Machine Learning
and Pattern Recognition

Basics

Fabio G. Cozman

August 5, 2019

Overview

- Two exams (P1 and P2), plus a few assignments (average T).
- Grading: (P1 + 2 P2 + T)/4.
- Textbook: G. James, D. Witten, T. Hastie, R. Tibshirani. *An Introduction to Statistical Learning*, Springer 2013.
- Data: UCI repository, Kaggle.

Other books

- D. Barber, *Bayesian Reasoning and Machine Learning*, Cambridge 2012.
- C. M. Bishop. *Pattern Recognition and Machine Learning*, Springer 2006.
- K. R. Murphy. *Machine Learning: A Probabilistic Perspective*, MIT Press, 2012.

Topics

- Basic technology, kNN.
- Probability, naive Bayes; Bayesian networks.
- MLE; Bayesian estimation; EM; MCMC.
- Statistical tests; bootstrap; comparison between methods.
- Regression (linear, lasso, etc).
- Decision trees, random forests.
- SVMs; boosting.
- Clustering (k-means, dendrograms; latent models).
- PCA, NMF; recommendation systems.
- MDPs and reinforcement learning.
- Neural networks; deep learning; deep reinforcement learning.

What is machine learning?

- T. Mitchell, *Machine Learning*, McGraw-Hill 1997:

  > *The field of machine learning is concerned with the question of how to construct computer programs that automatically improve with experience.*

- Comment: this is very general. A computer might learn by watching a human perform a task once. Or by running some sort of "genetic evolution" simulation. Or a computer might learn by collecting and analyzing data (*statistical machine learning*).

What is statistical machine learning?

- Barber:

  Machine Learning is the study of data-driven methods capable of mimicking, understanding and aiding human and biological information processing tasks.

- James et al.:

  Statistical learning refers to a vast set of tools for understanding data.

- Comment: this definition equates statistical learning and the whole of Statistics. Are they different?

A definition

- Statistical machine learning focuses on how to build computer programs that can improve by collecting and analyzing data.

- Today, there is a sea of data waiting to be analyzed automatically, hence the current success of statistical machine learning.

- How is this effort related to other areas?
  - Artificial intelligence.
  - Statistics.
  - Econometrics, control theory, databases, etc.

A bit of AI

- Definition (??)
  - Systems that think/act like humans (or at least with some degree of intelligence).
  - Systems that think/act *rationally*.
- Some history:
  - Beginning in the 40s/50s: interest in solving operations research problems,
  - Great expectations in the 60s/70s; great difficulties in the 80s; some maturity after 1990.
- Topics: Search and optimization; Knowledge representation (certain, uncertain); Planning; Machine Learning; Vision, sensing, acting, robotics; Communication and language.

Digression on Turing...

- Turing test proposed in 1950: the imitation game. Many arguments; however, not necessary, not sufficient.
- But note Turing's comments on learning:

TURING TEST EXTRA CREDIT: CONVINCE THE EXAMINER THAT HE'S A COMPUTER.

YOU KNOW, YOU MAKE SOME REALLY GOOD POINTS.

I'M ... NOT EVEN SURE WHO I AM ANYMORE.

Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's? If this were then subjected to an appropriate course of education one would obtain the adult brain. (...) The idea of a learning machine may appear paradoxical to some readers.

Machine learning in AI

- At first, probabilities were considered inadequate for AI. Machine learning focused on analogical learning, abduction and other logically motivated techniques.
- Situation changed with the appearance of probabilistic and decision models; the availability of data and the success of neural networks learning from data.
- Machine learning gradually became interwined with statistics.
- Today, statistical machine learning pervades many areas of AI.

Statistical learning and Statistics

- In many ways, statistical learning seems to be just Statistics; however, it is worth noting that the former is interested in computer programs, and the latter is more interested in human consultants.

- Also, statistical learning mostly focuses on building some function that approximates a system, so as to predict or to classify outcomes; there is less focus on modeling and interpreting the model that captures reality — but there is more and more overlap between areas.

- Also, many differences are just in using different names for the same concepts.

Glossary

Machine learning	Statistics
network, graphs	model
weights	parameters
learning	fitting
generalization	test set performance
supervised learning	regression/classification
unsupervised learning	density estimation, clustering
large grant = \$1,000,000	large grant= \$50,000
nice place to have a meeting: Snowbird, Utah, French Alps	nice place to have a meeting: Las Vegas in August

Other connections

- Most of supervised learning is classification, also the object of *pattern recognition*.
- Most of reinforcement learning parallels old efforts in control theory, with new ideas and twists (and focus on computers).

The success of statistical machine learning

- Automated classification and clustering have produced tools that enhance business intelligence, recommendation systems, user support systems, autonomous financial systems.
- Natural language and text processing is heavily based on statistical learning / deep learning (for instance, Google's translator, IBM's Watson).
- Computer vision is dominated by classification and clustering techniques; recently there has been success with deep learning techniques.
- Robotics currently bases navigation and localization on statistical models that require statistical learning.

Read:

- **A. Halevy, P. Norvig, F. Pereira, The unreasonable effectiveness of data, IEEE Intelligent Systems, 24(2):8–12, 2009.**

Big success: translation

- Based on statistical techniques, and now on deep learning.
- Old times (around 1990): English to Russian to English...

> *The spirit is willing, but the flesh is weak.*

Big success: translation

- Based on statistical techniques, and now on deep learning.
- Old times (around 1990): English to Russian to English...

> *The spirit is willing, but the flesh is weak.*
> *The vodka is strong, but the steak is lousy.*

Big success: translation

- Based on statistical techniques, and now on deep learning.
- Old times (around 1990): English to Russian to English...

    The spirit is willing, but the flesh is weak.
    The vodka is strong, but the steak is lousy.

- Current success: data-based.

- The spirit is willing, but the flesh is weak.

- The spirit is willing, but the flesh is weak.
- Ruh istakka tayyor bo'lib, go'shti hafta edi.

But... English - Uzbeque (2018)

- The spirit is willing, but the flesh is weak.
- Ruh istakka tayyor bo'lib, go'shti hafta edi.
- The spirit was ready for the meal, and the flesh was weeks.

- The spirit is willing, but the flesh is weak.
- Ruh istakka tayyor bo'lib, go'shti hafta edi.
- The spirit was ready for the meal, and the flesh was weeks.
- Ruh go'shtga tayyor edi, go'sht esa bir necha hafta edi.

- The spirit is willing, but the flesh is weak.
- Ruh istakka tayyor bo'lib, go'shti hafta edi.
- The spirit was ready for the meal, and the flesh was weeks.
- Ruh go'shtga tayyor edi, go'sht esa bir necha hafta edi.
- The spirit was ready for meat, and the meat was only a few weeks old.

But... English - Uzbeque (2018)

- The spirit is willing, but the flesh is weak.
- Ruh istakka tayyor bo'lib, go'shti hafta edi.
- The spirit was ready for the meal, and the flesh was weeks.
- Ruh go'shtga tayyor edi, go'sht esa bir necha hafta edi.
- The spirit was ready for meat, and the meat was only a few weeks old.
- Ruh tantana qilishga tayyor edi, va bu mushak faqat haftalik keksalikda edi.

But... English - Uzbeque (2018)

- The spirit is willing, but the flesh is weak.
- Ruh istakka tayyor bo'lib, go'shti hafta edi.
- The spirit was ready for the meal, and the flesh was weeks.
- Ruh go'shtga tayyor edi, go'sht esa bir necha hafta edi.
- The spirit was ready for meat, and the meat was only a few weeks old.
- Ruh tantana qilishga tayyor edi, va bu mushak faqat haftalik keksalikda edi.
- The spirit was ready to celebrate, and this muscle was only a weekly old man.