MA2823: Foundations of Machine Learning

École Centrale Paris — Fall 2016

Chloé-Agathe Azencott
Centre for Computational Biology, Mines ParisTech
chloe-agathe.azencott@mines-paristech.fr

Inserm
Institut national
de la santé et de la recherche médicale

institutCurie

MINES
ParisTech

- **TAs:**
  - **Benoît Playe** benoit.playe@mines-paristech.fr
  - **Mihir Sahasrabudhe** mihir.sahasrabudhe@centralesupelec.fr

- **Course material & contact**

  http://tinyurl.com/ma2823-2016

  https://github.com/chagaz/ma2823_2016

  chloe-agathe.azencott@mines-paristech.fr

Slides thanks to Ethem Alpaydi, Matthew Blaschko, Trevor Hastie, Rob Tibshirani and Jean-Philippe Vert.

What is (Machine) Learning?

Why Learn?

- **Learning**: Modifying a behavior based on experience. [F. Benureau]
- **Machine learning**: Programming computers to optimize a performance criterion using example data.
- There is no need to "learn" to calculate payroll.
- Learning is used when
  - Human expertise does not exist (bioinformatics);
  - Humans are unable to explain their expertise (speech recognition, computer vision);
  - Solutions change in time (routing computer networks);
  - Solutions need adapting to new cases (user biometrics).

Artificial Intelligence

Machine Learning is a subfield of **Artificial Intelligence:**

- A system that lives in a **changing environment** must have the ability to **learn** in order to **adapt.**
- ML algorithms are building blocks that make computers behave more intelligently by **generalizing** rather than merely storing and retrieving data (like a database system would do).

What we talk about when we talk about learning

- **Learning general models from particular examples** (data)
  - **Data** is (mostly) cheap and abundant;
  - **Knowledge** is expensive and scarce.
- Example in retail:

  From customer transactions to consumer behavior

  People who bought “Game of Thrones” also bought “Lord of the Rings” [amazon.com]
- Goal: **Build a model that is a good and useful approximation to the data.**

Data mining: Applying ML to (large) databases

- **Retail:** Market basket analysis, Customer relationship management (CRM).
- **Finance:** Credit scoring, fraud detection.
- **Manufacturing:** Control, optimization, troubleshooting.
- **Medicine:** Medical diagnosis.
- **Telecommunications:** Spam filters, intrusion detection, network optimization, routing.
- **Science:** Analyze large amounts of data in physics, astronomy, biology.
- **Web:** Search engines.

What is machine learning?

- **Optimizing a performance criterion using example data or past experience.**
- Role of **Statistics:**

  Build mathematical models to make inference from a sample.
- Role of **Computer Science:** Efficient algorithms to
  - Solve the optimization problem;
  - Represent and evaluate the model for inference.

Classes of machine learning problems

- **Association rule learning:** Discover relations between variables
- **Supervised learning:** Predict **outcome** from **features**
  - Classification
  - Regression
  - Ranking
  - Ordered categories (e.g. scores)
- **Unsupervised learning:** Find **patterns** in the data
  - Dimensionality reduction
  - Clustering
- **Semi-supervised learning:** Predict outcome for unlabeled but known instances
- **Reinforcement learning:** Maximize **cumulative reward**

Learning associations

Market basket analysis:

X, Y: products/services

P (Y | X): probability that somebody who buys X also buys Y.

Example:

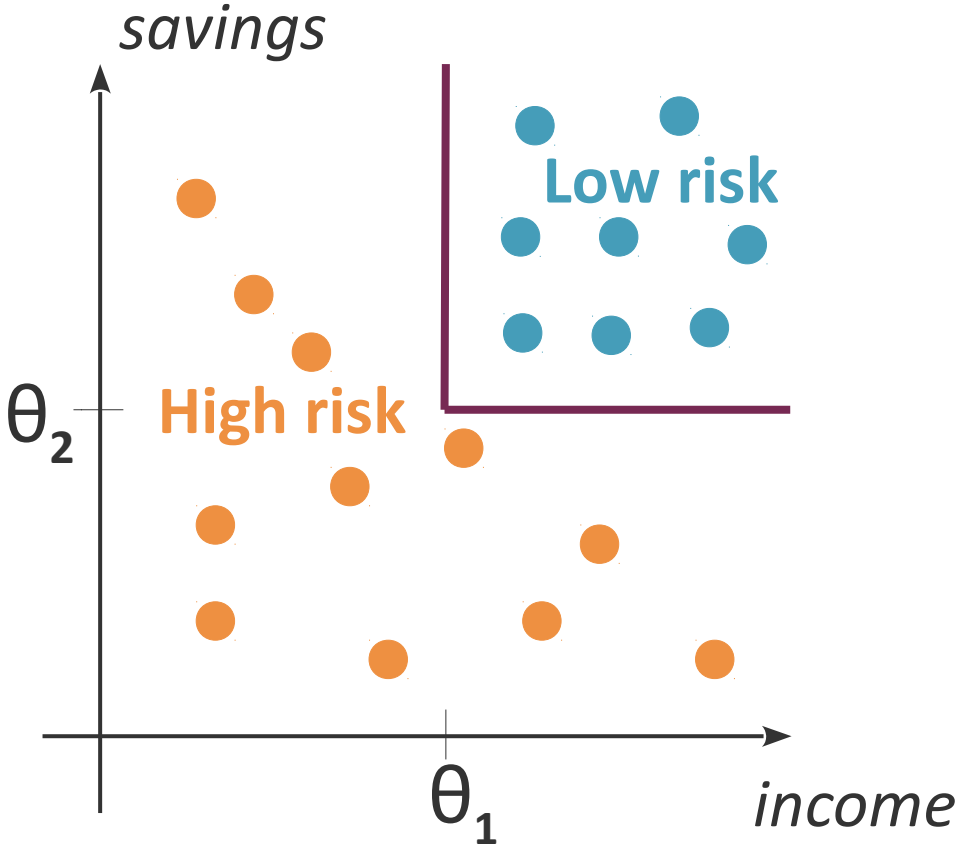
P (chips | beer) = 0.70

P (chips | beer, M, age < 30) = 0.85

Classification (or pattern recognition)

Example: **Credit scoring**

- Differentiate between **low-risk** and **high-risk** customers, from their income and savings



Discriminant: IF *income* > θ_1 AND *savings* > θ_2
THEN low-risk ELSE high-risk

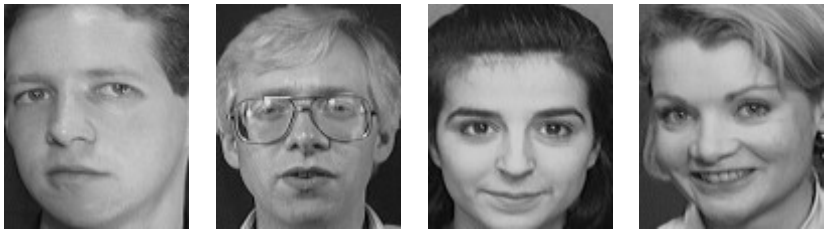
Classification: Applications

- **Face recognition:** independently of pose, lighting, occlusion (glasses, beard), make-up, hair style.
- **Character recognition:** independently of different handwriting styles.
- **Speech recognition:** account for temporal dependency.
- **Medical diagnosis:** from symptoms to illnesses.
- **Precision medicine:** from clinical & genetic features to diagnosis, prognosis, response to treatment.
- **Biometrics:** recognition/authentication using physical or behavioral characteristics: Face, iris, signature...

Face recognition example

- Training examples (one person)

- Test images

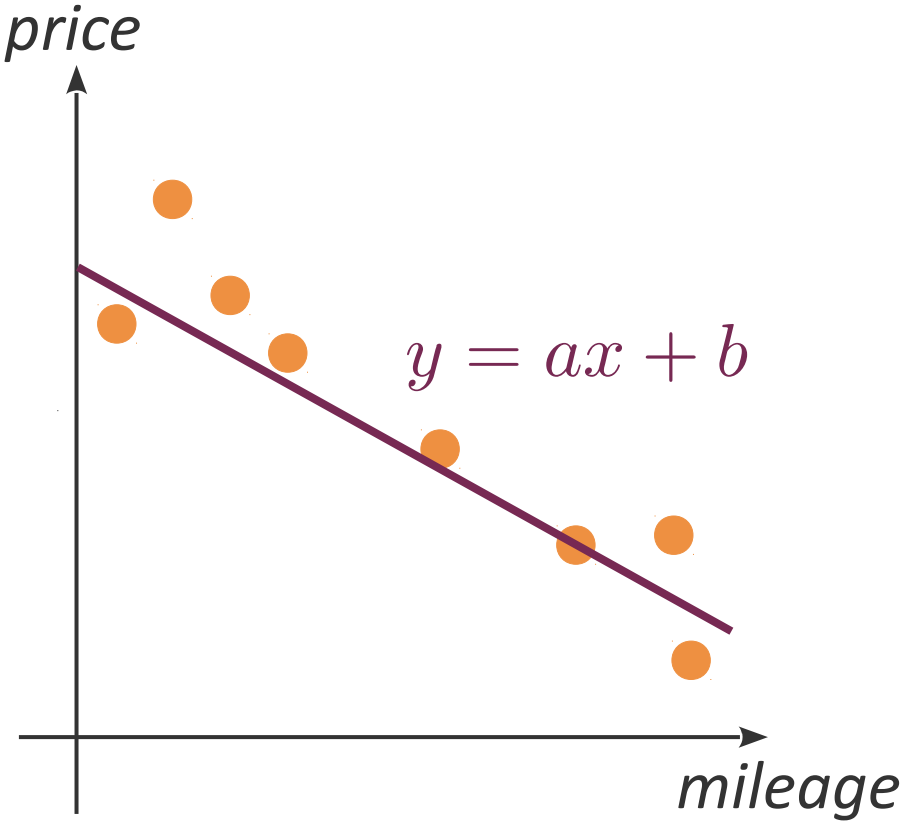


ORL dataset, AT&T Labs, Cambridge (UK).

Regression

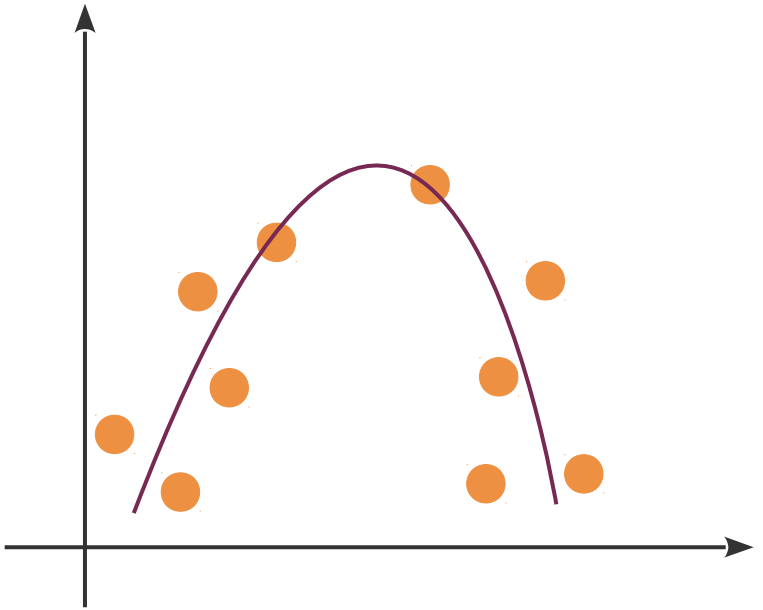
Example: **Price of a used car**

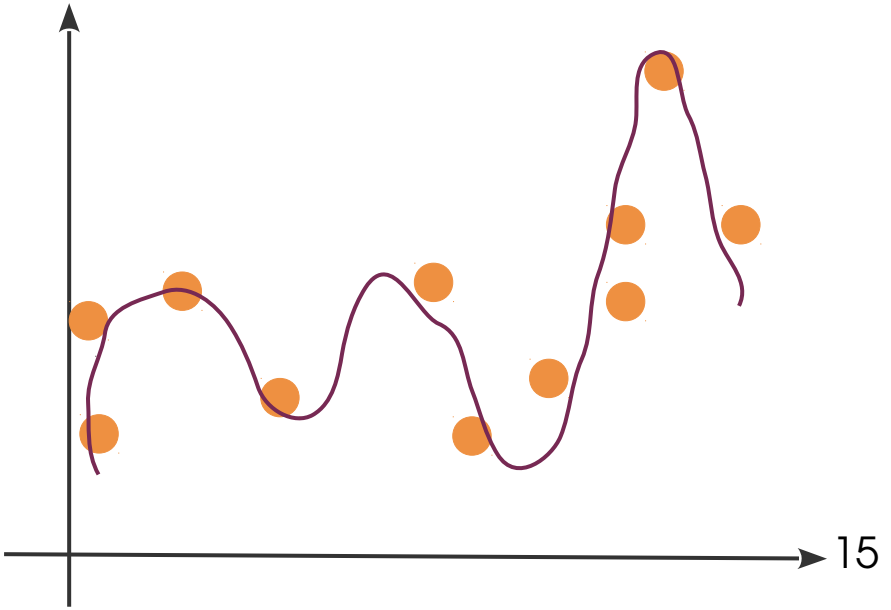
- x: milage
- y: price
- y = f(x|θ)
  - f = model
  - θ = parameters
  - θ = (a, b)



Regression: Applications

- Car navigation: angle of steering
- Kinematics of a robot arm
- Binding affinities between molecules
- Age of onset of a disease
- Solubility of a chemical in water
- Yield of a crop
- Direction of a forest fire





Ranking

- Find a function that **orders** a given list of items of X
- As a classification problem: **pairwise approach**
  - Is {x1, x2} correctly ordered?
- As a regression problem: **pointwise approach**
  - $f(x)$ such that $x_1 \prec x_2 \Leftrightarrow f(x_1) \prec f(x_2)$
- **Listwise approach:** directly optimize for the best list.

Ranking: applications

- **Chemoinformatics:**
  - virtual screening
  - scoring 3D structures.
- **Recommander systems:**
  - rank movies according to how likely a user is to view them.
  - rank items according to how likely a consumer is to buy them.
- **Information retrieval:**
  - document retrieval
  - web search.

Uses of supervised learning

- **Prediction** of future cases:
  - Use the rule to predict the output for new inputs.
- **Knowledge extraction:**
  - Interpret the rule.

    Assumes the rule is easy to understand.
- **Compression:**
  - The rule is simpler than the data it explains.
- **Outlier detection:**
  - Find exceptions that are not covered by the rule (they might be fraud, intrusion, errors).

Unsupervised learning

- No output
- Learn “what normally happens”

- **Density estimation:** Find the structure (regularities) in the data
- **Clustering**
- **Dimensionality reduction.**

Clustering

- Goal: **Group objects** into **clusters**, i.e. classes that were unknown beforehand.
- Objects in the same cluster are "**more similar**" to each other than to objects in a different cluster.
- Motivation:
  - Understanding **general characteristics** of the data;
  - **Visualization**;
  - **Infering some properties** of an object based on how it relates to other objects.

Clustering: applications

- **Customer relationship management:** Customer segmentation
- **Image compression:** Color quantization
- **Document clustering:** Group documents by topics (bag-of-words)
- **Bioinformatics:** Learning motifs.

Dimensionality reduction

Goal: Reduce the number of input variables

- **Feature selection:**

  Only keep the **features** (= variables) that are relevant

- **Feature extraction:**

  Transform the data into a space of lower dimension.

  Goals:

  - Reduce storage **space** & computational **time**
  - Remove **colinearities**
  - **Visualization** (in 2 or 3 dimensions) and **interpretability**.

Reinforcement learning

- Output = sequence of actions
- Learn a **policy:** sequence of correct actions to reach the goal
- No supervised output but **delayed reward**
- Examples
  - Game playing
  - Robot in a maze
- Issues: multiple agents, partial observability, ...

Artificial intelligence

Electrical engineering

Signal processing

Pattern recognition

Engineering

Optimization

Knowledge discovery
in databases

Computer science

Inference

Data mining

Discriminant analysis

Big data

Business

Statistics

Data science

Induction

Learning objectives

After this course, you should be able to

- **Identify problems** that can be solved by machine learning;
- Given such a problem, **identify** and **apply** the most appropriate classical algorithm(s);
- **Implement** some of these algorithms yourself;
- **Evaluate** and **compare** machine learning algorithms for a particular task.

Course Syllabus

- Sep 7
  - 1. Introduction
  - **2. Supervised learning**
- Sep 14
  - **3. Model evaluation & selection**
  - Lab: scientific Python
- Sep 21
  - **4. Bayesian decision theory**
  - Lab: scikit-learn
- Sep 30
  - **5. Linear and logistic regression**
  - Lab: Intro to Kaggle challenge
- Oct 7
  - **6. Regularized linear regression**
  - Lab: Regularized linear regression

Parametric methods

- Oct 14
  - **7. Nearest neighbor methods**
  - Lab: Nearest neighbor methods
- Nov 4
  - **8. Tree-based methods**
  - Lab: Tree-based methods
- Nov 18
  - **9. Support vector machines**
  - **Guest talk:** Beyrem Khalfaoui
- Nov 25
  - **10. Neural networks**
  - Lab: Support vector machines
- Dec 04
  - **11. Dimensionality reduction**
  - Lab: Dimensionality reduction
- Dec 11
  - **12. Clustering**
  - Lab: Kaggle challenge

Non-parametric methods

Unsupervised Learning

kaggle challenge project

How Many Bikes? Challenge

`https://www.kaggle.com/c/how-many-bikes`

- **Predict the number of shared bikes** that are rented in an American city
  - Regression
  - From weather, holiday, date & time.
- **Evaluation on**
  - Insights learned
  - Prediction performance.

Evaluation

- **Final exam (60 pts)** — **December 16, 2016**
  - Pen and paper
  - Closed book
- **Kaggle project (30 pts)** — **December 16, 2016**
  - Written report (25 pts)
  - Position in the leaderboard (5pts)
  - Introduction: September 30, 2016
- **Homework (10 pts)** — **1 problem each week**
  - To get the points: turn it in!
- **Scribe extra credit (5 pts)** — **once in the course**
  - Write and share lecture notes.

Scribes

- **What** are scribes?
  - 2-4 students / week
  - Create & share **written notes** from the lecture of that week.
- **Why** be a scribe?
  - **Learn** by focusing on one chapter
  - **Common good:** benefit from the notes taken by others
  - Experience using **LaTeX** and **github**
  - Extra credit.
- **Github repository** `https://github.com/chagaz/ma2823_2016`
  - See example of Chap. 2
- **When?** `https://framadate.org/omVzzIPfaHHgm881`
  - Sign up at least the day before class
  - Turn in notes at most the day before class.

Homeworks

- **One problem per week**
  - Similar to the questions you'll be asked at the exam
  - Turn it in at the beginning of the lecture
  - Solution will be posted after the lecture
  - Worth 1pt if you turn it in.

Resources

- **Course website**

  `http://tinyurl.com/ma2823-2016`

  - **Syllabus**
  - **2 days before the lecture:** printable lecture handout
  - **Shortly after the lecture:**
    - HW Problem n+1
    - Lecture slides
    - HW Solution n.
- **GitHub repository**

  `https://github.com/chagaz/ma2823_2016`

  - **Lab notebooks**
  - **Lecture notes.**

Resources: Datasets

- **UCI Repository:**
  `http://www.ics.uci.edu/~mlearn/MLRepository.html`
- **KDnuggets Datasets:**
  `http://www.kdnuggets.com/datasets/index.html`
- **ImageNet:** `http://www.image-net.org/`
- **Enron Email Dataset:** `http://www.cs.cmu.edu/~enron/`
- **Million Song Dataset:**
  `http://labrosa.ee.columbia.edu/millionsong/`
- **IMDB Data:** `http://www.imdb.com/interfaces`
- **Données publiques françaises:** `https://www.data.gouv.fr/`
- **TunedIT:** `http://www.tunedit.org/`
- **Knoema:** `https://knoema.com/`

Resources: Journals

- **Journal of Machine Learning Research** http://jmlr.csail.mit.edu/
- **IEEE Transactions on Pattern Analysis and Machine Intelligence** https://www.computer.org/portal/web/tpami
- **Annals of Statistics** http://imstat.org/aos/
- **Journal of the American Statistical Association** http://www.tandfonline.com/toc/uasa20/current
- **Machine Learning** http://link.springer.com/journal/10994
- **Neural Computation** http://www.mitpressjournals.org/loi/neco
- **Neural Networks** http://www.journals.elsevier.com/neural-networks
- **IEEE Transactions on Neural Networks and Learning Systems** http://cis.ieee.org/ieee-transactions-on-neural-networks-and-learning-systems.html

Resources: Conferences

- International Conference on Machine Learning (ICML) http://www.icml.cc/
- Neural Information Processing Systems (NIPS) http://www.nips.cc/
- International Conference on Learning Representations (ICLR) http://www.iclr.cc/
- European Conference on Machine Learning (ECML) http://www.ecmlpkdd.org/
- International Conference on AI & Statistics (AISTATS) http://www.aistats.org/
- Uncertainty in Artificial Intelligence (UAI) http://www.auai.org/
- Computational Learning Theory (COLT) http://www.learningtheory.org/past-conferences-2/
- Knowledge Discovery and Data Mining (KDD) http://www.kdd.org/
- International Conference on Pattern Recognition (ICPR) http://www.icpr2016.org/