Advanced Statistical Inference
Introduction

Maurizio Filippone
Maurizio.Filippone@eurecom.fr

Department of Data Science
EURECOM

Why do we call it Bayesian?

- Reverend Thomas Bayes (London 1701 – Kent 1761)
- Logic and theology degree from University of Edinburgh in 1722

Why do we call it Bayesian?

- Published works
  - "Divine Benevolence, or an Attempt to Prove That the Principal End of the Divine Providence and Government is the Happiness of His Creature" in 1731
  - "An Introduction to the Doctrine of Fluxions, and a Defence of the Mathematicians Against the Objections of the Author of The Analyst" in 1736

Why do we call it Bayesian?

- Published works
  - "Divine Benevolence, or an Attempt to Prove That the Principal End of the Divine Providence and Government is the Happiness of His Creature" in 1731
  - "An Introduction to the Doctrine of Fluxions, and a Defence of the Mathematicians Against the Objections of the Author of The Analyst" in 1736
- Thanks to Richard Price: "An Essay towards solving a Problem in the Doctrine of Chances" read to the Royal Society in 1763

Probabilities before Bayes

- "The Doctrine of Chances" (1718) by Abraham de Moivre (1667–1754)

Probabilities before Bayes

- "The Doctrine of Chances" (1718) by Abraham de Moivre (1667–1754)
- "Liber de ludo aleae" (1564 – published in 1663) by Gerolamo Cardano (1501–1576)

Some Historical Context - Renaissance

- Leonardo da Vinci (1452 – 1519)
- Nicolaus Copernicus (1473 – 1543)
- Niccolò Fontana Tartaglia (1499 – 1557)
- Galileo Galilei (1564 – 1642)
- Johannes Kepler (1571 – 1630)
- Blaise Pascal (1623 – 1662)
- Isaac Newton (1642 – 1726)
- Gottfried Wilhelm von Leibniz (1646 – 1716)

Some Historical Context - Age of Enlightment

- Diderot and D'Alembert "Encyclopdie ou Dictionnaire raisonn des sciences, des arts et des métiers" (1751)

Some Historical Context - Age of Enlightment

- Diderot and D'Alembert "Encyclopdie ou Dictionnaire raisonn des sciences, des arts et des métiers" (1751)
- Age of Enlightenment brought impact on sciences and societal changes
  - Birth of Economics and Chemistry
  - Steam engine by James Watt (Commercialized in 1776)
  - Industrial Revolution
  - French and American Revolutions
  - U.S. Constitution (1789) - influenced by James Maddison, Benjamin Franklin, and Thomas Jefferson. George Washington as President.

Some Historical Context - Age of Enlightment

- A.o.E. in parallel with Baroque and Neoclassicism
- Arts
  - "The History of Art in Antiquity" (1764) by Johann Joachim Winckelmann (1717 – 1768)
  - "Oath of the Horatii" (1784) Jacques-Louis David (1748 – 1825)
  - "Cupid's Kiss" (1787) by Antonio Canova (1757 – 1822)

Some Historical Context - Age of Enlightment

- Age of Enlightenment in parallel with Baroque and Neoclassicism
- Music
  - Johann Sebastian Bach (1685 – 1750)
  - Well Tempered Clavier (Book 1 & 2 in 1722 & 1742)
  - Wolfgang Amadeus Mozart (1756 – 1791)
  - Ludwig van Beethoven (1770 – 1827)

Some Historical Context

- First Statistics Department at University College London (1911)
  - First professor of Statistics Karl Pearson (1857 – 1936) - PCA
  - Following the death of Francis Galton (1822 – 1911) - Eugenics

Some Historical Context

- First Statistics Department at University College London (1911)
  - First professor of Statistics Karl Pearson (1857 – 1936) - PCA
  - Following the death of Francis Galton (1822 – 1911) - Eugenics
- World War I (1914 – 1918)

Some Historical Context

- Other key statisticians
  - Charles Spearman (1863 – 1945) - Rank test
  - Ronald Fisher (1890 – 1962) - Fisher Information, F and Von Mises distributions, LDA
  - Bruno de Finetti (1906 – 1985) - Philosophy of probabilities, exchangeability
  - John Tukey (1915 – 2000) - FFT
  - Calyampudi Radhakrishna Rao (1920 –) - Cramér-Rao bound
  - David Cox (1924 –) - Cox processes, Box-Cox transform

The Birth of Artificial Intelligence

- Alan Turing (1912 – 1954)
  - First abstraction of a machine that can do any computations (1936)
  - The Turing test (1950)
- World War II (1939 – 1945)

The Birth of Artificial Intelligence

- Alan Turing (1912 – 1954)
  - First abstraction of a machine that can do any computations (1936)
  - The Turing test (1950)
- World War II (1939 – 1945)
- Computers become a reality
  - John von Neumann (1903 – 1957) - inspired the design of moder computers

The First Neural Networks

- First neural network model - the Perceptron
  - Frank Rosenblatt (1958)
  - Minsky & Papert (1969)

The First Neural Networks

- First neural network model - the Perceptron
  - Frank Rosenblatt (1958)
  - Minsky & Papert (1969)
- High expectations that this would develop into models of an actual brain

Statistical Learning Theory

- "On the uniform convergence of relative frequencies of events to their probabilities" by Vapnik & Chervonenkis (1968)
  - Risk minimization
  - Regularization
  - VC dimension

Statistical Learning Theory

- "On the uniform convergence of relative frequencies of events to their probabilities" by Vapnik & Chervonenkis (1968)
  - Risk minimization
  - Regularization
  - VC dimension
- "Support Vector Networks" Cortes & Vapnik
- "Support Vector Machines" Cristianini & Shawe-Taylor, Schölkopf & Smola

Statistical Learning Theory

- "On the uniform convergence of relative frequencies of events to their probabilities" by Vapnik & Chervonenkis (1968)
  - Risk minimization
  - Regularization
  - VC dimension
- "Support Vector Networks" Cortes & Vapnik
- "Support Vector Machines" Cristianini & Shawe-Taylor, Schölkopf & Smola
- "Gaussian processes" O'Hagan 1978, Neal 1996, Williams & Rasmussen 1996, Williams & Barber 1998

Machine Learning

- First definition in 1959 by Arthur Samuel (1901 – 1990)
- "Field of study that gives computers the ability to learn without being explicitly programmed"

Machine Learning

- First definition in 1959 by Arthur Samuel (1901 – 1990)
- "Field of study that gives computers the ability to learn without being explicitly programmed"
- Today ML is the field which embraces
  - Statistics
  - Neural Networks
  - Statistical Learning Theory

Machine Learning

- First definition in 1959 by Arthur Samuel (1901 – 1990)
- "Field of study that gives computers the ability to learn without being explicitly programmed"
- Today ML is the field which embraces
  - Statistics
  - Neural Networks
  - Statistical Learning Theory
- **Function estimation**

Bayesian Machine Learning

What will you learn in this course?

- Function estimation using the philosophy of Bayes
- Conditioning on data and modeling assumption
- Offers quantification of uncertainty (due to the lack of data and imprecise knowledge of the environment)

Who uses it?

Companies with lots of data for which traditional models don't exist:

Google, Microsoft, **Amazon**, etc

Who uses it?

Companies with lots of data for which traditional models don't exist:

Google, Microsoft, **Amazon**, etc

- e.g. Recommendations

Who uses it?

Companies with lots of data for which traditional models don't exist:

Google, Microsoft, **Amazon**, etc

- e.g. Recommendations
- Can't write down an equation that describes what I like

Who uses it?

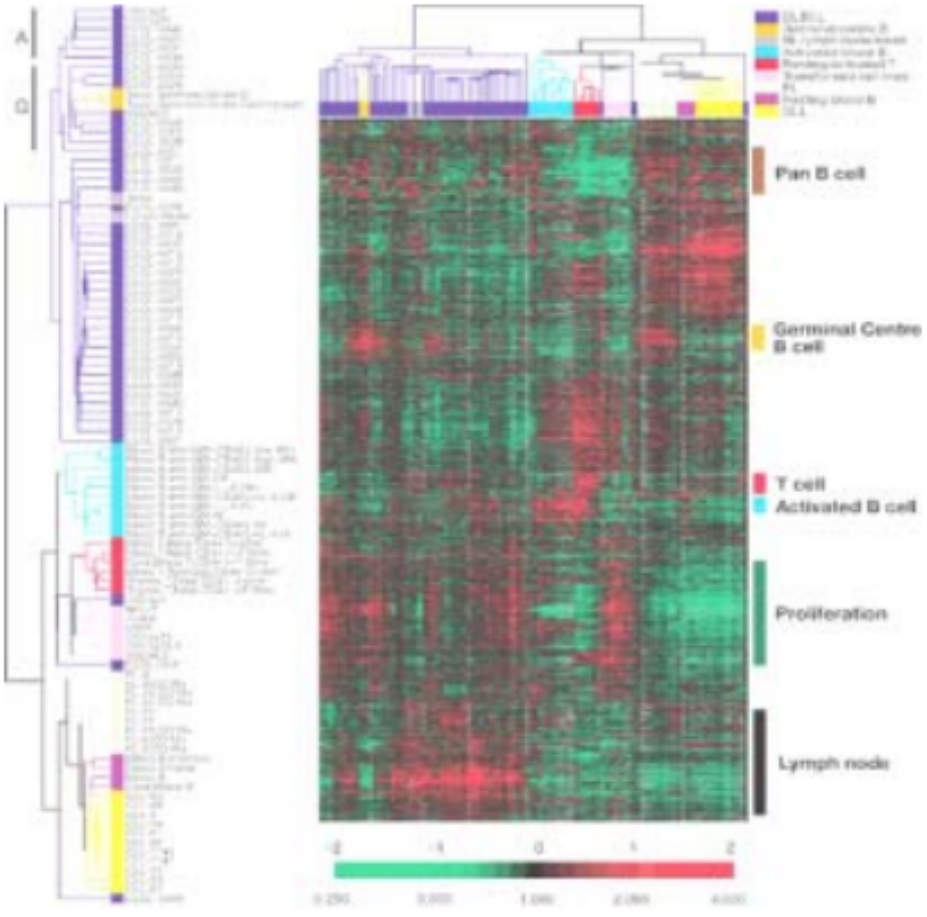
Companies with lots of data for which traditional models don't exist:

Google, Microsoft, **Amazon**, etc

- e.g. Recommendations
- Can't write down an equation that describes what I like
- But we can look for **patterns** in what I buy....
- ...and in what others buy.

Who uses it?

Biotech companies who want to diagnose patients and discover **biomarkers**.



Some examples within EURECOM

Life and Environmental Sciences

- Diagnosis and progression of neurological disorders
- Expensive simulators (climate, tsunami)
- Medical imaging

Some examples within EURECOM

Life and Environmental Sciences

- Diagnosis and progression of neurological disorders
- Expensive simulators (climate, tsunami)
- Medical imaging

Industrial applications

- Fraud detection
- Finance
- Automotive

Course overview

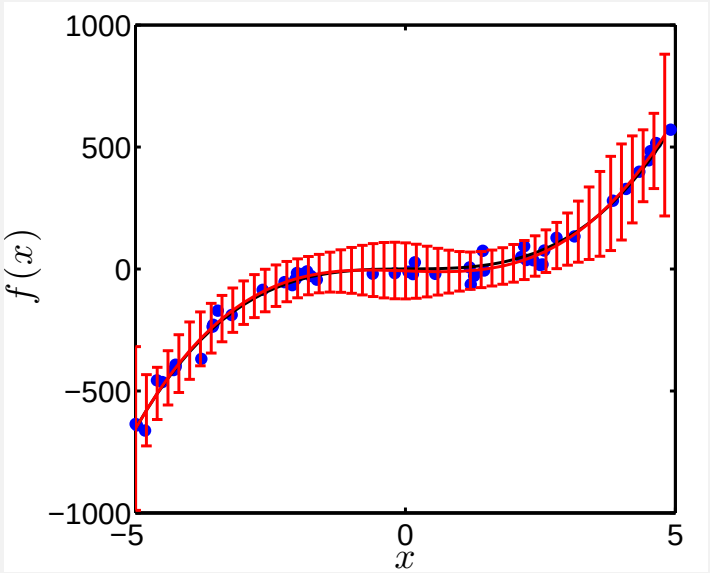
Supervised Learning

Unsupervised Learning

Supervised Learning

Regression

Learning a continuous function from a set of examples.



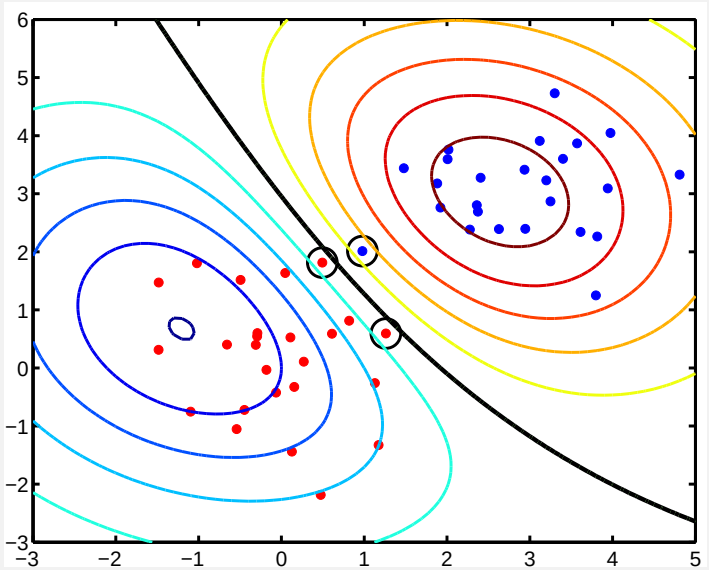
Example

Predicting stock prices (x might be time or some other variable of interest).

Supervised Learning

Classification

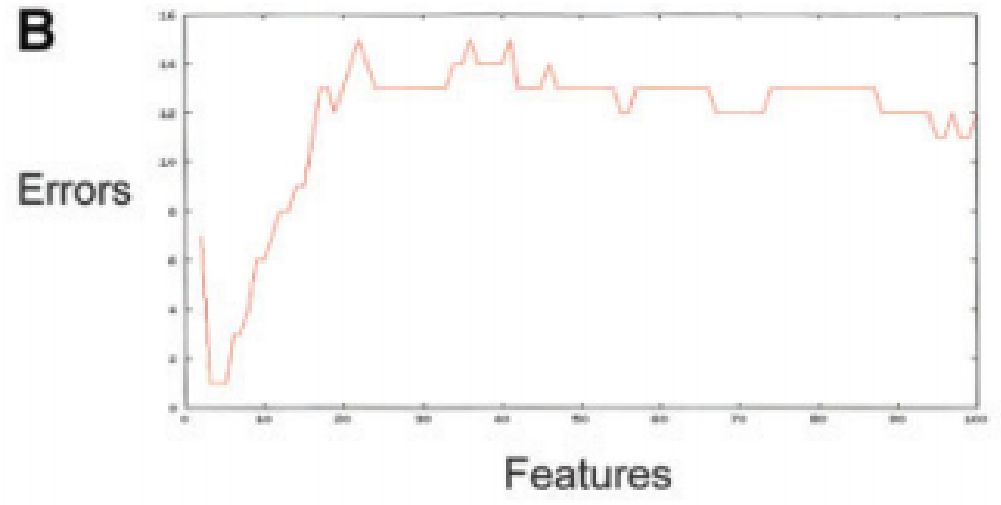
Learning a rule that can separate objects of different types from one another.



Examples

Disease diagnosis, spam email detection.

Predicting relapse of Wilms tumours



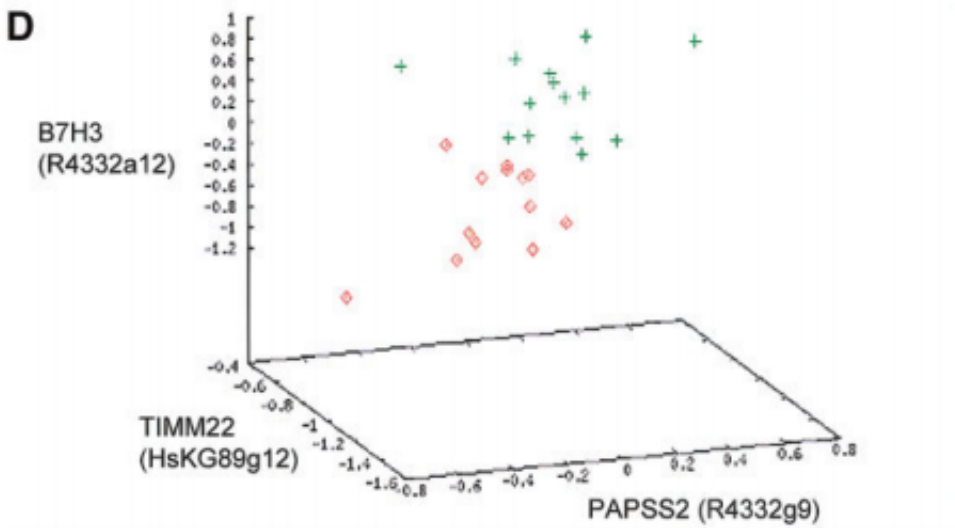
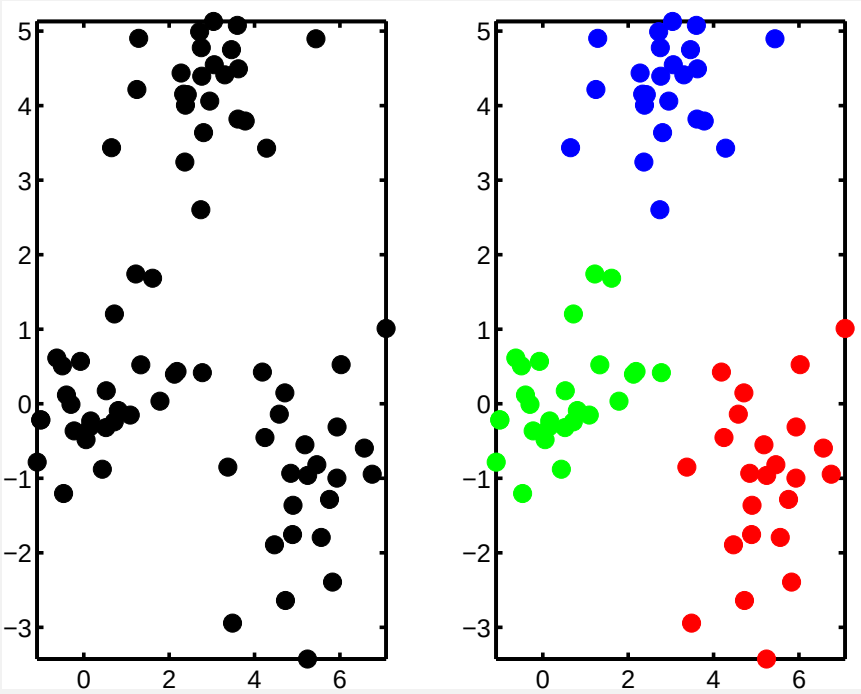


Figure 1. (Continued)

Unsupervised Learning

Clustering

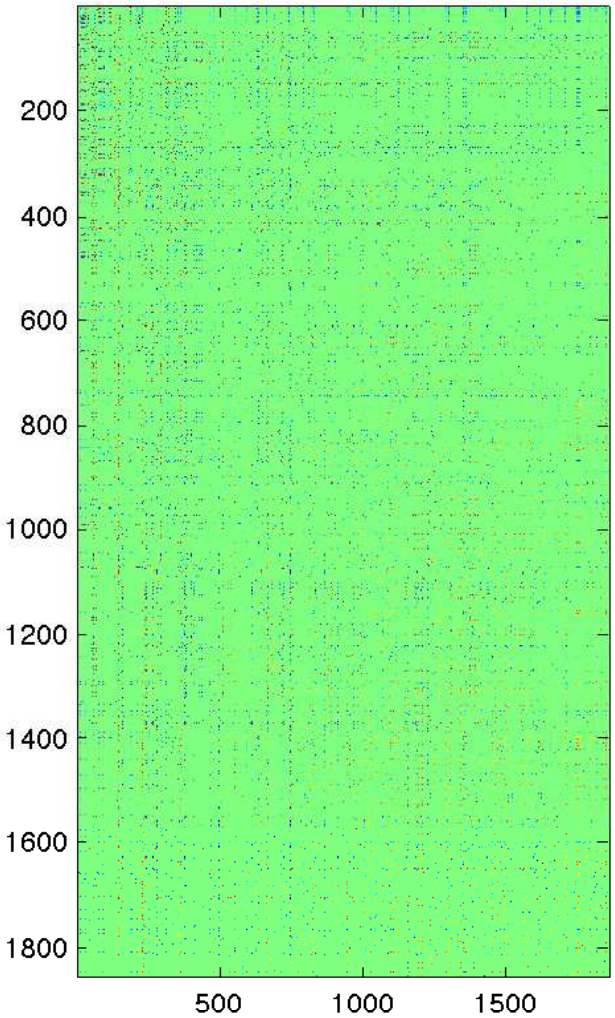
Finding groups of similar objects.

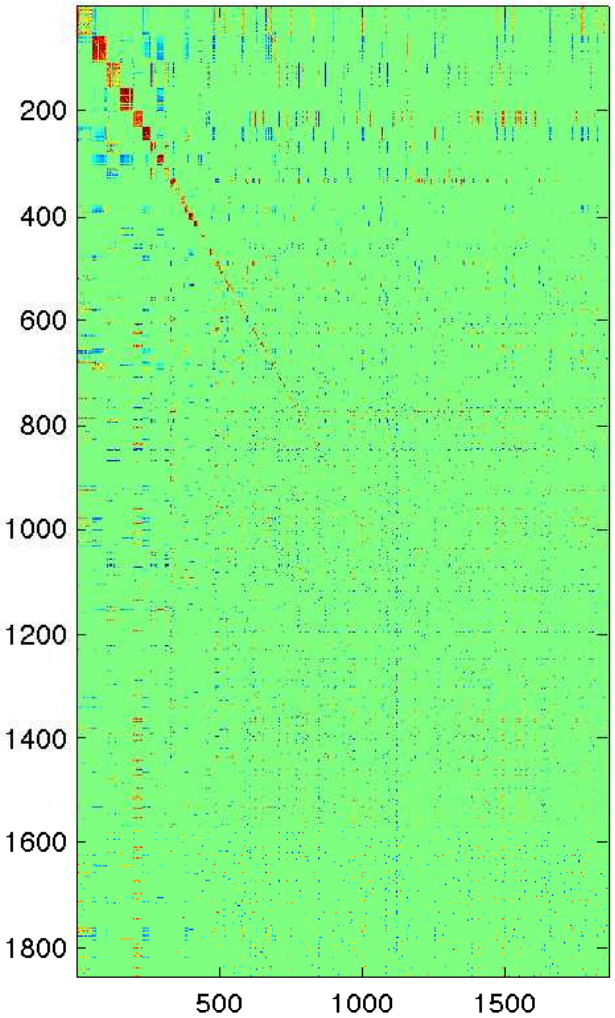


Examples

People with similar 'taste', genes with similar function.

Clustering Example

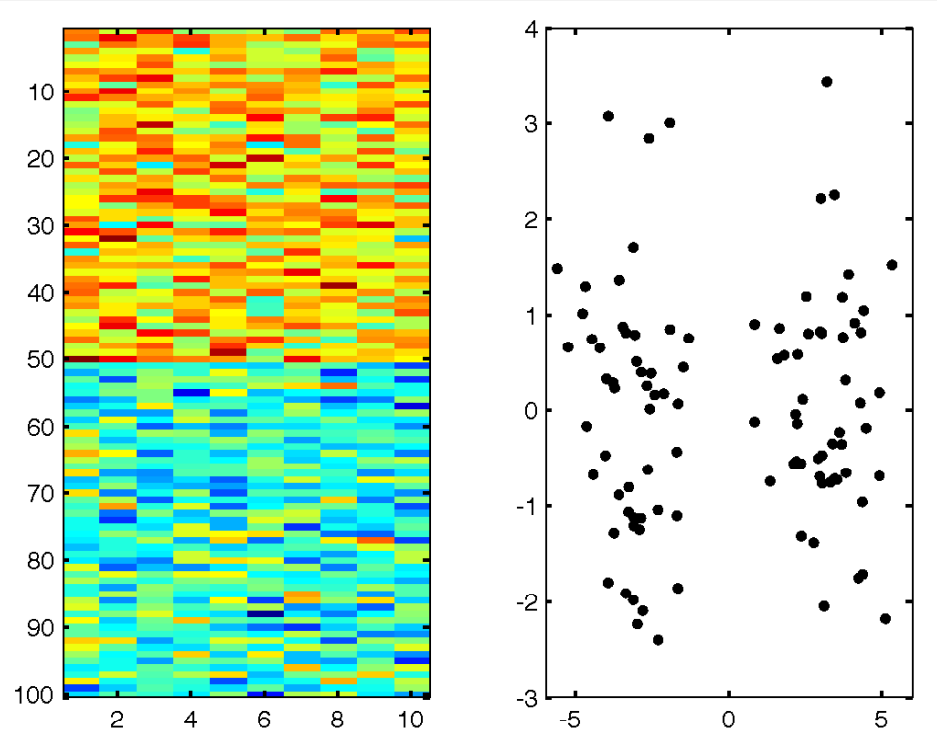




Unsupervised Learning

Projection

Reducing the number of variables – e.g. from 10 to 2.



Examples

Visualizing complex data.

Maths

- We represent objects as vectors/matrices (arrays of numbers), so we have to do maths.
- Being familiar with calculus (function analysis)
- Good understanding probabilities
- Good understanding of linear algebra

ASI schedule - Thursdays 9am-12pm

Lectures

$\times$ 9

ASI schedule - Thursdays 9am-12pm

Lectures

$\times$ 9

Labs

$\times$ 5

Assessment

- Coursework (25%)
  - Dates TBD
- Exam (75%)
- More details of both will be provided in due course.

Contacts

- Collaborative space
- Virtually
  - Maurizio.Filippone@eurecom.fr
- In person
  - Office 419
  - `https://mauriziofilippone.youcanbook.me`

Aside note

- I **do not** write recommendation letters to ASI students

Aside note

- I **do not** write recommendation letters to ASI students
- . . . unless they work on projects under my supervision

Suggested readings

A First Course in Machine Learning

S. Rogers and M. Girolami

Pattern Recognition and Machine Learning

C. Bishop

Information Theory, Inference, and Learning Algorithms

D. MacKay

Machine Learning: A Probabilistic Perspective

K. P. Murphy

Suggested readings

Bayesian Data Analysis

Andrew Gelman

Bayesian Reasoning and Machine Learning

David Barber

Machine Learning

Peter Flach