Machine learning

BERRIMI Mohamed
Master in Data engineering and web technologies
Machine learning enthusiast

University of Ferhat Abbas 1 , Setif
Setifian Scientific Club SeSC
Setifian Developers Group SDG

22nd May, 2019

What is covered in this lecture

What is covered in this lecture

Introduction

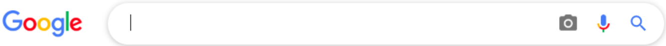
"People don't know what they are looking for untill you show them" Steve Jobs.

Introduction

"people don't know what they are looking for untill you show them"S. Jobs.
It is certain that you have used at least one Machine learning method this day

Introduction -suite-

Google search

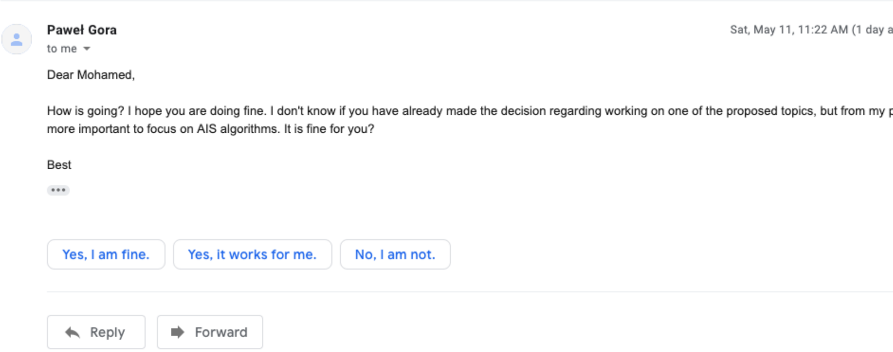


Introduction -suite-

Google image search

Introduction -suite-

Gmail reply recommendation

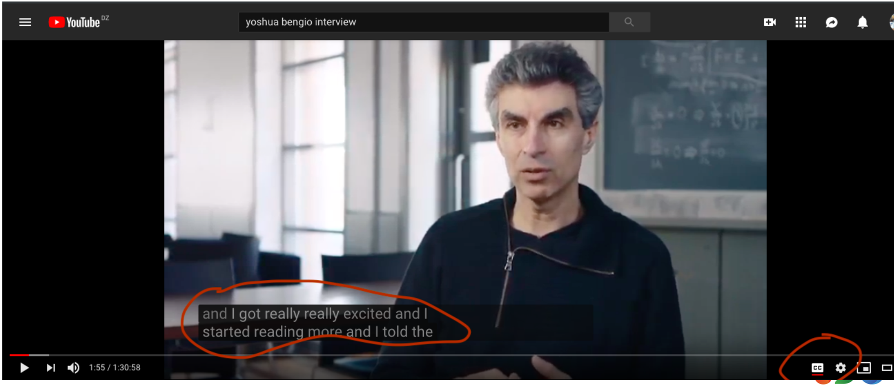


Introduction -suite-

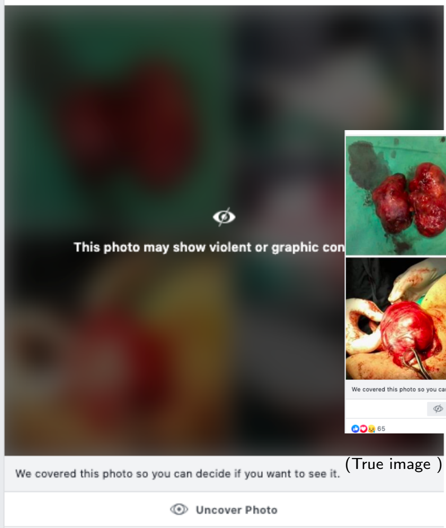
Phones assistants

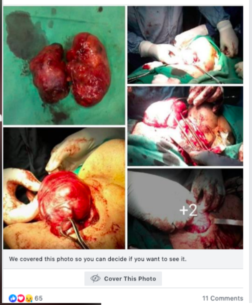
Introduction -suite-

Youtube Captions



introduction -suite-





(True image)

Introduction - suite - Facebook ML

Facebook image prediction

(True image)

HOW ?

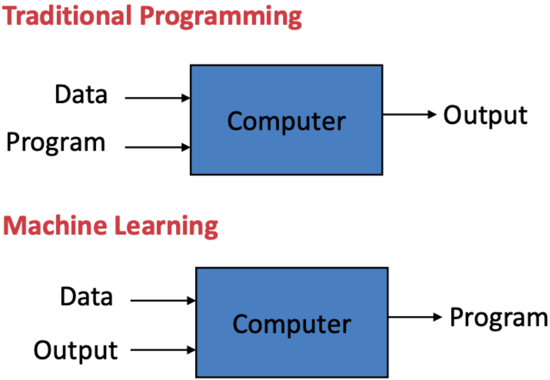
How the HECK Google and Facebook are doing that ???

What is covered in this lecture

Definitions

Learning is any process by which a system improves performance from experience.

Definitions



Definitions..

Experience

Task

Performance

A computer program is said to learn from **experience E** *with respect to some* **task T** *and some* **performance measure P**, *if its performance on T, as measured by P, improves with experience E.*

Tom Mitchell – 1998

What is covered in this lecture

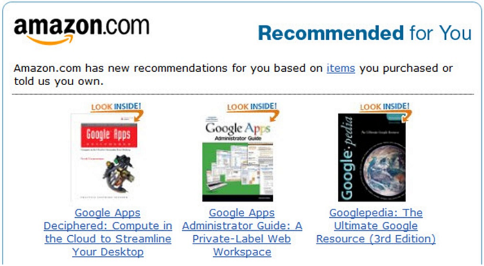
Where Machine learning is applied ?

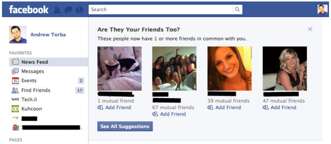
The use of machine learning is becoming very important and essential in many companies such as Google, Facebook, Apple , Amazon, Netflix .. etc

Where Machine is applied ?

The use of machine learning is becoming very important and essential in many companies such as Google, Facebook, Apple , Amazon, Netflix .. etc
Machine learning can also be applied in our life, it can improve many sides of our daily life

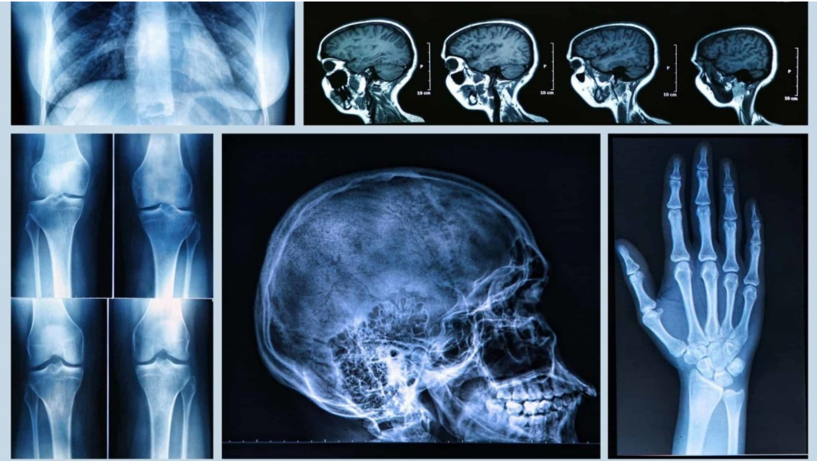
Recommendation systems



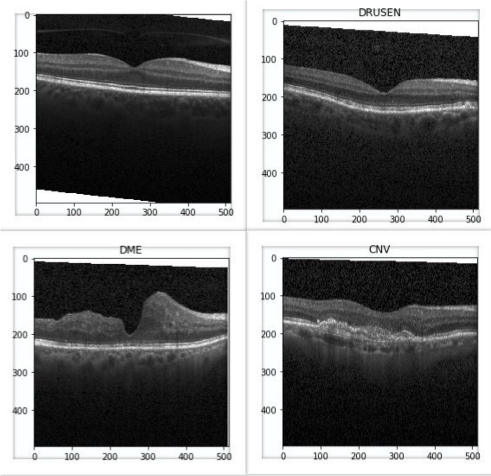


Recommendation systems

Medical imaging analysis



Medical imaging



Machine learning in security

Self driving cars

Voice recognition

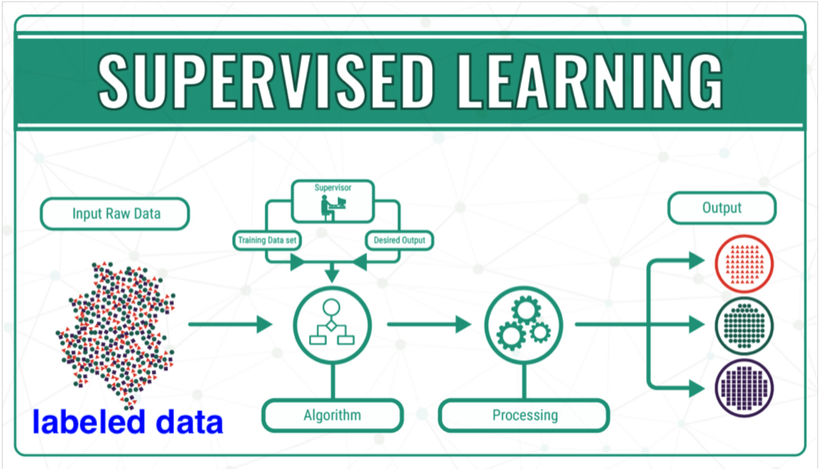
Search engines

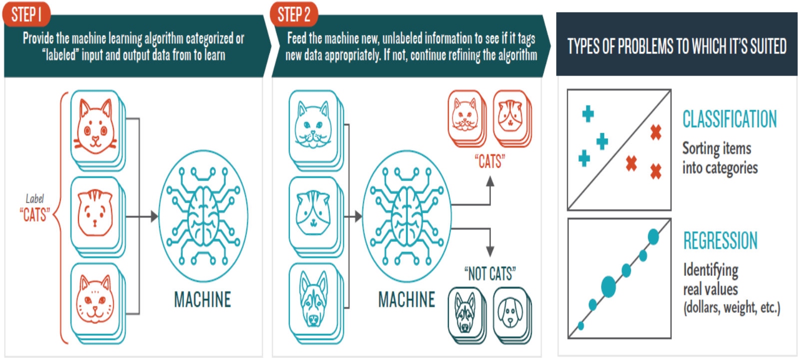
Coputer vision

Everywhere ...

If you have enogh data, You surely can apply machine learning to it

What is covered in this lecture

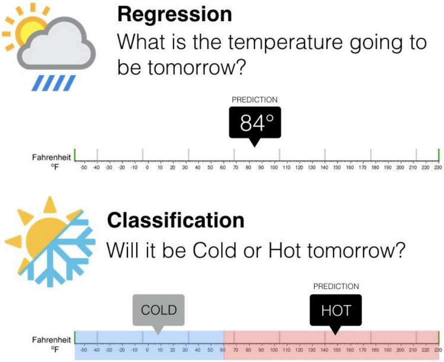

SUPERVISED LEARNING
Input Raw Data
Supervisor
Training Data set
Desired Output
Output
labeled data
Algorithm
Processing


STEP 1
Provide the machine learning algorithm categorized or "labeled" input and output data from to learn
Label "CATS"
MACHINE
STEP 2
Feed the machine new, unlabeled information to see if it tags new data appropriately. If not, continue refining the algorithm
"CATS"
MACHINE
"NOT CATS"
TYPES OF PROBLEMS TO WHICH IT'S SUITED
CLASSIFICATION
Sorting items into categories
REGRESSION
Identifying real values (dollars, weight, etc.)

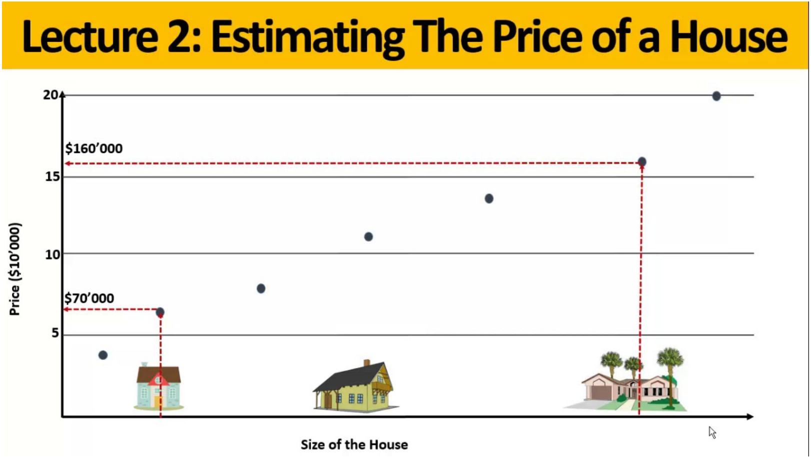
Regression

Regression is a technique from statistics that is used to predict values of a desired target **quantity** when the target quantity is **continuous**

Regression


Regression
What is the temperature going to be tomorrow?
PREDICTION
84°
Fahrenheit
°F
-50 -40 -30 -20 -10 0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230
Classification
Will it be Cold or Hot tomorrow?
COLD
PREDICTION
HOT
Fahrenheit
°F
-50 -40 -30 -20 -10 0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230

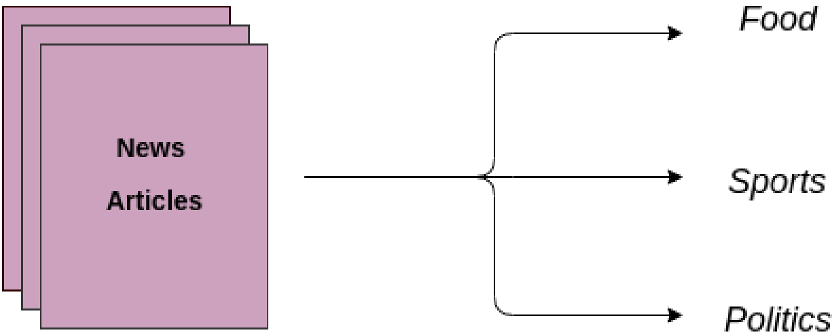
Regression


Lecture 2: Estimating The Price of a House
20
15
10
5
Price ($10'000)
$160'000
$70'000
Size of the House

Classification

Classification is the process of predicting **the class** of given data points. Classes are sometimes called as **targets/ labels or categories**. Classification predictive modeling is the task of approximating a mapping function (f) from input variables (X) to **discrete output** variables (y).

Classification examples



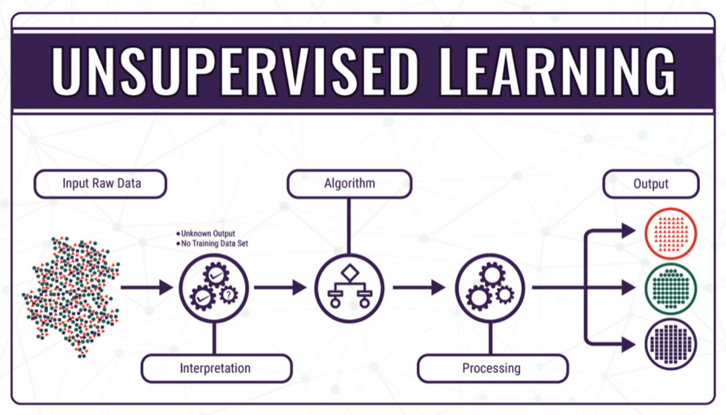
Unsupervised learning

Unsupervised Learning is a class of Machine Learning techniques to find the patterns in data. The data given to unsupervised algorithm are **not labelled**, which means only the input variables(X) are given with no corresponding output variables.

Unsupervised learning

Unsupervised Learning is a class of Machine Learning techniques to find the patterns in data. The data given to unsupervised algorithm are **not labelled**, which means only the input variables(X) are given with no corresponding output variables. Given an **Unlabled data** to an algorithm , this algorithm tries to find if there are **groups, clusters** within the given data.

Unsupervised learning



Clustering

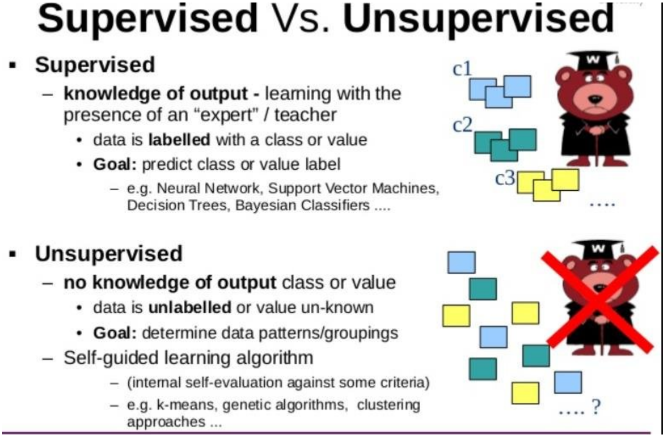
Clustering is a Machine Learning technique that involves the **grouping** of data points. Given a set of **unlabled data**, we can use a clustering algorithm to **group** each data point into a **specific group**. In theory, data points that are in the **same** group should have **similar properties and/or features**, while data points in **different** groups should have **highly dissimilar properties and/or features**

Association rules

Association rule mining is the data **mining** process of **finding the rules** that may govern associations and causal objects between sets of items. So in a given transaction with multiple items, it tries to find the rules that govern **how or why such items** are often **bought together**.

Association rules mining

Unsupervised learning Vs. Supervised learning



Where to start ML ?

- Basics are essential
- Theorical background is essential in ML , use frameworks for practice
- Step by step
- Statistics and interpretations
- Love what you are doing

Usefull ressources

AI Pionneers are alive !

Usefull ressources

- Machine learning course at corsera, full course about Machine learning, provided by Andrew NJ - Professor at Stanford University and AI cheig at Baidu.
- MIT opencourseware
- Medium and Towards Data science Blogs
- Google things ..

Advanced ML applications


ANY
QUESTIONS???
quickmeme.com

What is covered in this lecture

Data split

Overfitting and underfitting

Cross validation

Boosting

What is covered in this lecture

Ambiente de experimentação

- *Hardware*: 2 CPU AMD@1.7Ghz, 12 cores/CPU e 47GB RAM.
- *Software*: Ubuntu x64 12.04, Hadoop 2.2.0, Sun JDK 1.7.

Configuração dos experimentos

Texto texto texto texto texto texto texto texto

	Coluna1	Coluna2
Linha1	1	2
Linha2 Vcores	3	4

Resultados de desempenho

figs/Figura12-totalCores.png

What is covered in this lecture

Conclusão

- Item
- Item
- Item

Trabalhos Futuros

- Item
- Item
- Item

What is covered in this lecture

Referências

- DEY, A. K. Understanding and Using Context. Personal Ubiquitous Comput., London, UK, UK, v.5, n.1, p.4-7, Jan. 2001.
- MAAMAR, Z.; BENSLIMANE, D.; NARENDRA, N. C. What can context do for web services? Commun. ACM, New York, NY, USA, v.49, n.12, p.98-103, Dec. 2006.
- KUMAR, K. A. et al. CASH: context aware scheduler for hadoop. In: INTERNATIONAL CONFERENCE ON ADVANCES IN COMPUTING, COMMUNICATIONS AND INFORMATICS, New York, NY, USA. Proceedings. . . ACM, 2012. p.52-61. (ICACCI '12).
- ZAHARIA, M. et al. Improving MapReduce performance in heterogeneous environments. In: USENIX CONFERENCE ON OPERATING SYSTEMS DESIGN AND IMPLEMENTATION, 8., Berkeley, CA, USA. Proceedings. . . USENIX Association, 2008. p.29-42. (OSDI'08).
- TIAN, C. et al. A Dynamic MapReduce Scheduler for Heterogeneous Workloads. In: EIGHTH INTERNATIONAL CONFERENCE ON GRID AND COOPERATIVE COMPUTING, 2009., Washington, DC, USA. Proceedings. . . IEEE Computer Society, 2009. p.218-224. (GCC '09).
- CHEN, Q. et al. SAMR: a self-adaptive mapreduce scheduling algorithm in heterogeneous environment. In: IEEE INTERNATIONAL CONFERENCE ON COMPUTER AND INFORMATION TECHNOLOGY, 2010., Washington, DC, USA. Proceedings. . . IEEE Computer Society, 2010. p.2736-2743. (CIT '10).

Referências (Continuação)

- RASOOLI, A.; DOWN, D. G. Coshh: a classification and optimization based scheduler for heterogeneous hadoop systems. In: SC COMPANION: HIGH PERFORMANCE COMPUTING, NETWORKING STORAGE AND ANALYSIS, 2012., Washington, DC, USA. Proceedings. . . IEEE Computer Society, 2012. p.1284-1291. (SCC '12).
- ISARD, M. et al. Quincy: fair scheduling for distributed computing clusters. In: ACM SIGOPS 22ND SYMPOSIUM ON OPERATING SYSTEMS PRINCIPLES, New York, NY, USA. Proceedings. . . ACM, 2009. p.261-276. (SOSP '09).
- XIE, J. et al. Improving MapReduce performance through data placement in heterogeneous Hadoop clusters. In: PARALLEL AND DISTRIBUTED PROCESSING, WORKSHOPS AND PHD FORUM (IPDPSW). Anais. . . IEEE International Symposium, 2010.
- HADOOP, A. Arquitetura do HDFS. http://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-hdfs/Federation.html, Acesso em novembro de 2013.
- HADOOP, A. Arquitetura do YARN. http://hadoop.apache.org/docs/current/hadoop-yarn/hadoop-yarn-site/YARN.html, Acesso em novembro de 2013.

Machine learning

BERRIMI Mohamed
Master in Data engineering and web technologies
Machine learning enthusiast

University of Ferhat Abbas 1 , Setif
Setifian Scientific Club SeSC
Setifian Developers Group SDG

22nd May, 2019