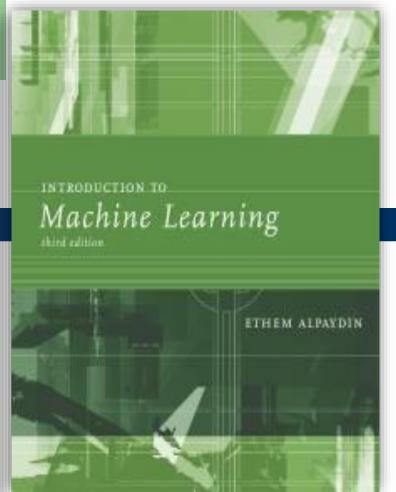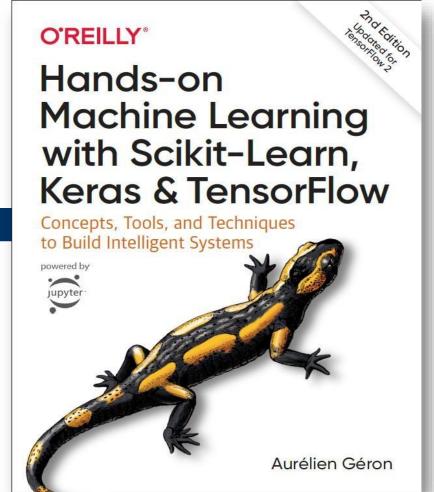# *Machine Learning with Python*

Dr. Sherif Eletriby

❑ INTRODUCTION TO Machine Learning 3rd Edition

❑ Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition by Aurélien Géron

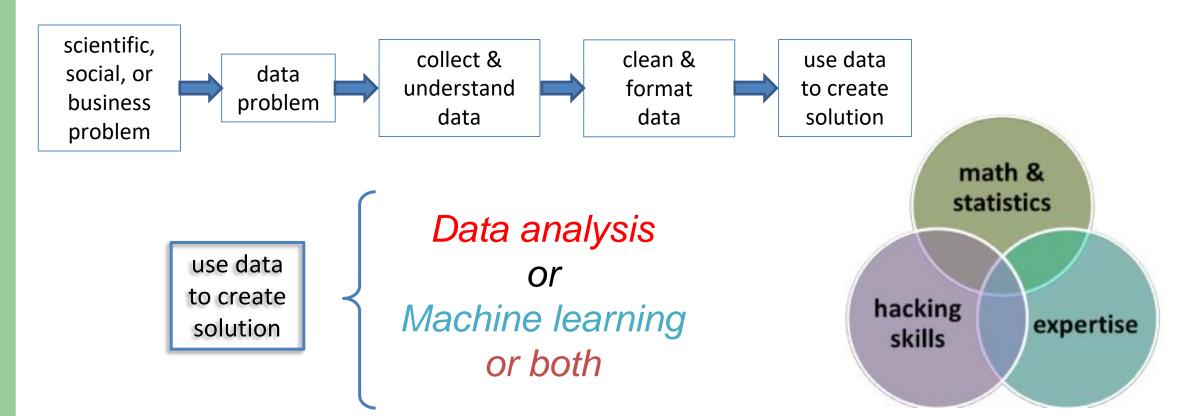❑ Deep Learning with Python Book by François Chollet

Topic	Content	
Python from scratch		
Data preprocessing & Visualization	• How to read dataset • Visualization using Seaborn {Line Plot, Scatter Plot ,Box plot} • Analyze data using pandas_profiling report	
Data preprocessing	Data encoding {label encoder , one hot encoder} • Data normalization {stander scaler, min, max} • Dealing with missing values {simple imputer,…}	
Data preprocessing	• Feature selection and dimensionality reduction {Recursive Feature Elimination , PCA} • Dealing with imbalanced data {oversampling (SMOTE) }	
Regression	• Linear regression • How to split data into (x, y & train test) • How to evaluate regression model	
Classification	• Support Vector Machine • Explain SVM kernel trick • Confusion matrix • Evaluation matrix { accuracy, precision, recall, f1-measure, roc curve , AUC} • Print classification report	
	• Knn classifier • Decision tree classifier	
Ensemble learning	Voting classifiers	
Clustering	K_mean clustering	

PROJECTS

❖ PROJECT EXAMPLES

  ▪ 1- ...

  ▪ 2- ...

❖ FINAL PROJECT DISCUSSION

WHAT IS DATA SCIENCE?

…solving problems with data…

scientific, social, or business problem → data problem → collect & understand data → clean & format data → use data to create solution

use data to create solution {

Data analysis
or
Machine learning
or both

math & statistics

hacking skills

expertise

WHAT IS DATA ANALYSIS?

…using data to discover useful information…

- **data**: anything you can *measure* or *record*

- **statistics**: summarize (and visualize) *main characteristics* of the data

- **algorithms**: apply algorithms to find *patterns* in the data

WHAT IS MACHINE LEARNING?

…creating and using models that learn from data…

- **data**: anything you can *measure* or *record*

- **model**: specification of a (mathematical) *relationship* between different variables
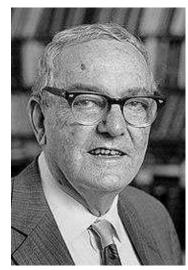
- **evaluation**: how well does the model *work*?

Machine Learning

- **Herbert Alexander Simon**:
"**Learning** is any process by which a system improves performance from experience."

- "**Machine Learning** is concerned with computer programs that automatically improve their performance through experience. "

Herbert Simon
Turing Award 1975
Nobel Prize in Economics 1978

What is Machine Learning?

Machine Learning is:

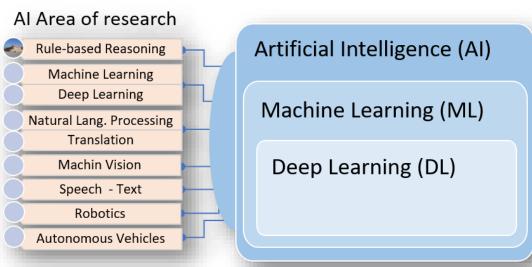- *The subfield of computer science that "gives computers the ability to learn without being explicitly programmed".*
(Arthur Samuel, 1959)

- *A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E."*
(Tom Mitchell, 1997)

What is Machine Learning?
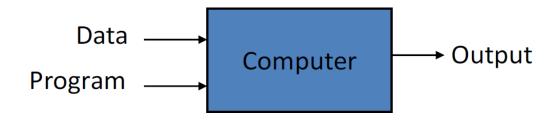
Machine Learning is:

An application of artificial intelligence that involves algorithms and data that automatically analyse and make decision by itself without human intervention.
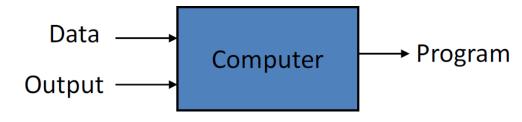
AI Area of research

- Rule-based Reasoning
- Machine Learning
- Deep Learning
- Natural Lang. Processing
- Translation
- Machin Vision
- Speech - Text
- Robotics
- Autonomous Vehicles

Artificial Intelligence (AI)

Machine Learning (ML)

Deep Learning (DL)

Traditional programming vs ML

The difference between **Traditional Programming** and **Machine Learning** is that a human developer hasn't given codes that instructs the system how to react to situation, instead it is being trained by a large number of data.
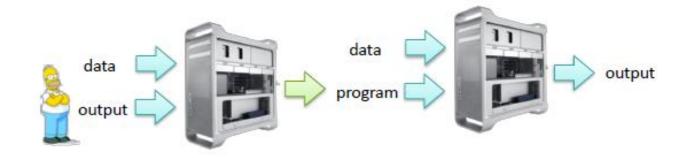
Traditional Programming

Machine Learning

WHAT IS MACHINE LEARNING?

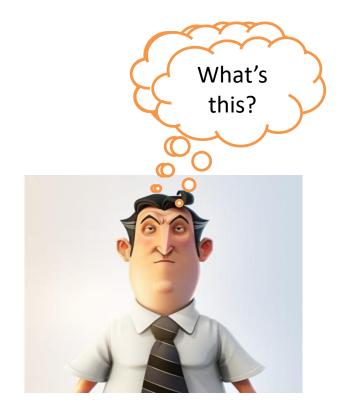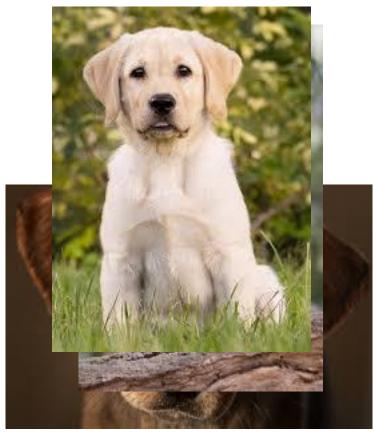- Traditional CS

- Machine Learning

Size: 250
Price: 130,000

Size: 220
Price: 115,000

Size: 200
Price: 100,000

Size: 150
Price: 70,000

Size: 100
Price: 50,000

Types of Machine Learning

- **Supervised (inductive) learning**

    Given: training data + desired outputs (labels)

- **Unsupervised learning**

    Given: training data (without desired outputs)

- **Semi-supervised learning**

    Given: training data + a few desired outputs

- **Reinforcement learning**

    Rewards from sequence of actions

Types of Machine Learning

Supervised Learning   التعلم بواسطة الإشراف

Supervised

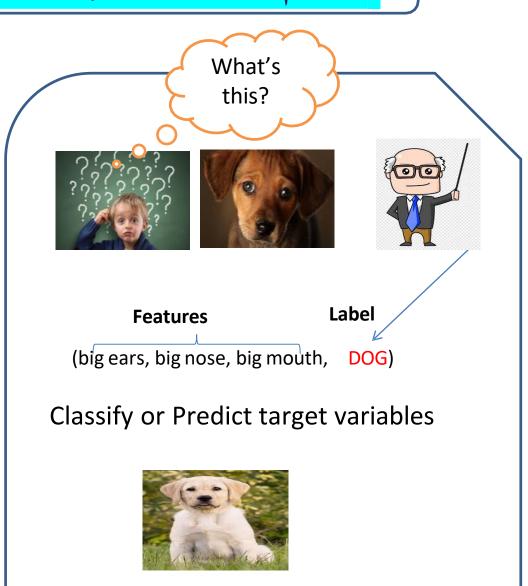→ Learn through examples of which we know the desired output (what we want to predict).

- *Is this a cat or a dog?*

- *Are these emails spam or not?*

- *Predict the market value of houses*
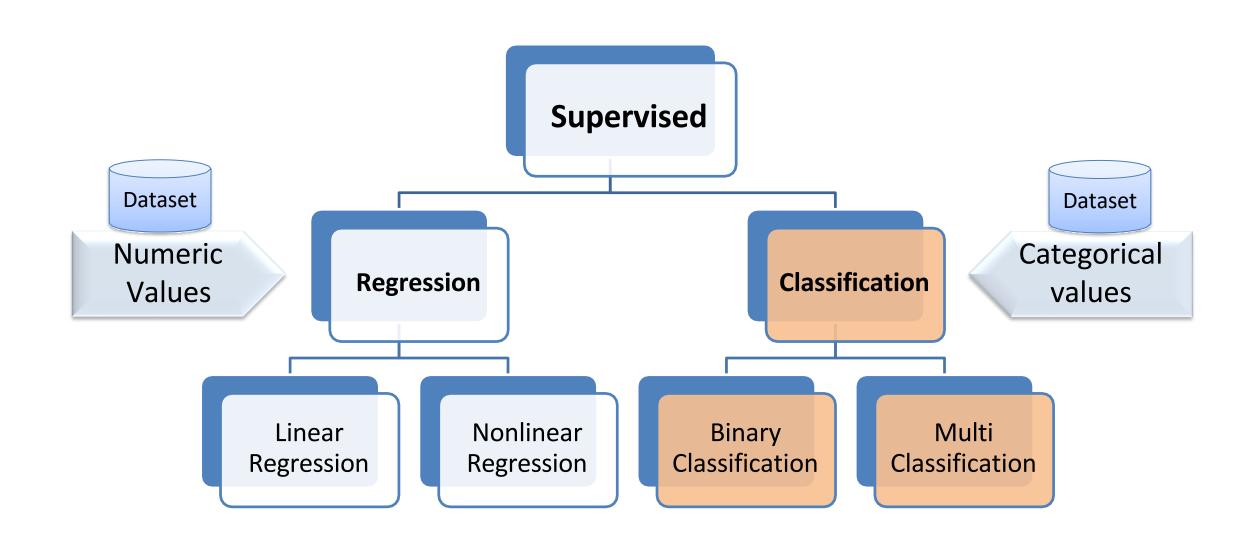
Supervised Learning التعلم بواسطة الإشراف

Supervised ➡ **Uses Labeled data**

- **What is this?**

What's this?

Features   **Label**

(big ears, big nose, big mouth,   DOG)

Classify or Predict target variables

Supervised Learning التعلم بواسطة الإشراف

Regression

Classification

Features | Label

Size	Room#	Age	Floor	Price
90	3	3	1	50
120	5	15	3	70
150	4	10	2	100
200	6	5	4	150
250	6	2	5	160
170	5	4	3	??

Training example
Training Data set

Testing Data set

Unseen data

Tweet	Class
I like it.	Positive
It is annoying	Negative
Awesome	Positive
The worst ever	Negative
It was bad product	??

Supervised Learning   التعلم بواسطة الإشراف

Regression

Features or Attributes   Label

Size	Room#	Age	Floor	Price
90	3	3	1	50
120	5	15	3	70
150	4	10	2	100
200	6	5	4	150
250	6	2	5	160
170	5	4	3	??

Training example
Training Data set

Testing Data set

Unseen data

attribute values
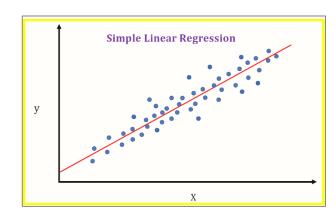
Classification

Tweet	Class
I like it.	Positive
It is annoying	Negative
Awesome	Positive
The worst ever	Negative
It was bad product	??

17
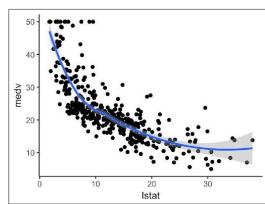
Supervised Learning التعلم بواسطة الإشراف

Regression

- Given $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
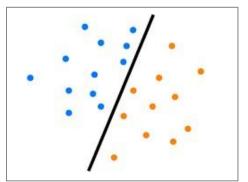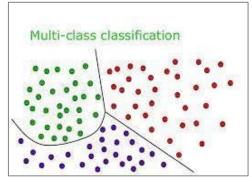 - ❑ **y** is real-valued

Output is **continuance** values
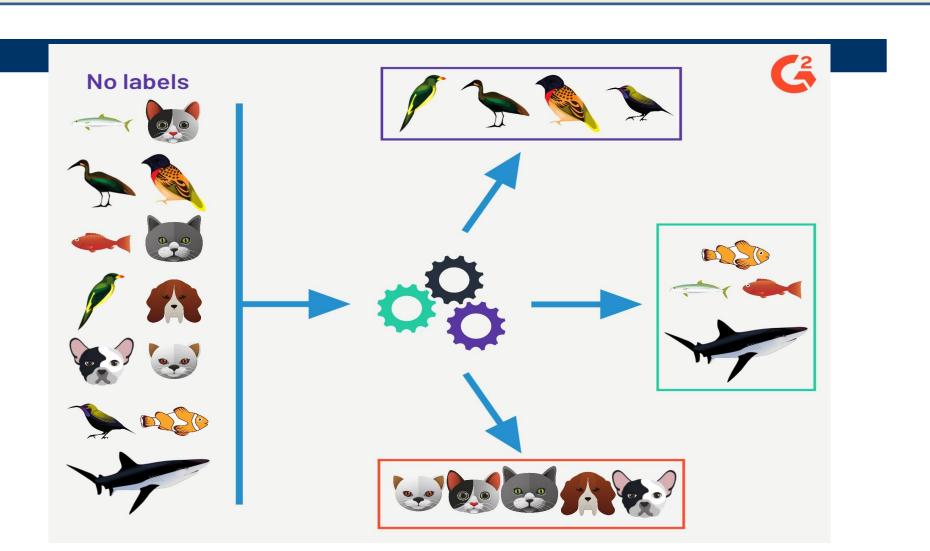(e.g. price/temperature, ..)

Classification

- Given $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
  - ❑ **y** is categorical

Output is **discrete** values
(e.g. Cat/Dog, ..)

Unsupervised Learning التعلم بدون إشراف
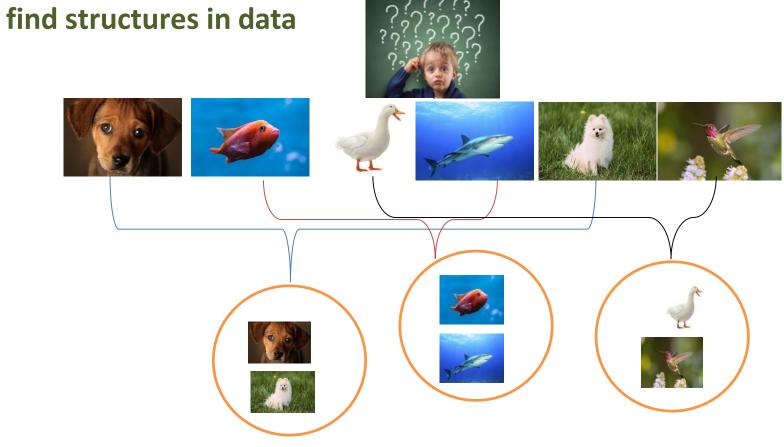
Unsupervised Learning

Unsupervised

There is no *desired* output. Learn something about the data. *Latent* relationships.

- *I have photos and want to put them in 20 groups.*
- *I want to find anomalies in the credit card usage patterns of my customers.*

Unsupervised Learning   التعلم بدون إشراف

❖ **The machine learns through observation &**
find structures in data
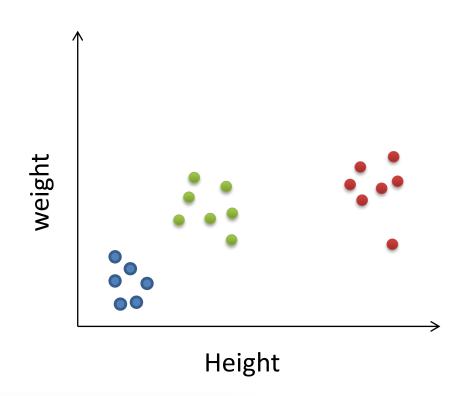
Uncovering hidden Patterns
(**Clustering**)

❖ **The machine learns through observation &**
   find structures in data

Height	Weight
180	97
160	50
155	51
177	100
67	25
54	19

Clustering Algorithm

- Given x_1, x_2, ..., x_n (without labels)
- Output hidden structure behind the x's
  – E.g., clustering

weight

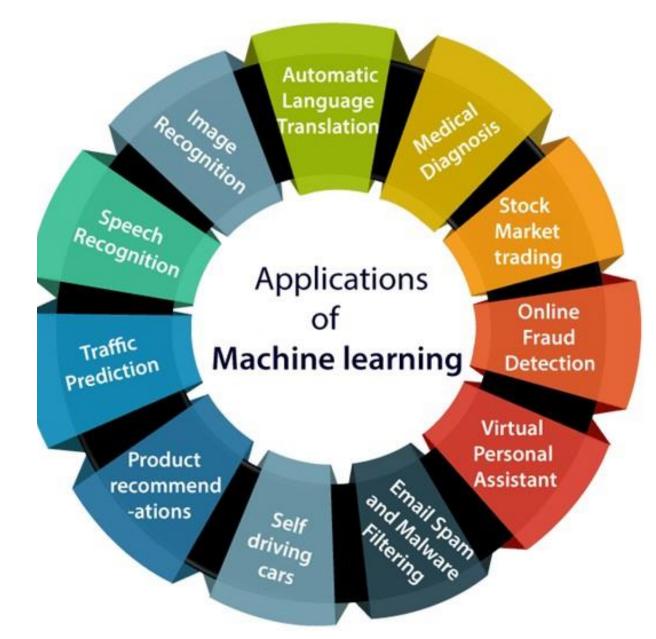Height

المقارنة بين التعلم الإشرافي و غير الإشرافي :

supervised	unsupervised
Regression, classification	clustering
more evaluation methods	fewer evaluation methods
controlled environment	less controlled environment

Machine Learning more types

- Semi-supervised learning
  - mix of Supervised and Unsupervised learning
  - usually small part of data is labelled

- Reinforcement learning
  - Model learns from a series of actions by maximizing a reward function
  - The reward function can either be maximized by penalizing bad actions and/or rewarding good actions
  - Example - training of self-driving car using feedback from the environment

Machine Learning Applications

- Image Recognition

- Speech Recognition

- Traffic prediction

- Product recommendations

- Self-driving cars

- Email Spam and Malware Filtering

- Virtual Personal Assistant

- Online Fraud Detection

Machine Learning Applications

- Loan decisions (القروض)

- Diagnose patients (تشخيص المرضى)

- Autocompleting texts (الإكمال التلقائي للنصوص)
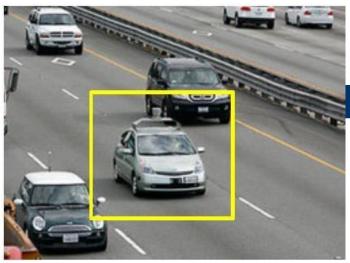
- Face Detect (تحديد الوجه)
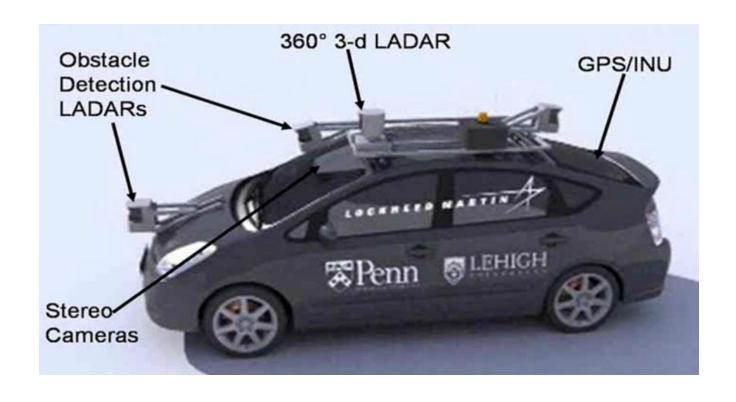

Mark Zuckerberg

Autonomous Cars

- Nevada made it legal for autonomous cars to drive on roads in June 2011

- As of 2013, four states (Nevada, Florida, California, and Michigan) have legalized autonomous cars

Penn's Autonomous Car →
(Ben Franklin Racing Team)

Autonomous Car Sensors

Algorithm Selection & Training

Supervised

- Linear classifier
- Naive Bayes
- Support Vector Machines (SVM)
- Decision Tree
- Random Forests
- k-Nearest Neighbors

Neural Networks (Deep learning)

Unsupervised
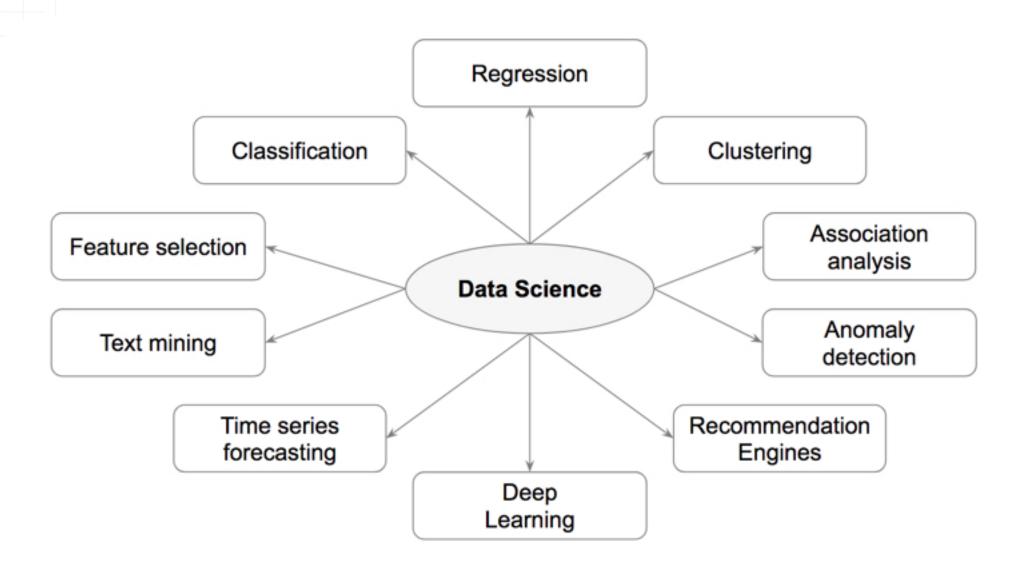
- PCA
- t-SNE
- k-means
- DBSCAN

Reinforcement

- SARSA$-\lambda$
- Q-Learning

The End

Questions?

Types of Data Science

Tasks	Description	Algorithms	Examples
Classification	Predict if a data point belongs to one of predefined classes. The prediction will be based on learning from known data set.	Decision Trees, Neural networks, Bayesian models, Induction rules, K nearest neighbors	Assigning voters into known buckets by political parties eg: soccer moms. Bucketing new customers into one of known customer groups.
Regression	Predict the numeric target label of a data point. The prediction will be based on learning from known data set.	Linear regression, Logistic regression	Predicting unemployment rate for next year. Estimating insurance premium.
Anomaly detection	Predict if a data point is an outlier compared to other data points in the data set.	Distance based, Density based, LOF	Fraud transaction detection in credit cards. Network intrusion detection.
Time series	Predict if the value of the target variable for future time frame based on history values.	Exponential smoothing, ARIMA, regression	Sales forecasting, production forecasting, virtually any growth phenomenon that needs to be extrapolated
Clustering	Identify natural clusters within the data set based on inherit properties within the data set.	K means, density based clustering - DBSCAN	Finding customer segments in a company based on transaction, web and customer call data.
Association analysis	Identify relationships within an itemset based on transaction data.	FP Growth, Apriori	Find cross selling opportunities for a retailor based on transaction purchase history.

Process Basics

Data Science Process

Data Exploration

Model Evaluation

Core Algorithms

Classification

Decision Trees

Rule Induction

k-Nearest Neighbors

Naïve Bayesian

Artificial Neural Networks

Support Vector Machines

Ensemble Learners

Regression

Linear Regression

Logistic Regression

Association Analysis

Apriori

FP-Growth

Clustering

k-Means

DBSCAN

Self-Organizing Maps

Common Applications

Text Mining

Time Series Forecasting

Anomaly Detection

Feature Selection