



Traitement Numérique des Données

M1 – INF 2163

AIDN: Applications Interactives et
Données Numériques

Sylvie Gibet

1

1

Introduction à l'apprentissage automatique

cours inspiré du cours d'Andrew Ng, Stanford, Californie (Mooc coursera)

2

2

Qu'est-ce que le machine learning?

- Champ de l'Intelligence Artificielle (IA)

3

3

Qu'est-ce que le machine learning?

- Champ de l'Intelligence Artificielle (IA)
- Nouvelles possibilités pour les ordinateurs

4

4

Qu'est-ce que le machine learning?

- Champ de l'Intelligence Artificielle (IA)
- Nouvelles possibilités pour les ordinateurs
- Ce n'est pas une discipline ! Mais c'est important d'explorer ces techniques avec l'arrivée du Big Data

5

5

Qu'est-ce que le machine learning?

- Champ de l'Intelligence Artificielle (IA)
- Nouvelles possibilités pour les ordinateurs
- Ce n'est pas une discipline ! Mais c'est important d'explorer ces techniques avec l'arrivée du Big Data

Objectifs

Consiste à concevoir des algorithmes qui apprennent à partir des données, afin de **prédire** le comportement d'un système (phénomène), ou pour **classifier** des données, ou enfin pour **discriminer** différents "clusters" au sein des données.

6

6

Qu'est-ce que le machine learning?

Exemples

- ▣ Fouille de données (data mining)

7

7

Qu'est-ce que le machine learning?

- ▣ **Arthur Lee Samuel** (December 5, 1901 – July 29, 1990) was an American pioneer in the field of computer gaming and artificial intelligence (AI). He coined the term “machine learning” in 1959. “Field of study that gives the computer the ability to learn without being explicitly programmed

8

8

Qu'est-ce que le machine learning?

- **Arthur Lee Samuel** (December 5, 1901 – July 29, 1990) was an American pioneer in the field of computer gaming and artificial intelligence (AI). He coined the term "machine learning" in 1959. "Field of study that gives the computer the ability to learn without being explicitly programmed"
- **Tom Mitchell** (1998): 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.
- Un programme informatique est dit apprendre à partir de l'expérience E selon une tâche T et une mesure de performance P, si sa performance sur T, mesurée par P, s'améliore avec l'expérience E.

9

9

Qu'est-ce que le machine learning?

- **Arthur Lee Samuel** (December 5, 1901 – July 29, 1990) was an American pioneer in the field of computer gaming and artificial intelligence (AI). He coined the term "machine learning" in 1959. "Field of study that gives the computer the ability to learn without being explicitly programmed"
The Samuel Checkers-playing Program appears to be the world's first self-learning program, and as such a very early demonstration of the fundamental concept of AI.
- **Tom Mitchell** (1998): 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.
- Un programme informatique est dit apprendre à partir de l'expérience E selon une tâche T et une mesure de performance P, si sa performance sur T, mesurée par P, s'améliore avec l'expérience E.
- **Remarque : experience vue comme apprentissage (training)**

10

10

Qu'est-ce que le machine learning?

- Un programme informatique est dit apprendre à partir de l'expérience E selon une tâche T et une mesure de performance P, si sa performance sur T, mesurée par P, s'améliore avec l'expérience E.
 - Exercice : Supposez que votre application de messagerie scrute quels sont les emails marqués comme spam /pas spam, et sur cette base apprend comment mieux filtrer les spam. Comment définissez vous la tâche T dans cette situation?
- A. Elle classe les emails comme spam / pas spam
 - B. Elle vous observe marquer les emails comme spam / pas spam
 - C. Elle évalue le nombre (ou la proportion) d'emails correctement classés comme spam / pas spam
 - D. Aucune des assertions précédentes – ce n'est pas un problème de machine learning

11

11

Qu'est-ce que le machine learning?

- Un programme informatique est dit apprendre à partir de l'expérience E selon une tâche T et une mesure de performance P, si sa performance sur T, mesurée par P, s'améliore avec l'expérience E.
 - Exercice : Supposez que votre application de messagerie scrute quels sont les emails marqués comme spam /pas spam, et sur cette base apprend comment mieux filtrer les spam. Comment définissez vous la tâche T dans cette situation?
- A. Elle classe les emails comme spam / pas spam ← T
 - B. Elle vous observe marquer les emails comme spam / pas spam ← E
 - C. Elle évalue le nombre (ou la proportion) d'emails correctement classés comme spam / pas spam ← P
 - D. Aucune des assertions précédentes – ce n'est pas un problème de machine learning

12

12

Qu'est-ce que le machine learning?

Première définition

1. Observation d'un phénomène
2. Construction d'un modèle de ce phénomène à partir des données
3. Prédiction automatique et analyse de ce phénomène, sans intervention humaine

13

13

Machine learning

Modélisation mathématique

- Observation d'un phénomène-> data z_i dans Z
- Deux problèmes (en fonction des données et des tâches)
 - ▣ Apprentissage non supervisé (non supervised learning)
 - Pas de structure interne de Z connue
 - Clustering: regroupe les exemples qui sont corrélés
 - ▣ Apprentissage supervisé (Supervised learning)
 - $Z = (x,y)$ in $X \times Y$ (x : donnée ; y : classe)
 - Modéliser le lien entre x et y
 - Faire des prévisions : connaissant x , prédire y

14

14

Machine learning: deux étapes

- Pour un certain nombre de méthodes on a deux étapes dans le processus de machine learning
 - ▣ **Phase d'entraînement (training)** : l'algorithme apprend sur une partie des données
 - ▣ **Phase de test** : test sur une autre partie des données

15

15

Trois approches

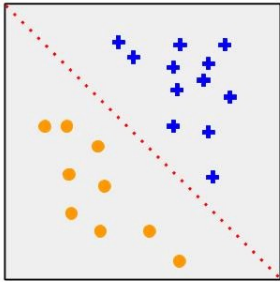
- **Classification (supervisée)**
 - ▣ Dans l'ensemble d'apprentissage (training), on sait a priori que chaque exemple (sample) de cet ensemble correspond à une classe spécifique
 - ▣ Dans l'ensemble de test on trouve automatiquement la classe associée à chaque nouvel exemple.
- **Régression (cas continu, supervisé)**
 - ▣ Les données sont reliées les unes aux autres à travers un processus de régression (une droite, une courbe, une tendance statistique).
- **Clustering (non supervisé)**
 - ▣ Comme pour l'apprentissage supervisé, mais on ne connaît pas a priori les classes associées aux exemples.

16

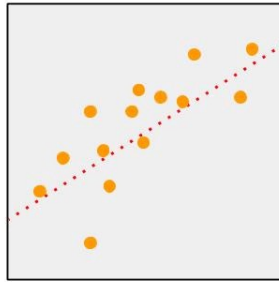
16

Trois méthodes principales

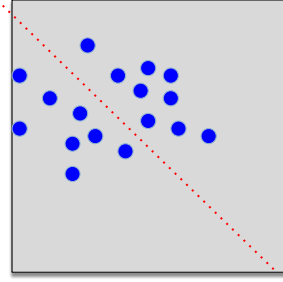
classification



régression



clustering



un

17

17

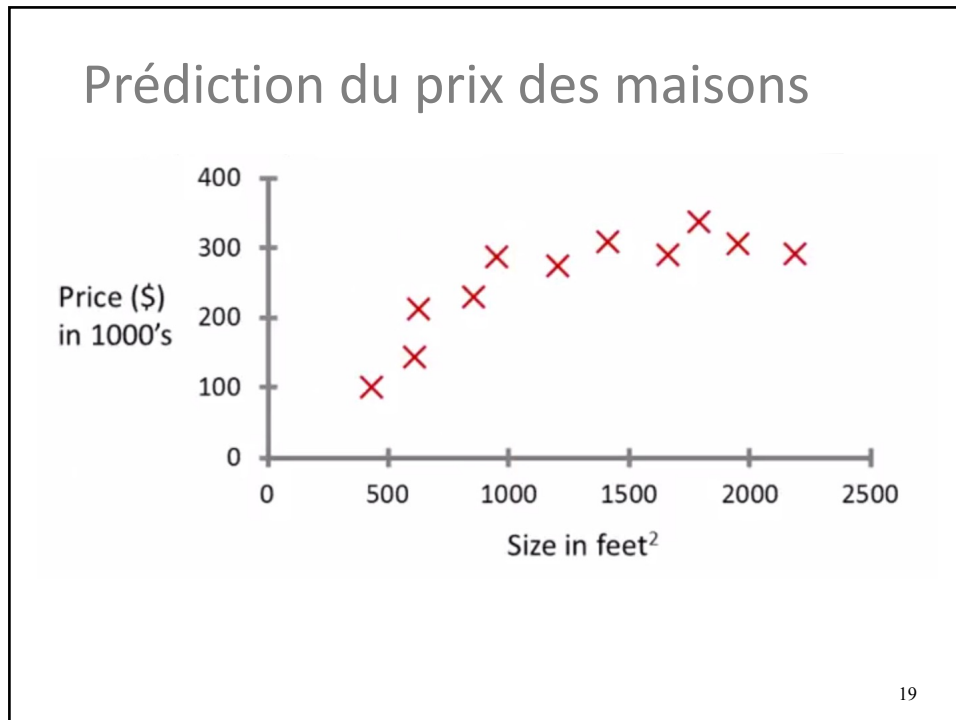


Introduction au Machine Learning

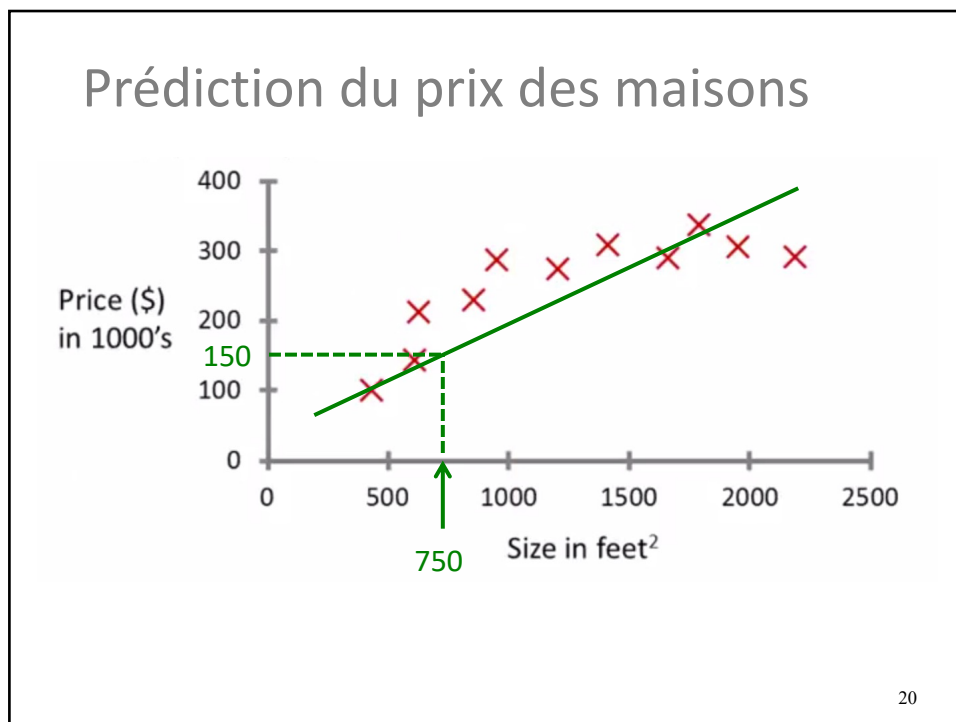
Apprentissage supervisé

18

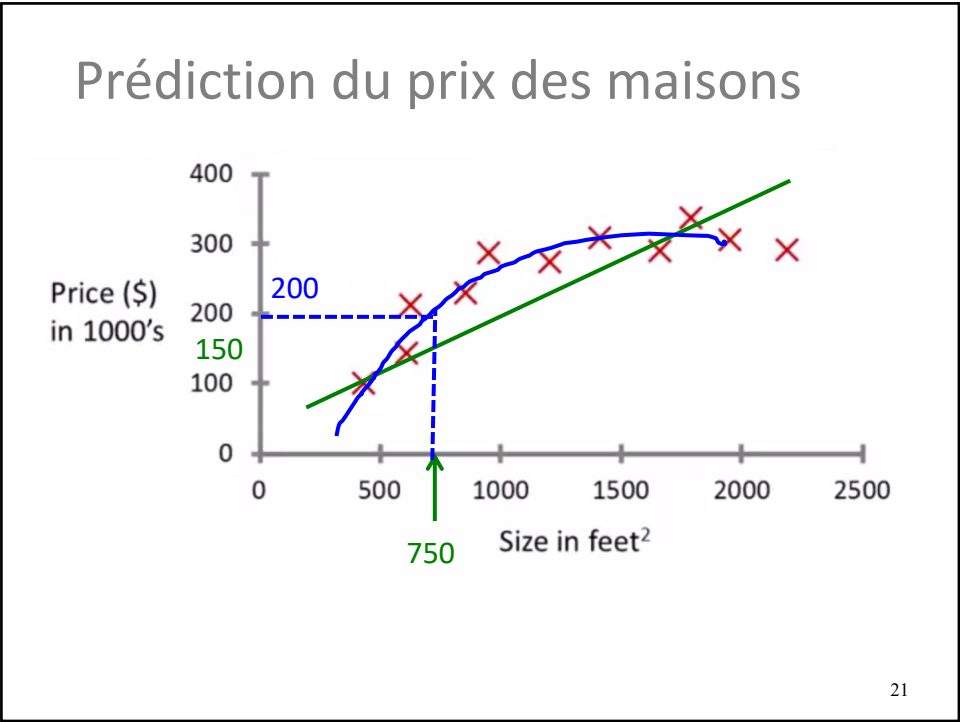
18



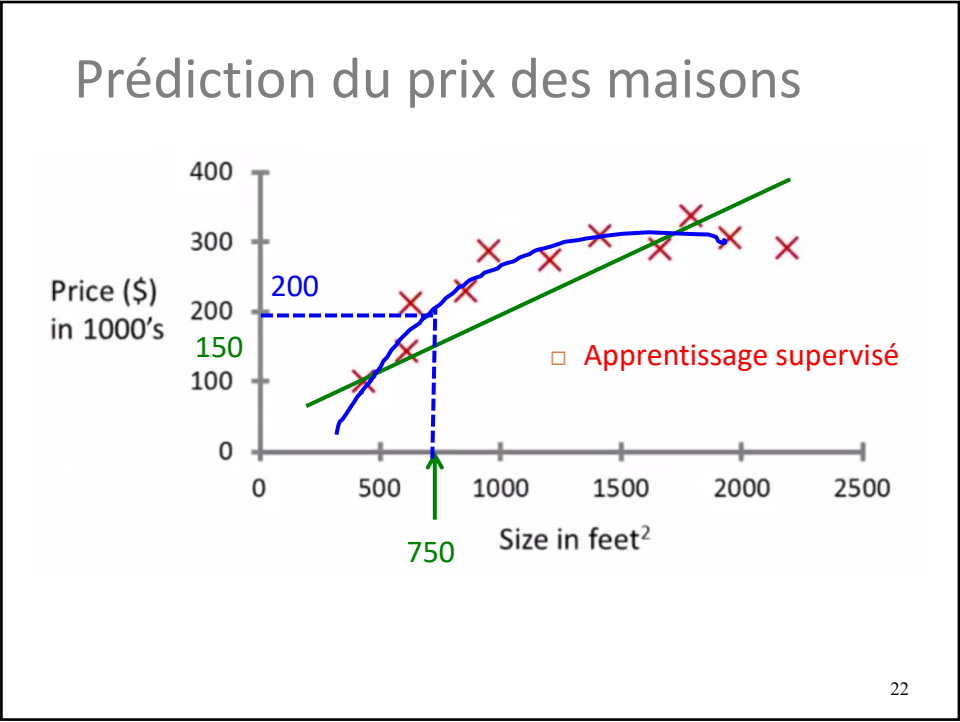
19



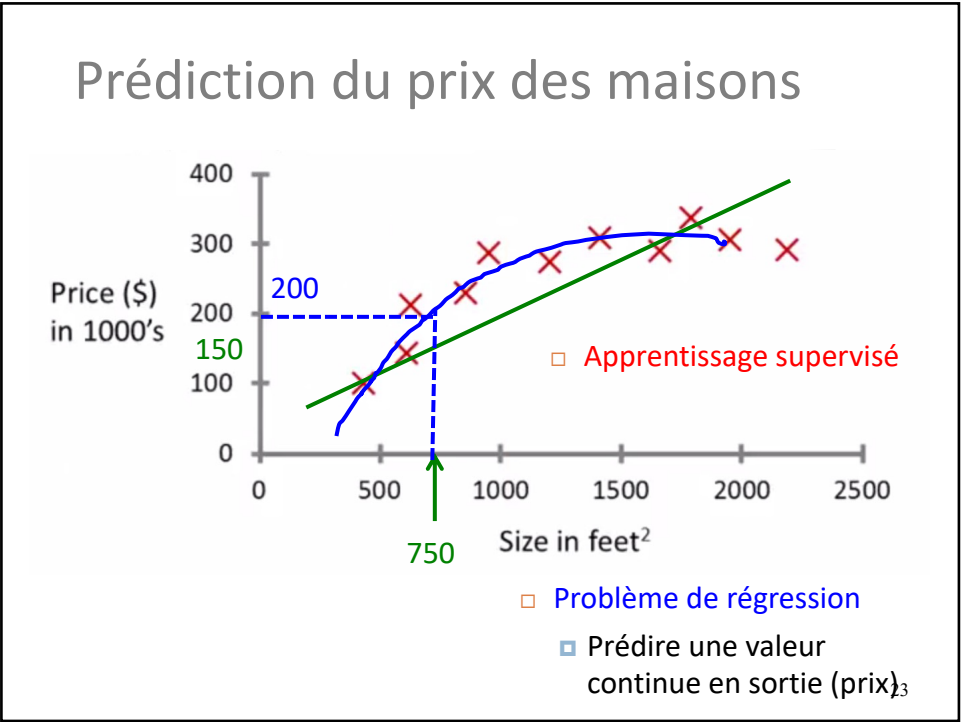
20



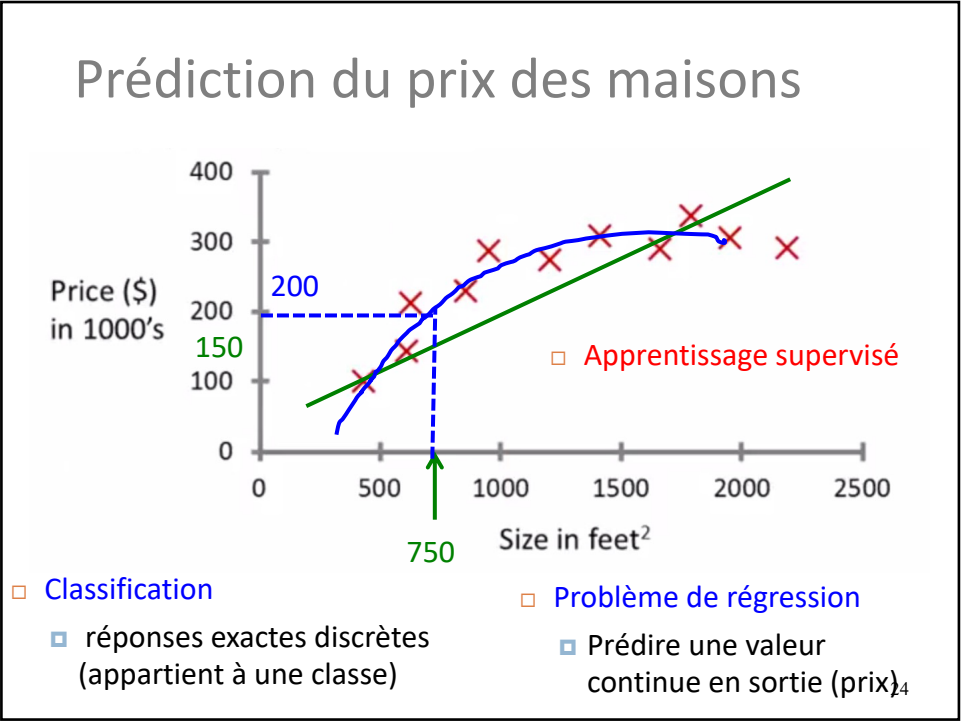
21



22

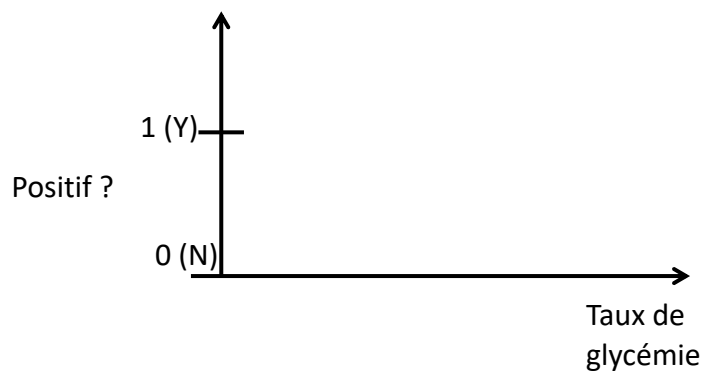


23



24

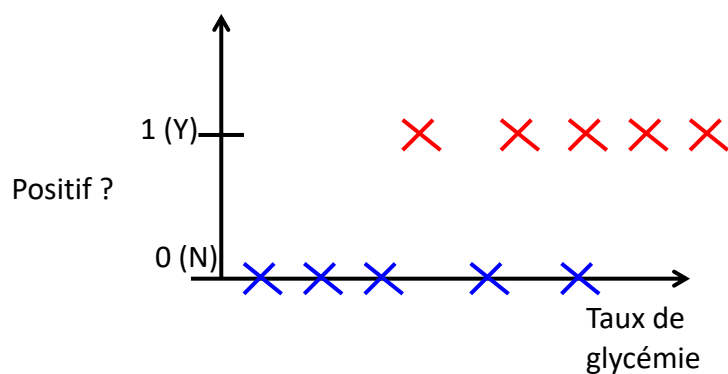
Autre exemple: diabète (positif, négatif)



25

25

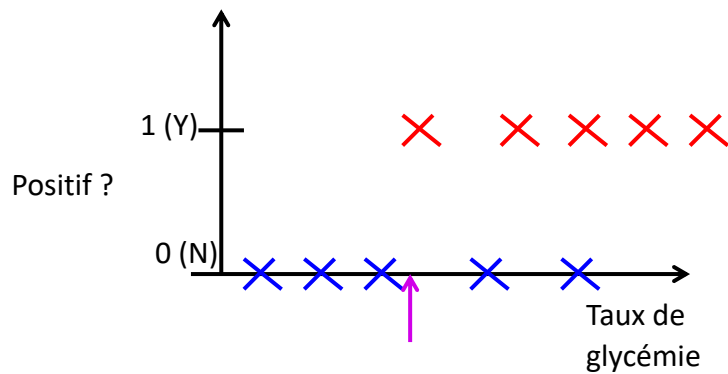
Autre exemple: diabète (positif, négatif)



26

26

Autre exemple: diabète (positif, négatif)

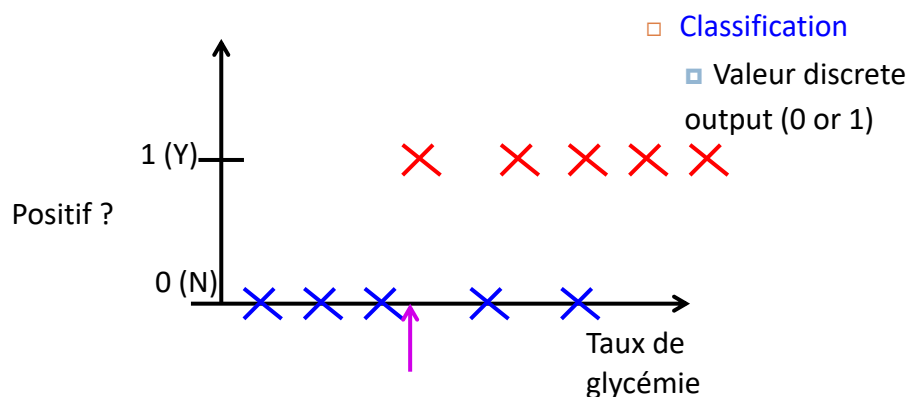


- Si l'on prend en compte le taux de glycémie, quelle est la probabilité d'avoir un diabète ?

27

27

Autre exemple: diabète (positif, négatif)

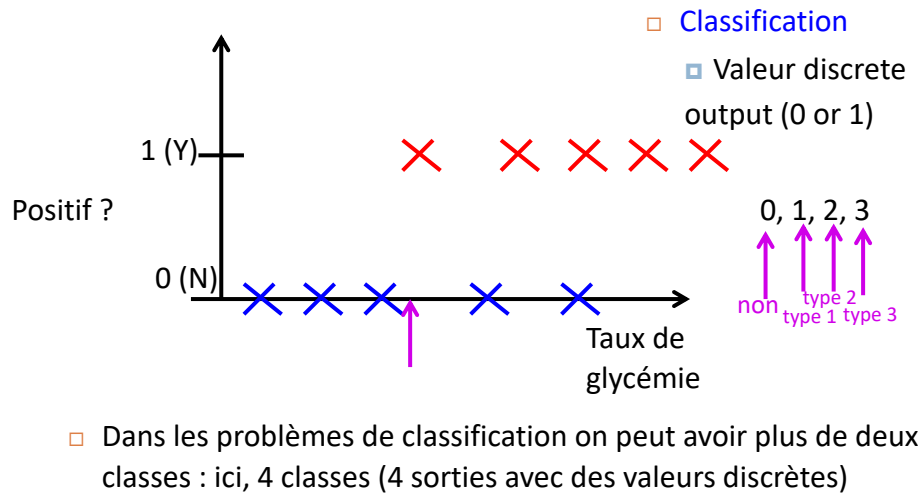


- Si l'on prend en compte le taux de glycémie, quelle est la probabilité d'avoir un diabète ?

28

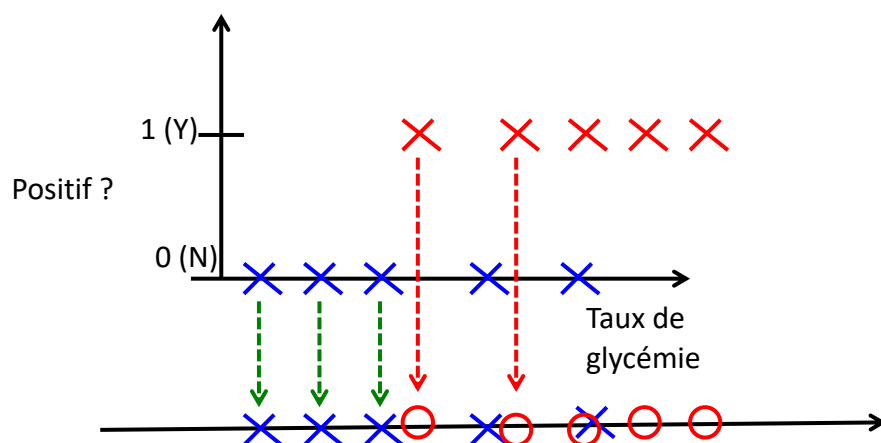
28

Autre exemple: diabète (positif, négatif)



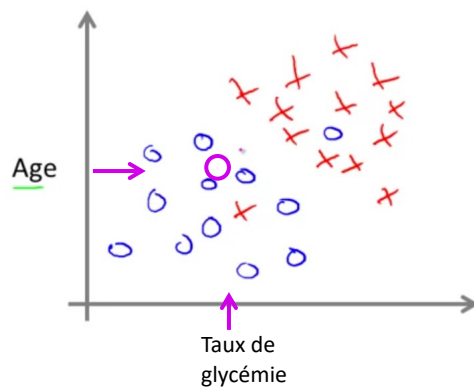
29

Autre exemple: diabète (positif, négatif)



30

Apprentissage supervisé - exemple

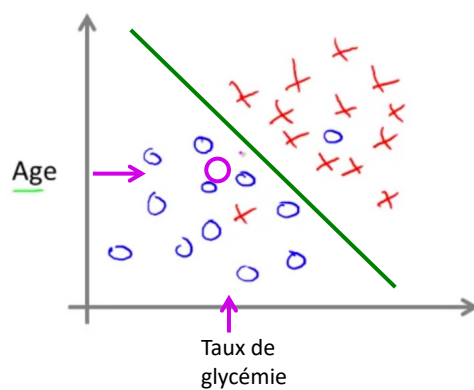


- On rajoute une caractéristique : l'âge du patient

31

31

Apprentissage supervisé - exemple

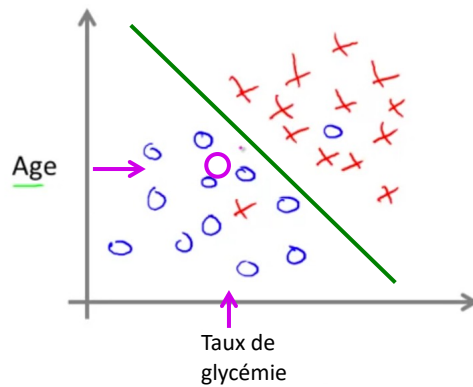


- On rajoute une caractéristique : l'âge du patient

32

32

Apprentissage supervisé - exemple

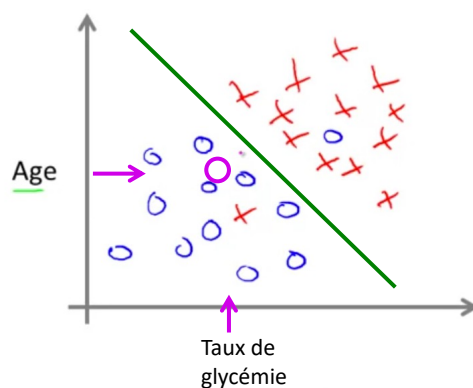


- On rajoute une caractéristique : l'âge du patient
- En fonction du taux de glycémie et de l'âge on veut prédire la probabilité d'avoir un diabète

33

33

Apprentissage supervisé - exemple



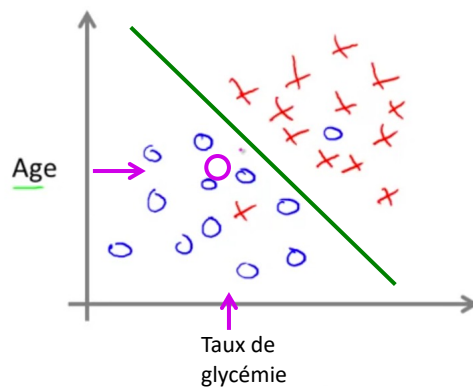
- Autres caractéristiques
 - ▣ Hyper-glycémie
 - ▣ Perte de poids
 - ▣ Altération de la vision
 - ▣ ...

- Deux caractéristiques : l'âge du patient, le taux de glycémie

34

34

Apprentissage supervisé - exemple



- Autres caractéristiques
 - Hyper-glycémie
 - Perte de poids
 - Altération de la vision
 - ...
- Possibilité de travailler avec un ensemble infini de caractéristiques (features)
- Deux caractéristiques : l'âge du patient, le taux de glycémie

35

35

Apprentissage supervisé

- Vous dirigez une entreprise et vous souhaitez développer des algorithmes d'apprentissage automatique pour résoudre les deux problèmes suivants
- **Problème 1 : vous disposez d'un grand inventaire d'articles identiques. Vous voulez prédire combien de ces articles seront vendus au cours des 3 prochaines mois.**
- Est-ce que vous devez traiter ce problème comme un problème de classification ou de régression ?

36

36

Apprentissage supervisé

- Vous dirigez une entreprise et vous souhaitez développer des algorithmes d'apprentissage automatique pour résoudre les deux problèmes suivants
- **Problème 1 : vous disposez d'un grand inventaire d'articles identiques. Vous voulez prédire combien de ces articles seront vendus au cours des 3 prochaines mois.**
- Est-ce que vous devez traiter ce problème comme un problème de classification ou de régression ?
- **Problème 1 : régression**
 - Il s'agit de prédire de manière continue la valeur des articles vendus les 3 prochains mois

37

37

Apprentissage supervisé

- Vous dirigez une entreprise et vous souhaitez développer des algorithmes d'apprentissage automatique pour résoudre les deux problèmes suivants
- **Problème 2 : vous aimeriez avoir un soft pour analyser les comptes individuels de clients et pour chaque compte décider s'il a été hacké / non hacké**
- Est-ce que vous devez traiter ce problème comme un problème de classification ou de régression ?

38

38

Apprentissage supervisé

- Vous dirigez une entreprise et vous souhaitez développer des algorithmes d'apprentissage automatique pour résoudre les deux problèmes suivants
- **Problème 2 : vous aimeriez avoir un soft pour analyser les comptes individuels de clients et pour chaque compte décider s'il a été hacké / non hacké**
- Est-ce que vous devez traiter ce problème comme un problème de classification ou de régression ?
- **Problème 2 : classification**

Apprentissage : associer à chaque compte individuel une étiquette hacké / non hacké

Test : Pour chaque nouveau client, dire s'il a été hacké / non hacké (2 classes)

39

39



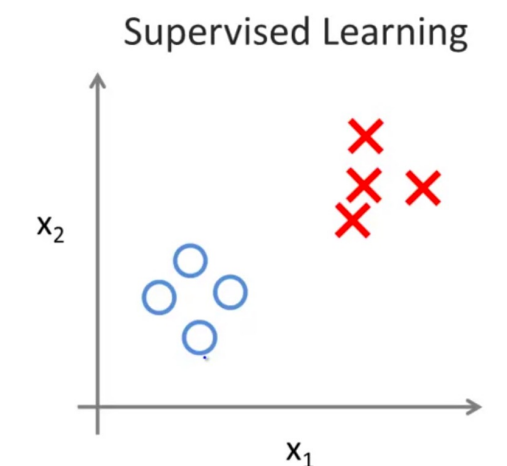
Introduction au Machine Learning

Apprentissage non supervisé

40

40

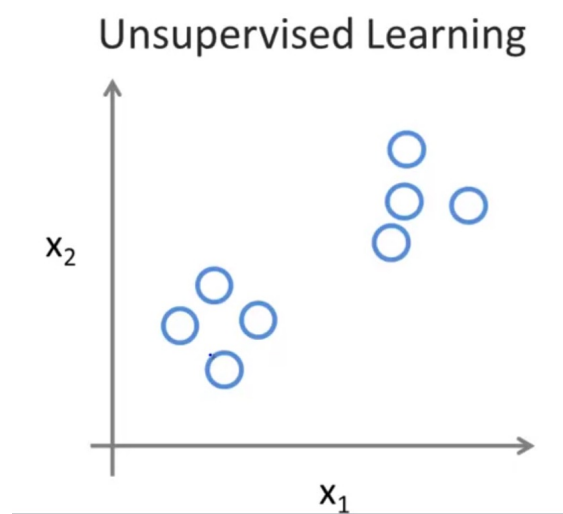
Apprentissage supervisé



41

41

Apprentissage non supervisé

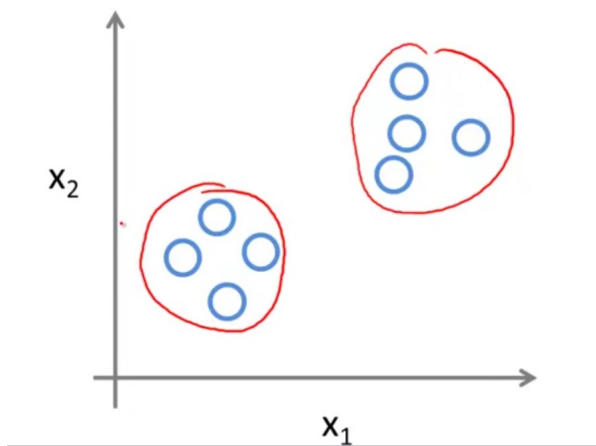


42

42

Apprentissage non supervisé : clustering

Unsupervised Learning




43

43

When Hurricane Maria hit Puerto Rico, 'everything ...
<https://www.washingtonpost.com/national/when-hurricane-maria-hit...>
SAN JUAN, PUERTO RICO — When things went bad during Hurricane Maria, they went bad all at once, across this entire island. Suddenly, everything ...

Category: Hurricanes in Puerto Rico - Wikipedia
https://en.wikipedia.org/wiki/Category:Hurricanes_in_Puerto_Rico
Pages in category "Hurricanes in Puerto Rico" The following 54 pages are in this category, out of 54 total. This list may not reflect recent changes


Reasons to Visit Puerto Rico in Hurricane Season - TripSavvy
<https://www.tripsavvy.com/visiting-puerto-rico-in-hurricane-season...>
Hurricane season isn't the best time to visit the Caribbean, but take a closer look at why Puerto Rico might be an excellent off-season destination.

Images of hurricane porto rico
<bing.com/images>

See more images of hurricane porto rico

Effects of Hurricane Georges in Puerto Rico - Wikipedia
https://en.wikipedia.org/wiki/Effects_of_Hurricane_Georges_in...
Effects of Hurricane Georges in Puerto Rico. The effects of Hurricane Georges in Puerto Rico included \$2 billion in damages and eight fatalities. Hurricane Georges was the first hurricane to cross the entire island since the San Ciprian Hurricane in 1932.^[1] Georges formed on September 15 as Tropical Depression Seven off the African coast. Georges strengthened into a Category 4-hurricane on September 19 as it made landfall in the Lesser Antilles. Georges made landfall on the island on September 21 as a Category 3 hurricane. Georges caused \$2 billion (1996 USD, \$2.7 billion 2010 USD) in damage to the island.

Hurricane Maria: Puerto Rico reels as storm turns toward US East Coast

By Joe Sterling and Dakin Anderson, CNN
Updated 22:53 GMT (06:53 HKT) September 23, 2017



Dominica

Article Talk Read Edit View history Search Wikipedia

Effects of Hurricane Georges in Puerto Rico

From Wikipedia, the free encyclopedia


The effects of Hurricane Georges in Puerto Rico included \$2 billion in damages and eight fatalities. Hurricane Georges was the first hurricane to cross the entire island since the San Ciprian Hurricane in 1932.^[1] Georges formed on September 15 as Tropical Depression Seven off the African coast. Georges strengthened into a Category 4-hurricane on September 19 as it made landfall in the Lesser Antilles. Georges made landfall on the island on September 21 as a Category 3 hurricane. Georges caused \$2 billion (1996 USD, \$2.7 billion 2010 USD) in damage to the island.

Contents [hide]

- 1 Background
- 2 Preparations
- 3 Impact
- 4 Aftermath
- 5 See also
- 6 References

Hurricane Georges

Category 3 major hurricane (SSHWS/NWS)



Hurricane Georges approaching Puerto Rico

Winds	1-minute sustained: 115 mph (185 km/h)
Pressure	968 mbar (hPa); 28.59 inHg
Fatilities	1 direct, 7 indirect
Damage	\$2 billion (1996 USD)
Areas affected	All of Puerto Rico, including Vieques and Culebra.

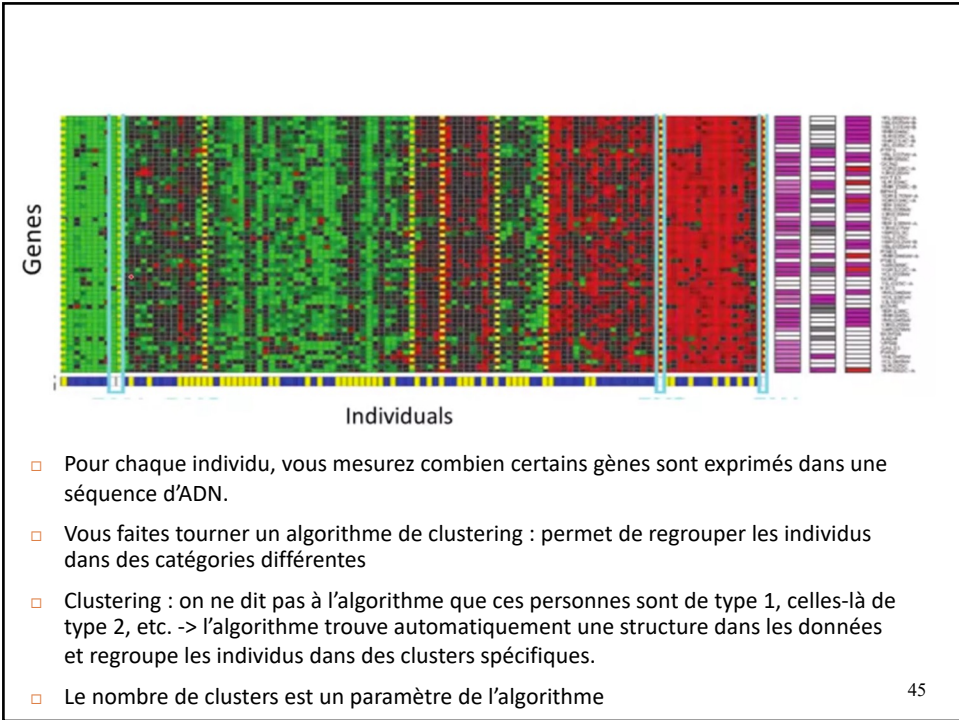
Part of the 1998 Atlantic hurricane season

History

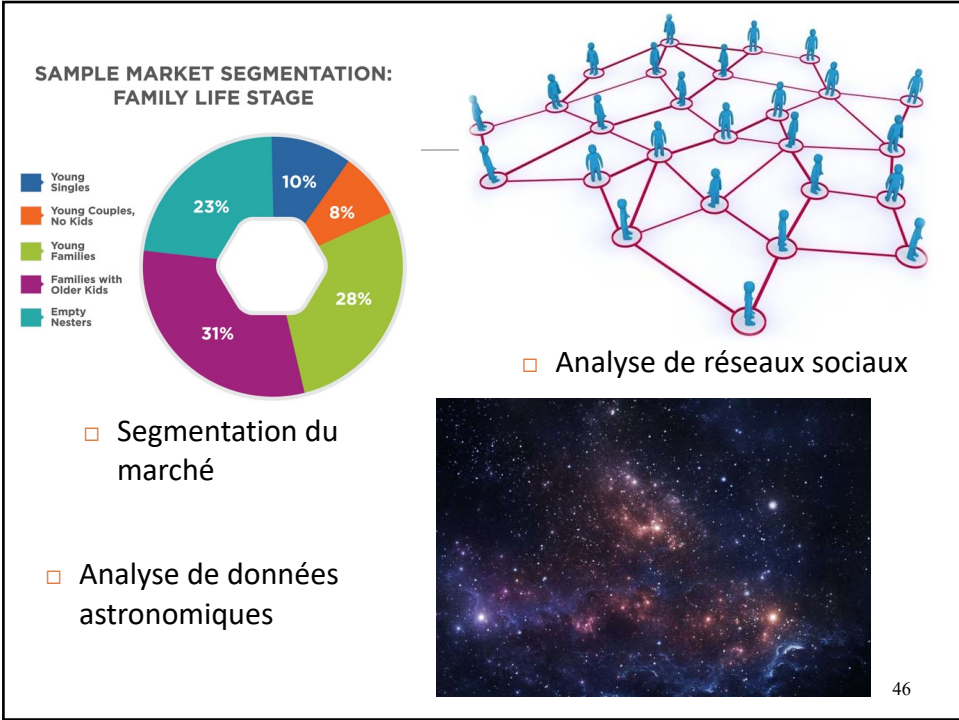
- Meteorological history

Effects

44



45



46

Parmi les exemples suivants, lesquels correspondent à votre avis à des problèmes d'apprentissage non supervisé ?

- A. Étant donné des emails étiquetés spam / non spam, apprendre un filtre de spam
- B. Étant donné un ensemble de nouveaux articles trouvés sur le web, les regrouper dans des sous-ensembles d'articles relatant la même histoire
- C. Étant donnée une base de données clients, découvrir automatiquement les segments du marché et regrouper les clients dans ces différents segments de marché
- D. Étant donné une base de données de patients qui ont été diagnostiqués comme ayant ou pas le covid, dire si de nouveaux patients sont atteints du covid ou pas.

47

47