# Machine Learning & Data Mining

INTRODUCTION

F.BAUDOIN

## Plan

- L'approche Machine learning
  - o apprentissage automatique
- Les applications et les enjeux

# Reconnaître les ours dans des images

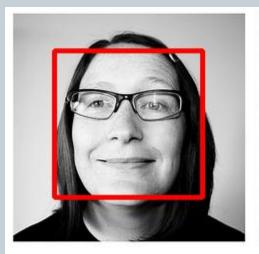


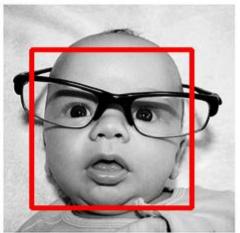




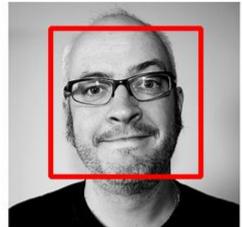


# Reconnaitre les visages (humains)









# Détecter les visages (humains)

#### • Phase 1:

- o Collecter beaucoup d'exemples (d'images) de visages
- o Collecter beaucoup d'exemples de non-visages

#### Phase 2

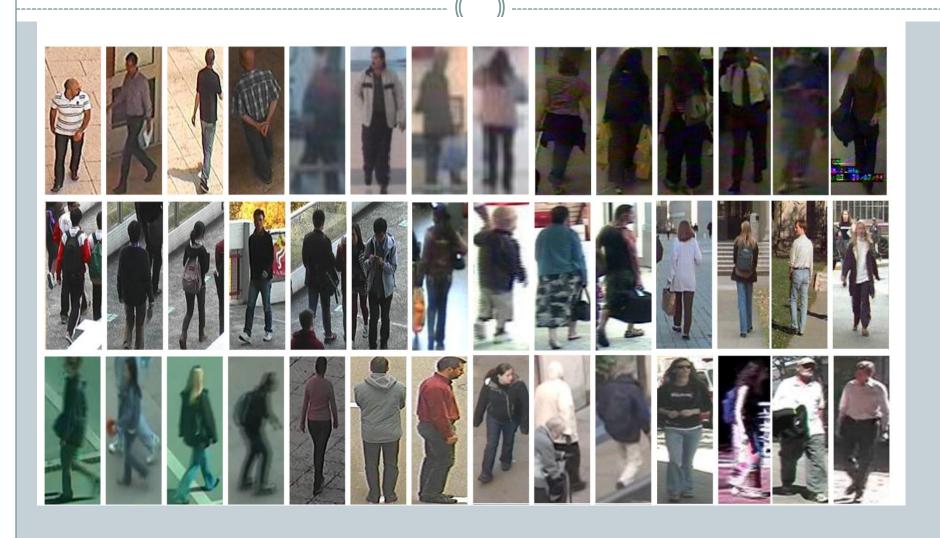
 Construire un classifieur qui prédit si une nouvelle image (image non vue) contient un visage ou non

# Application

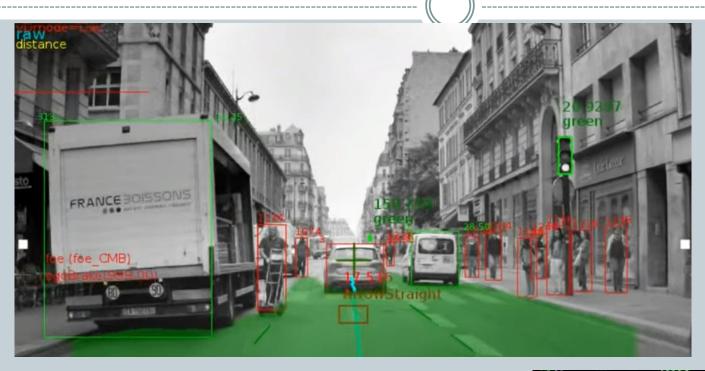
- Voiture intelligente : Conduite autonome (Tesla Mobileye)
  - o Détection de piétons
  - o Détection d'autres véhicules

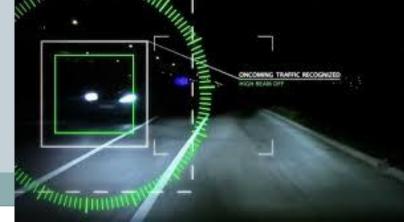


# Exemples positifs de piétons



# Application





# Cherchez les Tufas



## Apprentissage humain

- Une image de Tufa
  - Apprentissage rapide
  - o Features: pieds, tête, ...
  - o Raisonnement inductif, généralisation
- Approche machine Learning
  - Doit extraire les aspects (features)
    - × des données
    - Pour une tâche de prédiction

# Machine learning

## Différentes problématiques

- Prédiction météo (pour anticiper la consommation énergétique, la consommation de ketchup)
- Détection des anomalies (virus en mutation, fraude à la CB, délit d'initié)
- o Classifier (risque d'un crédit, diagnostique du cancer)
- Ordonner (Google, personnalisation)
- o Résumer (News, ...)
- Aide à la décision (IA, robotique, commerce, courtage financier)
- o Reconnaitre (des séquences ADN, des caractères manuscrits,

# Quand utiliser le machine learning?

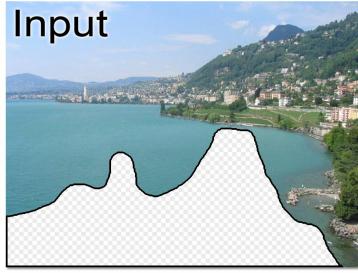
- L'expertise humaine est absente (Navigation sur mars)
- L'expertise humaine est impossible à expliquer, à rationnaliser (reconnaissance de la parole, vision)
- Les solutions évoluent avec le temps (préférences, ...)
- Le problème est trop vaste pour le raisonnement humain (calcul du rang des pages web, publicité sur facebook)

# ML dans le traitement du langage naturel

- Reconnaissance de la parole
- Traduction automatique
- Raison de la réussite : beaucoup de données avec étiquettes (labels)

# Complétion de scène



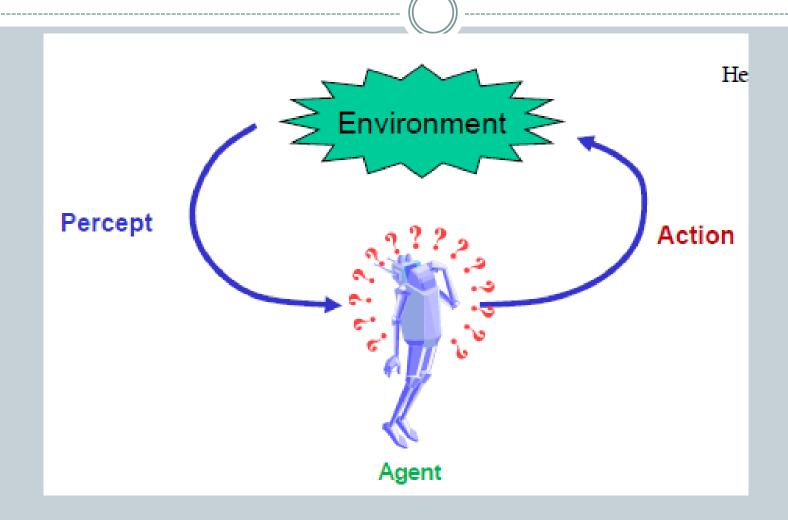




ESGI 2018 -



# Apprentissage



## Apprentissage: définition

#### Herbert Simon

 L'apprentissage est un processus par lequel un système accroit ses performances à partir de l'expérience

#### Tom Mitchell

 Un programme informatique apprend à partir d'une expérience E un ensemble de tâches T avec une performance P si P s'accroît avec E

# Les catégories d'apprentissage

- Apprentissage supervisé
- Apprentissage non supervisé
- Apprentissage par renforcement
- Optimisation continue
- Pourquoi des algorithmes particuliers ?
  - On ne peut pas décrire toutes les situations possibles (explosion combinatoire)

# Apprentissage supervisé

Data = exemples + labels

• Prédiction du label sur un nouvel exemple



 $\rightarrow$  ?

# Apprentissage supervisé

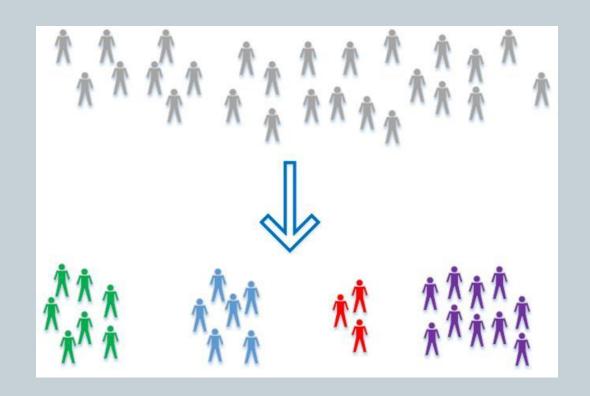
- Un oracle classe les exemples
- L'apprenant apprend a classer comme l'oracle
- Algorithmes
  - O Input : Data = exemples + labels
  - o Sortie: prédiction du label d'un nouvel exemple
- Exemple : reconnaissance de caractères
  - O Data = images + label

# Apprentissage supervisé

- Phase 1 OFFLINE : apprentissage
  - Construction d'un classifieur
    - = une fonction qui prédit la classe d'un exemple
- Phase 2 ONLINE : prédiction
  - Utilisation du classifieur pour prédire la classe d'un exemple non vu

# Apprentissage non supervisé

Exemple : segmenter une population de clients en clusters



# Apprentissage non supervisé

• L'apprenant apprend par lui-même à classer

#### Algorithmes

o Input : Data = exemples

o Sortie : exemples regroupés en clusters

# Les catégories d'apprentissage

## Apprentissage supervisé

- Input : Data = exemples + labels
- o Sortie: prédiction du label d'un nouvel exemple

## Apprentissage non supervisé

- O Input : Data = exemples
- Sortie : des clusters d'exemples

## Optimisation continue

- Input : Data = rien + une fonction C à maximiser
- Output : l'exemple qui maximize C