# Machine Learning

Introduction à l'apprentissage automatique – GIF-4101 / GIF-7005

Professor: Christian Gagné

Week 1

UNIVERSITÉ LAVAL

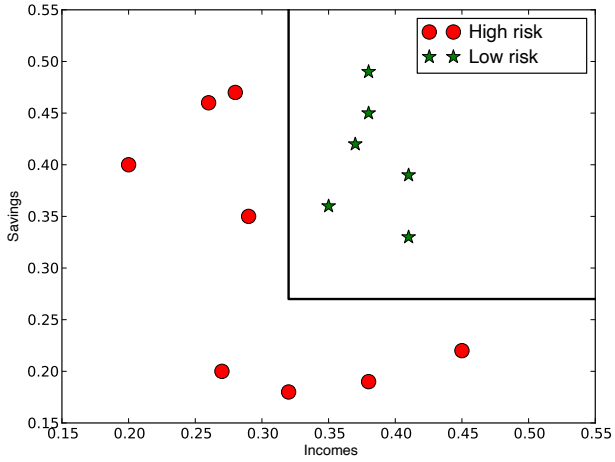
# Why machine learning?

- Machine learning consists in using computers in order to **optimize** an information processing **model** that treats data with regards to some **performance criteria** based on **observations**, whether from examples or past experiences.
- When we know the right **model** to use, there is no need to do learning!
- Machine learning is mostly useful when:
  - We lack the expertise for a specific task (e.g. a robot navigating on Mars)
  - We have an expertise that cannot be explained because it's implicit (e.g. face recognition)
  - The solutions to the problem are changing over time (e.g. packets routing)
  - The solutions must be personalized (e.g. biometrics)

# Example

- A credit business should automatically estimate the risk factor of its customers.
- Available data: client's income (variable $x_1$) and client's savings (variable $x_2$)
- Database filled of previous clients' data: high-risk clients (in red circles) and low-risk clients (in green stars)

# Example




**If** $x_1 > 0.32$ **and** $x_2 > 0.27$ **then** *low-risk* **else** *high-risk*

# Model and observations

- Objective: infer a **general treatment model** from **specific observations**
  - The inferred model must be a good and useful approximation of the observations
- The observations are available in sufficient quantities at an inexpensive cost; knowledge is expensive and rare
- Example: linking consumer transactions to their respective consumption behaviour.
  - Suggestions of similar items on Amazon (books, music), Netflix (movies), etc.

# Perspectives of machine learning

- Optimize the model on the observations with respect to the performance criteria.
- **Statistic's perspective**: inference from samples
- **Computer science's perspective**: implement algorithms and create efficient representations in order to build and evaluate models
- **Engineering's perspective**: solve problems without having to manually specify or specialize the models

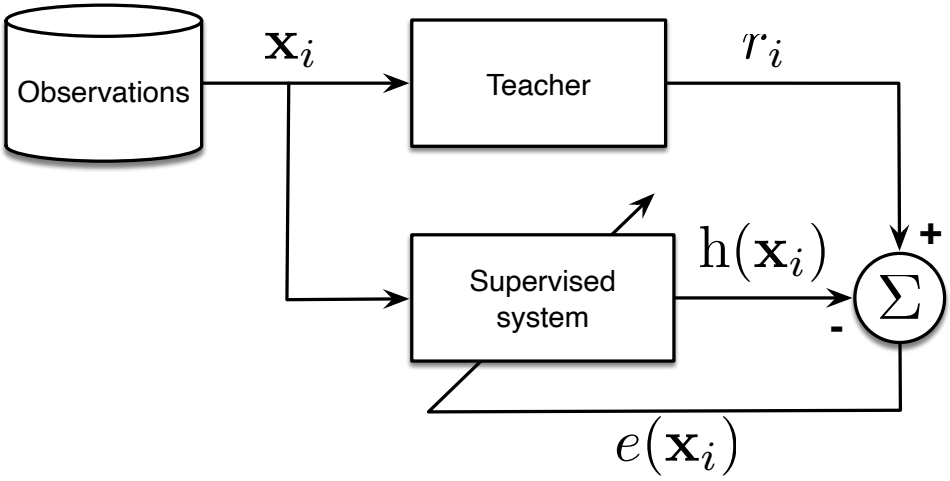
# Learning associations

- Analysis of a grocery cart
  - $P(Y|X)$ is the probability that a client who buys a product $X$ also buys $Y$, where $X$ and $Y$ are products or services
- Example: The probability that "beer" is selected knowing that "chips" already is: $P(chips|beer) = 0.7$

# Supervised learning

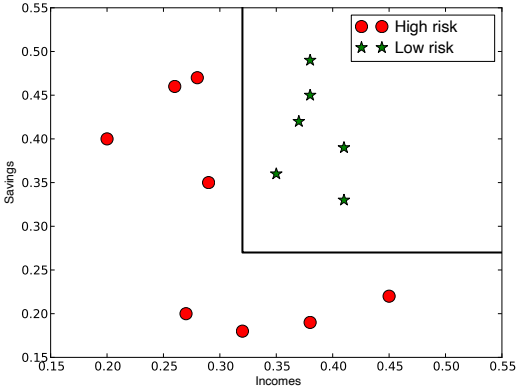
- Supervised learning
  - Objective: learn a projection between the input observations $X$ and the associated $Y$ output values
- Mathematical modelling
  - $y = \mathrm{h}(x|\theta)$
  - $\mathrm{h}(\cdot)$: general function of the model
  - $\theta$: model's parameter

# Supervised learning schema



# Classification

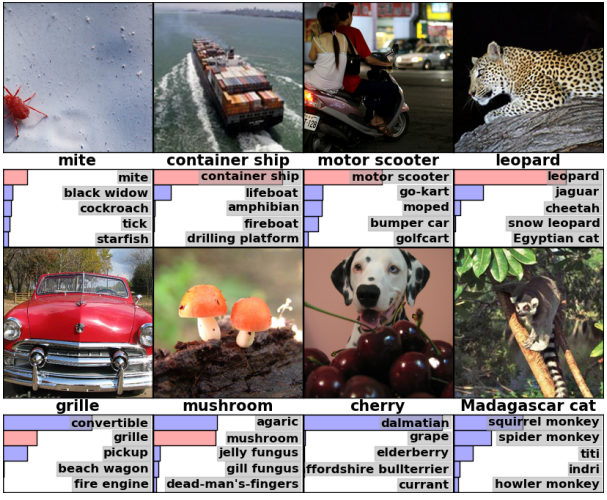
- $Y$ is discrete and corresponds to the class labels
- $\mathrm{h}(\cdot)$ is a discriminating function



# Classification application

- Pattern recognition
  - Objects recognition: recognize objects types that are present in an image even if the position or the pose of the objects are varying
  - Handwritten character recognition: recognize the characters despite the different styles of writing
  - Speech recognition: Time dependence of information, use valid words/structure dictionaries
- Natural language processing
- Medical diagnostic assistance
- Drugs discovery
- Biometrics
- Etc.

# Objects recognition

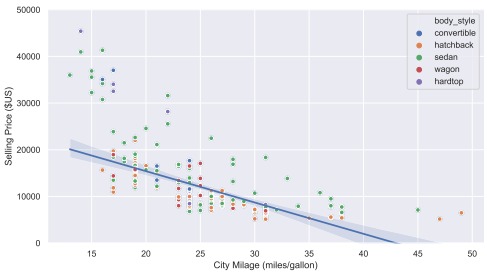



Taken from *A. Krizhevsky, I. Sutskever and G.E. Hinton. Imagenet Classification with Deep Convolutional Neural Networks. In Advances in Neural Information Processing Systems, 2012.*

# Characters recognition

# Regression

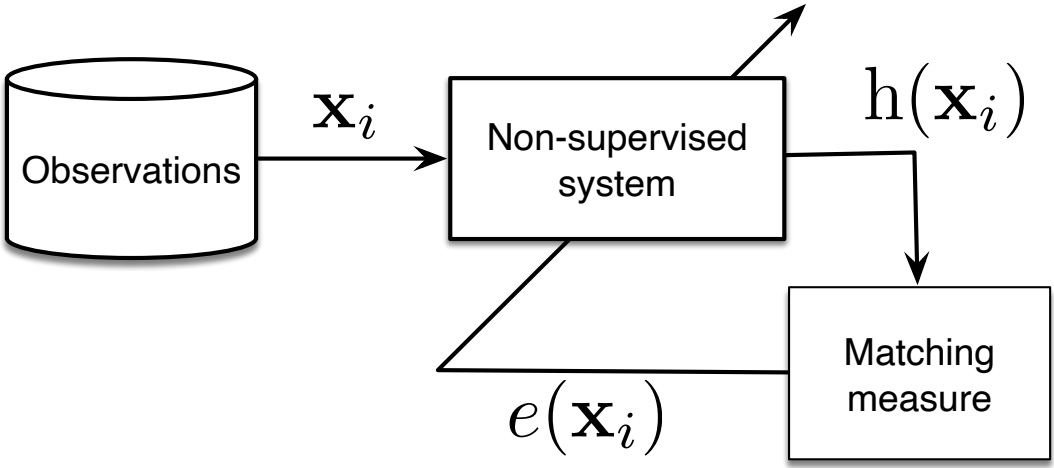
- $Y$ is a real value
- $\mathrm{h}(\cdot)$ is the regression function
- Example: prediction of the sale price of a used car based on the mileage travelled
  - Observations: mileage travelled ($x$)
  - Prediction: Sale price ($y$)
- Values prediction application
  - Finance and insurance
  - Natural phenomena (e.g. weather)
  - Offer and demand
- Risk and uncertainty assessment



# Unsupervised learning

- Unlike supervised learning, there are no output values
- Objective: discover regularities in the observations
  - *Clustering*: discover clusters of similar observations
- Applications
  - Segmentation of the users in a purchasing database
  - Bioinformatic: discover patterns in DNA
  - Image segmentation: define coherent regions of images

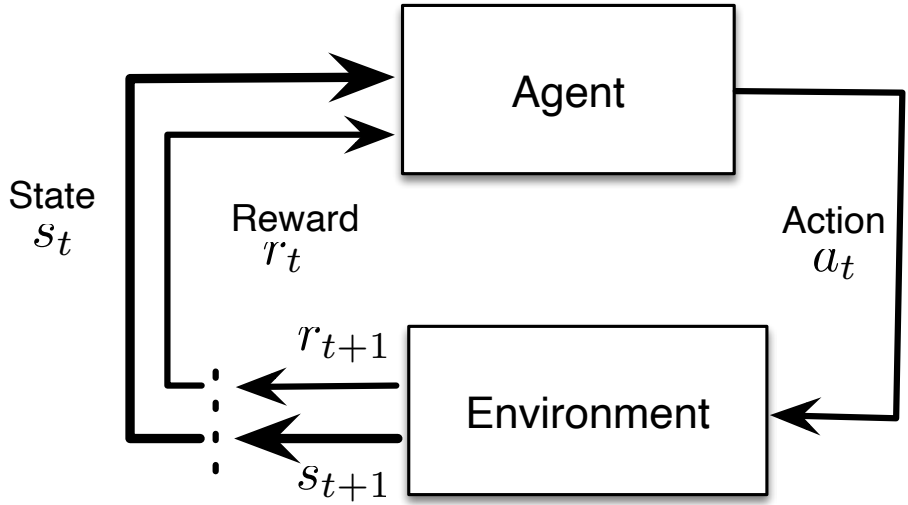
# Unsupervised learning schema



# Reinforcement learning

- Learn a policy: a state to action mapping that leads to a reward
- The learning is not supervised; a reward is given, but with a delay
- Problem of credit assignment: which sequence of actions has led to a reward?
- Applications
  - Games, with one or many players
  - Robotic: navigation within an environment
  - Agents: decision-making

# Reinforcement learning schema



# Data and challenges

- UCI Machine Learning Repository: `http://archive.ics.uci.edu/ml/`
- Kaggle
  - Challenges: `https://www.kaggle.com/competitions`
  - Databases: `https://www.kaggle.com/datasets`
- ImageNet: `http://www.image-net.org/`
- COCO (Common Objects in Context): `http://cocodataset.org/`
- Open data
  - USA: `https://www.data.gov/`
  - Europe: `http://data.europa.eu/euodp/fr/data/`
  - Canada: `http://ouvert.canada.ca/fr/donnees-ouvertes`
  - Quebec: `https://www.donneesquebec.ca`