

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

Introduction

Machine learning

ML applications

Learning types

Machine learning concepts

Advanced ML applications

Ressrouces

Questions



What is covered in this lecture

Introduction

Machine learning

ML applications

Learning types

Machine learning concepts

Advanced ML applications

Ressrouces

Questions



Introduction

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



Introduction

"people don't know what they are looking for until 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



Pawel Gora

to me ▾

Sat, May 11, 11:22 AM (1 day ago)

Dear Mohamed,

How is going? I hope you are doing fine. I don't know if you have already made the decision regarding working on one of the proposed topics, but from my point of view it would be more important to focus on AIS algorithms. It is fine for you?

Best

[Yes, I am fine.](#)

[Yes, It works for me.](#)

[No, I am not.](#)

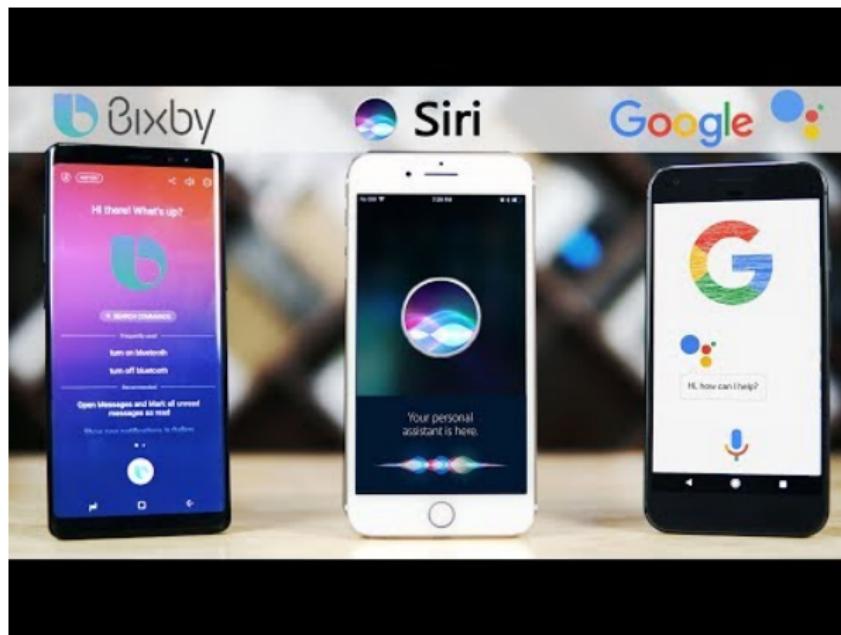
Reply

Forward



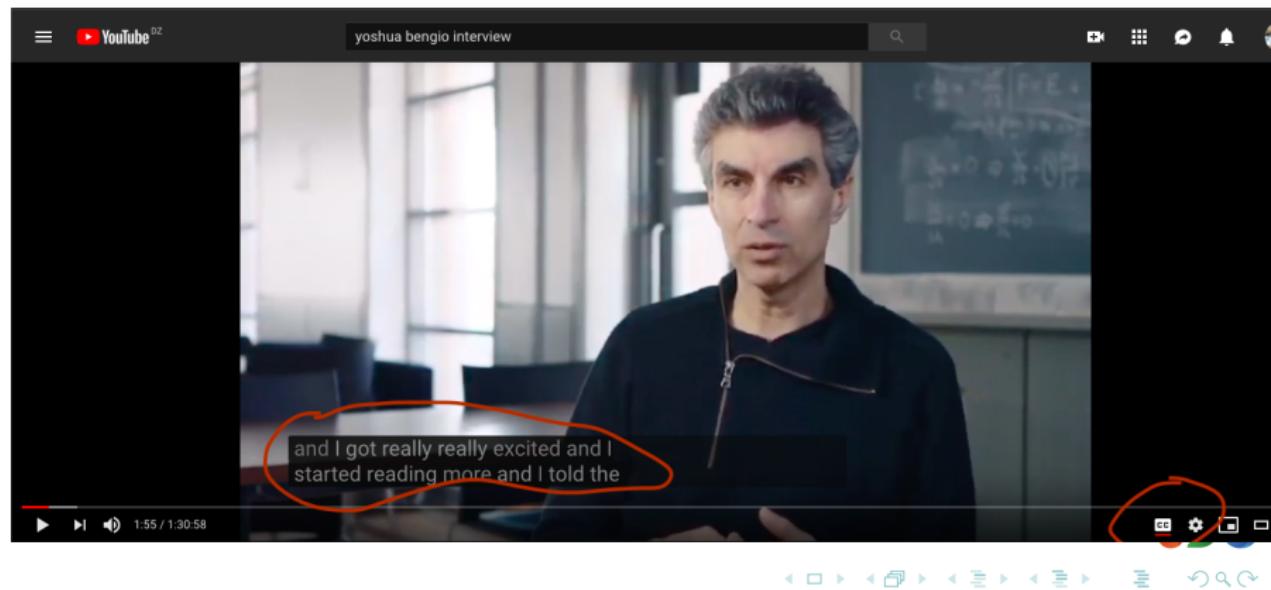
Introduction -suite-

Phones assistants

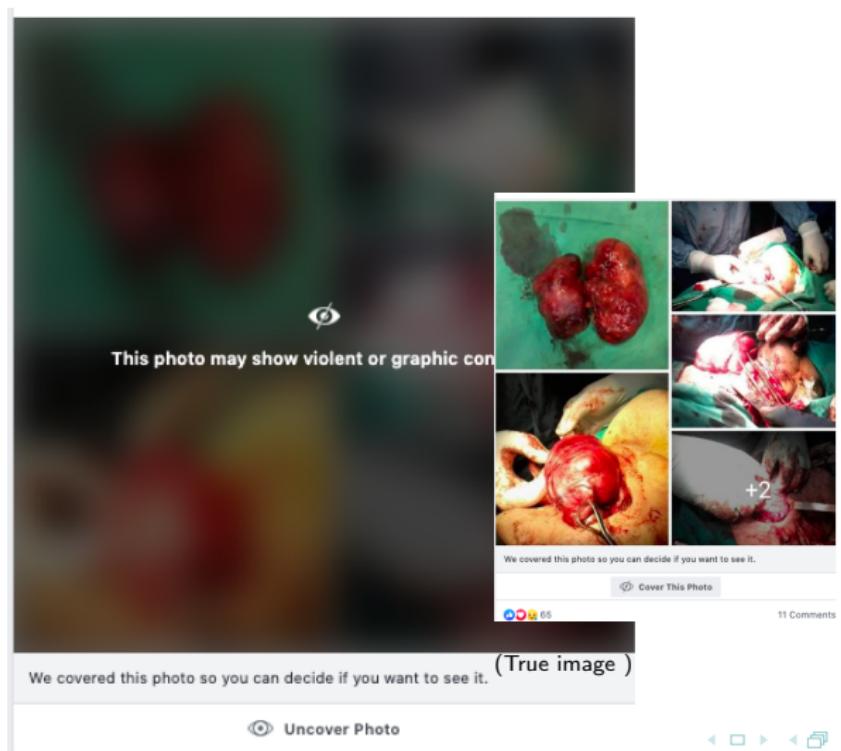


Introduction -suite-

Youtube Captions



introduction -suite-



Introduction - suite - Facebook ML



Mohamed Berrimi added a new photo.

April 5 ·

Image may contain: 1 person, sunglasses and outdoor



(True image)

Like

Comment

Share

Facebook image prediction

SDG



HOW ?

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



What is covered in this lecture

Introduction

Machine learning

ML applications

Learning types

Machine learning concepts

Advanced ML applications

Ressrouces

Questions



Definitions

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



Definitions

Traditional Programming



Machine Learning



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

Introduction

Machine learning

ML applications

Learning types

Machine learning concepts

Advanced ML applications

Ressrouces

Questions



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

The screenshot shows the Amazon.com homepage with a "Recommended for You" section. It features three book covers with "LOOK INSIDE!" buttons:

- Google Apps Deciphered: Compute in the Cloud to Streamline Your Desktop**
- Google Apps Administrator Guide: A Private-Label Web Workspace**
- Googlepedia: The Ultimate Google Resource (3rd Edition)**

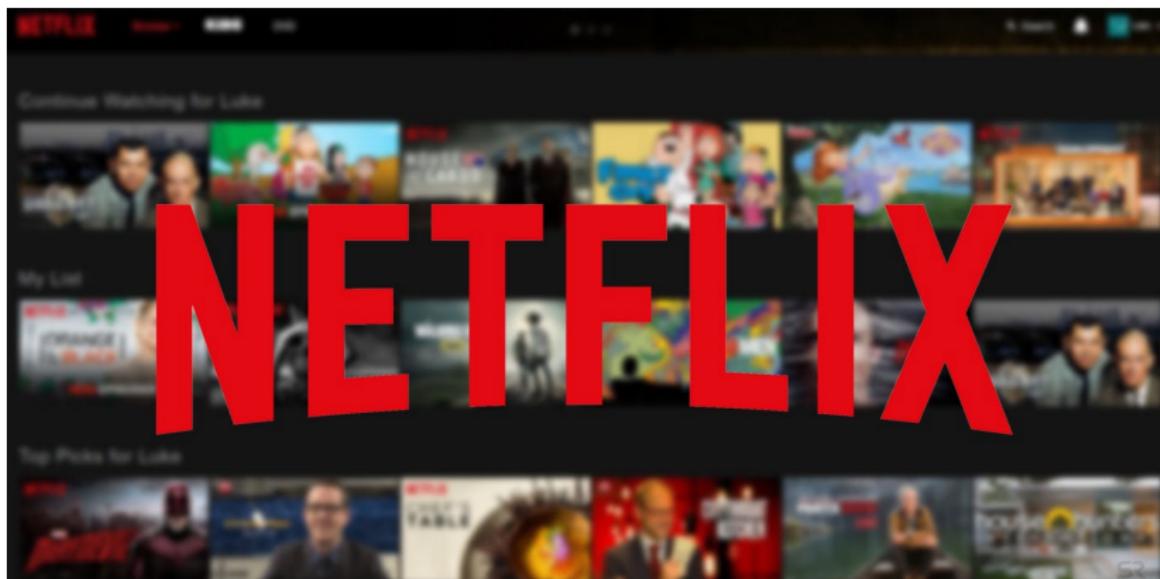
Below the books, there is descriptive text: "Amazon.com has new recommendations for you based on items you purchased or told us you own."

The screenshot shows a Facebook profile for Andrew Tolia. On the left, under "FAVORITES", is a list of links: News Feed, Message, Events, Find Friends, Tech3, Kachoo!, and a few other items. On the right, there is a "Are They Your Friends Too?" section with four suggestions:

- 1 mutual friend (1) Add Friend
- 1 mutual friend (1) Add Friend
- 39 mutual friends (39) Add Friend
- 47 mutual friends (47) Add Friend

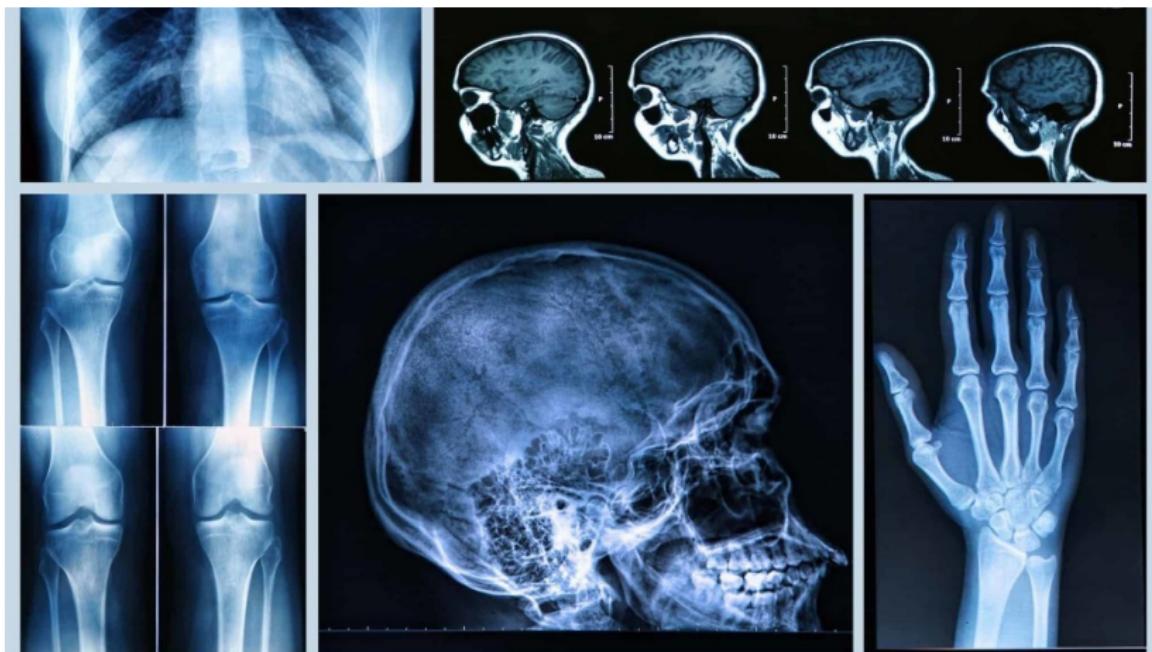
At the bottom, there are navigation icons for back, forward, search, and other functions.

Recommendation systems



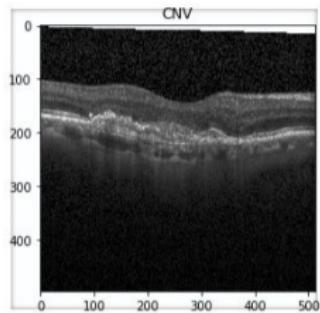
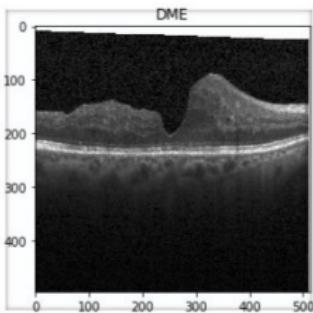
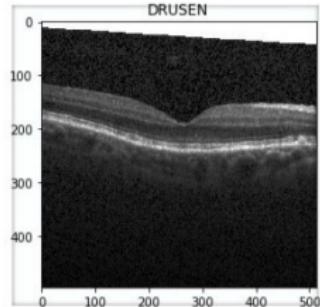
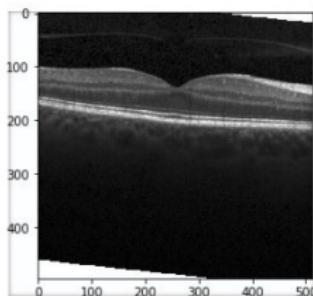
SDG

Medical imaging analysis



SDG

Medical imaging



Machine learning in security



SDG

Self driving cars



Voice recognition



Search engines

Bing



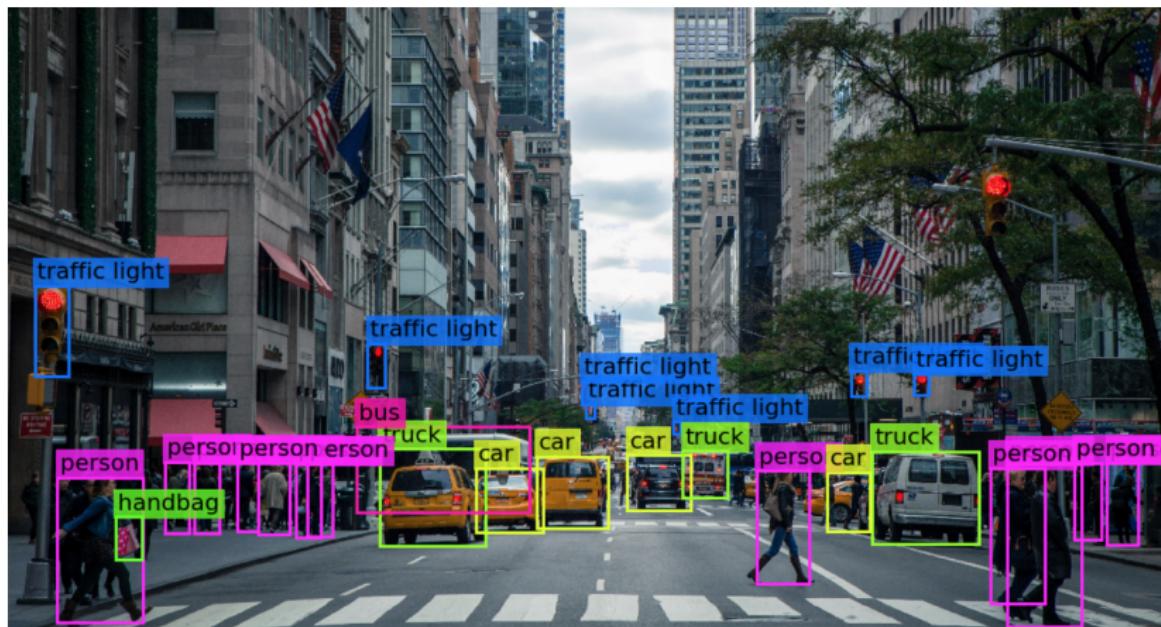
DuckDuckGo

Baidu 百度

YAHOO!



Computer vision



SDG

Everywhere ...

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



What is covered in this lecture

Introduction

Machine learning

ML applications

Learning types

Machine learning concepts

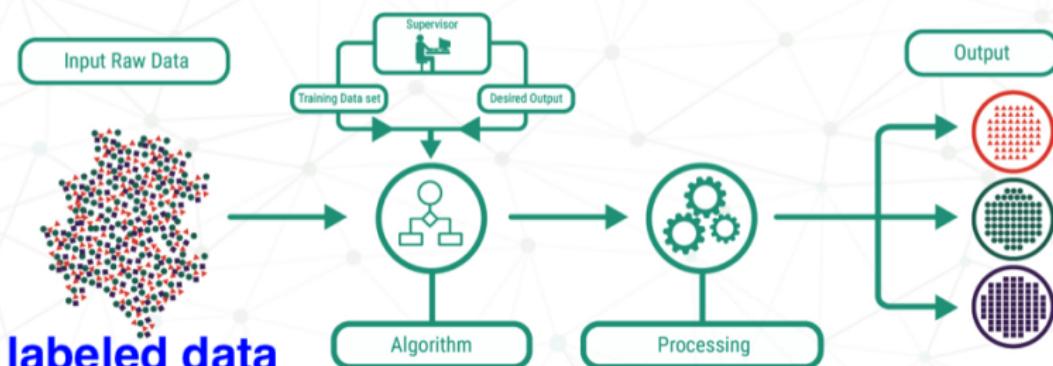
Advanced ML applications

Ressrouces

Questions



SUPERVISED LEARNING

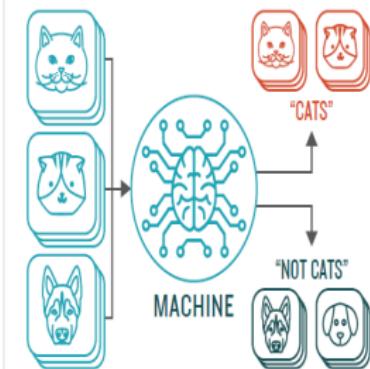
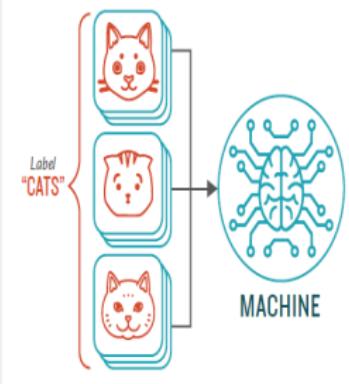
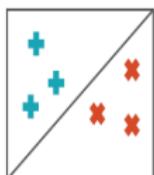


STEP 1

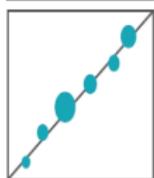
Provide the machine learning algorithm categorized or "labeled" input and output data from to learn

STEP 2

Feed the machine new, unlabeled information to see if it tags new data appropriately. If not, continue refining the algorithm

**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



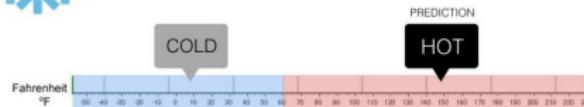
Regression

What is the temperature going to be tomorrow?



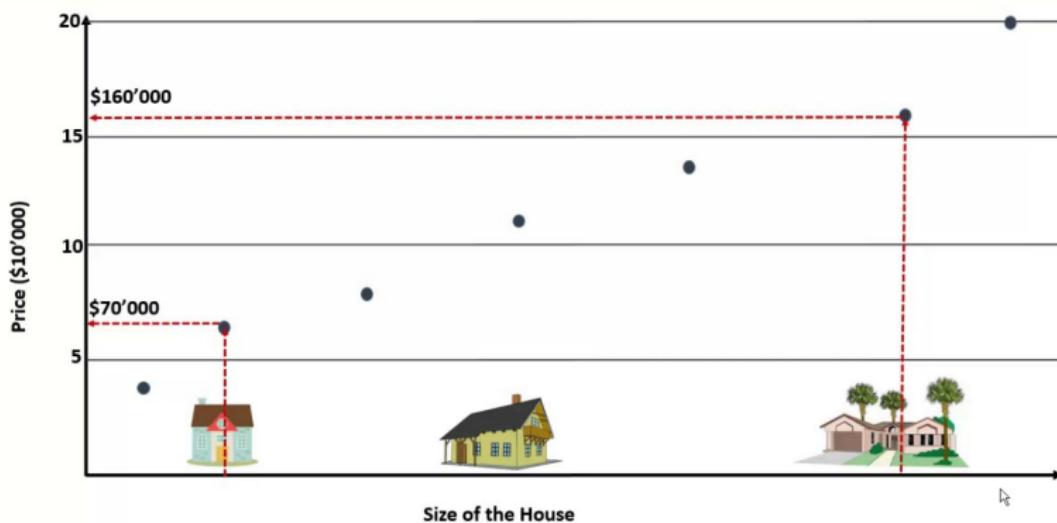
Classification

Will it be Cold or Hot tomorrow?



Regression

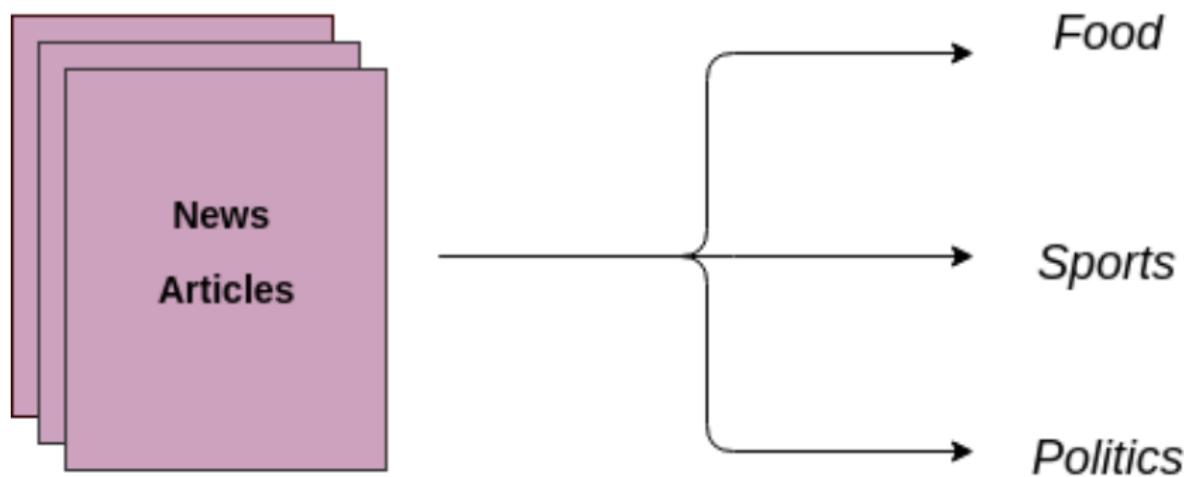
Lecture 2: Estimating The Price of a House



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



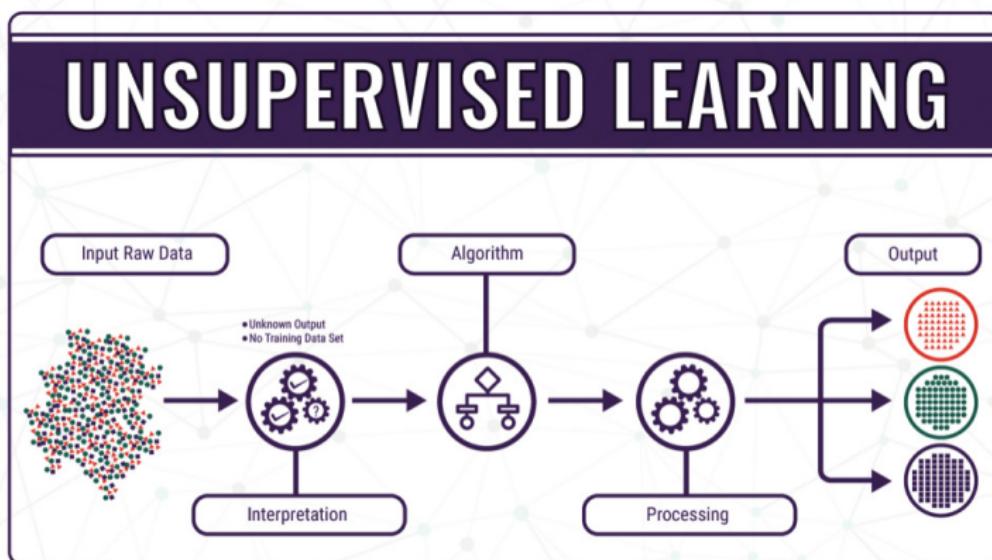
Unsupervised learning

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 **Unlabelled 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 **unlabeled 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

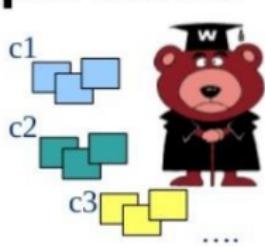


Unsupervised learning Vs. Supervised learning

Supervised Vs. Unsupervised

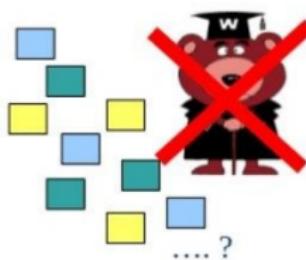
▪ Supervised

- **knowledge of output** - learning with the presence of an “expert” / teacher
 - data is **labelled** with a class or value
 - **Goal:** predict class or value label
 - e.g. Neural Network, Support Vector Machines, Decision Trees, Bayesian Classifiers



▪ Unsupervised

- **no knowledge of output** class or value
 - data is **unlabelled** or value un-known
 - **Goal:** determine data patterns/groupings
- Self-guided learning algorithm
 - (internal self-evaluation against some criteria)
 - e.g. k-means, genetic algorithms, clustering approaches ...



Where to start ML ?

- ▶ Basics are essential
- ▶ Theoretical 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



What is covered in this lecture

Introduction

Machine learning

ML applications

Learning types

Machine learning concepts

Advanced ML applications

Ressources

Questions

Data split

Overfitting and underfitting

Cross validation

Boosting

What is covered in this lecture

Introduction

Machine learning

ML applications

Learning types

Machine learning concepts

Advanced ML applications

Ressources

Questions

Ambiente de experimentação

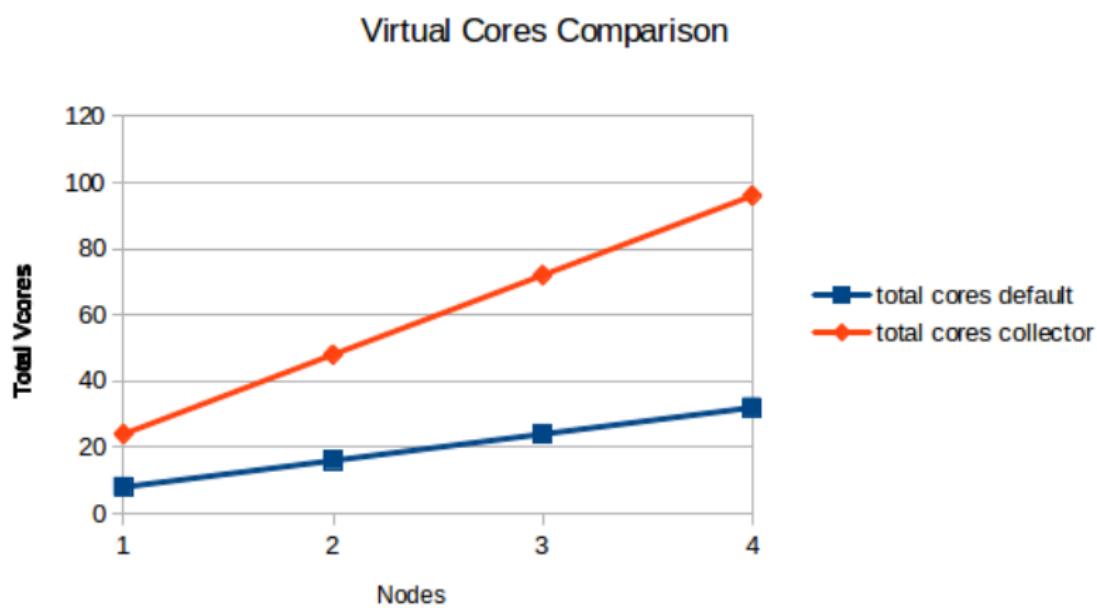
- ▶ *Hardware*: 2 CPU AMD@1.7Ghz, 12 cores/CPU e 47GB RAM.
- ▶ *Software*: Ubuntu x64 12.04, Hadoop 2.2.0, Sun JDK 1.7.

Configuração dos experimentos

Texto texto texto texto texto texto texto

	Coluna1	Coluna2
<i>Linha1</i>	1	2
<i>Linha2</i> Vcores	3	4

Resultados de desempenho



What is covered in this lecture

Introduction

Machine learning

ML applications

Learning types

Machine learning concepts

Advanced ML applications

Ressrouces

Questions

Conclusão

- ▶ Item
- ▶ Item
- ▶ Item

Trabalhos Futuros

- ▶ Item
- ▶ Item
- ▶ Item

What is covered in this lecture

Introduction

Machine learning

ML applications

Learning types

Machine learning concepts

Advanced ML applications

Ressrouces

Questions



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