Changelog: 14 Oct 2016, 5 Nov 2016,

What's Machine Learning?

Definitions and Examples

Part 2: Lecture 1

Marina Santini

Acknowledgements
Slides borrowed and adapted from:
Data Mining by I. H. Witten, E. Frank and M. A. Hall (2011)
Slides intergrated with: Daumé (2015); Alpaydin (2014); Mitchell (1997)

Lecture 1: Required Reading

- Handout
- Witten et al. (2011): Ch 1

Outline

- Definitions
- Machine learning & Statistics
- Data vs. Information
- Machine learning techniques
- Datasets & Problem Representation
  - contact lens dataset
- Machine Learning & Data Mining

Start by saying…

Machine learning is the subfield of computer science

Highly interdisciplinary

Machine learning is a wide field, with hundreds of different learning algorithms for solving different problems in virtually any fields.

What is Machine Learning?

1) "The field of study that gives computers the ability to *learn without being explicitly programmed"*

(Arthur Samuel, 1959)

2) "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."

Mitchell (1997: 2)

3) "Machine learning is programming computers to optimize a performance criterion using example data or past experience. "

Alpaydin (2014: 3)

Traditionally…

Solving a problem by creating an algorithm

Ex: sorting numbers

- the input is a set of unordered numbers
- the output is their ordered list

For some problems, the traditional approach is not appropriate…

For ex, consider the "spam" problem:

- input is: an email document that (in the simplest case) is a file of characters.
- output: yes/no label indicating whether the message is spam or not.

Oversimplifying:

Traditional programming:

Data → computer → Output
Program → computer

ML:

Data → computer → Program
Output → computer

The *mission of* ML

To detect certain patterns or regularities in data and construct a good and useful approximation.

That appoximation may not explain everything but may still be able to account for some part of the data.

Why "Learn" ?

There is no need to "learn" to calculate payroll

Learning is used when:

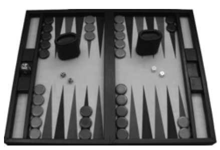
- Mars),
- Humans are unable to explain their expertise (spam detection, speech recognition, etc.)
- Solution changes in time (think in terms of "adaptability", for ex routing on a computer network, or misspellings viagra-v1a-gr-a)
- Solution needs to be adapted to particular cases (user biometrics)

What We Talk About When We Talk About "Learning"

- Learning general models from a data of particular examples
- Data is cheap and abundant (data warehouses, data marts); knowledge is expensive and scarce.
- *Build a model that is a good and useful approximation to the data.*

Complex problems such as…

- Learning to recognize spoken words
- Learning to identify linguistic structures
- Learning to drive autonomous vehicles
- Learning to classify new astronomical structures
- Learning to play games
- etc.



Example by Andrew Y. Ng:

"A computer program is said to *learn* from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E."

Suppose your email program watches which emails you do or do not mark as spam, and based on that learns how to better filter spam. What is the task T in this setting?

- Classifying emails as spam or not spam
- Watching you label emails as spam or not spam.
- The number (or fraction) of emails correctly classified as spam/not spam.
- None of the above—this is not a machine learning problem.

Example by Andrew Y. Ng: Answer

"A computer program is said to *learn* from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E."

Suppose your email program watches which emails you do or do not mark as spam, and based on that learns how to better filter spam. What is the task T in this setting?

- ○ Classifying emails as spam or not spam. T ←
- ○ Watching you label emails as spam or not spam. E ←
- ○ The number (or fraction) of emails correctly classified as spam/not spam.
- ○ None of the above—this is not a machine learning problem. P ←

ML is Multidisciplinary

- Artificial intelligence
  Learning symbolic representations of concepts. Machine learning as a search problem. Learning as an approach to improving problem solving. Using prior knowledge together with training data to guide learning.
- Bayesian methods
  Bayes' theorem as the basis for calculating probabilities of hypotheses. The naive Bayes classifier. Algorithms for estimating values of unobserved variables.
- Computational complexity theory
  Theoretical bounds on the inherent complexity of different learning tasks, measured in terms of the computational effort, number of training examples, number of mistakes, etc. required in order to learn.
- Control theory
  Procedures that learn to control processes in order to optimize predefined objectives and that learn to predict the next state of the process they are controlling.
- Information theory
  Measures of entropy and information content. Minimum description length approaches to learning. Optimal codes and their relationship to optimal training sequences for encoding a hypothesis.
- Philosophy
  Occam's razor, suggesting that the simplest hypothesis is the best. Analysis of the justification for generalizing beyond observed data.
- Psychology and neurobiology
  The power law of practice, which states that over a very broad range of learning problems, people's response time improves with practice according to a power law. Neurobiological studies motivating artificial neural network models of learning.
- Statistics
  Characterization of errors (e.g., bias and variance) that occur when estimating the accuracy of a hypothesis based on a limited sample of data. Confidence intervals, statistical tests.

Machine Learning, Tom Mitchell (1997: 4, Table 1.2)

More specifically, ML & Statistics

- Historical difference (grossly oversimplified):
  - Statistics: testing hypotheses
  - Machine learning: finding the right hypothesis
- But: huge overlap
  - Decision trees (C4.5 and CART)
  - Nearest-neighbour methods
- Today: perspectives have converged
  - Most ML algorithms employ statistical techniques

Data vs. Information

- Society produces huge amounts of data
  - Sources: business, science, medicine, economics, geography, environment, sports, …
- Potentially valuable resource
- Raw data is useless: need techniques to automatically extract information from it
  - Data: recorded facts
  - Information: patterns underlying the data

Basic steps

- Extracting ***implicit, previously unknown, potentially useful*** information from data
- Needed: programs that detect patterns and regularities in the data
- Strong patterns → good predictions
  - Problem 1: most patterns are not interesting
  - Problem 2: patterns may be inexact (or spurious)
  - Problem 3: data may be garbled or missing

Machine learning techniques

- *Algorithms for acquiring structural descriptions from examples*
- Structural descriptions represent patterns explicitly
  - Can be used to predict outcome in new situation
  - Can be used to understand and explain how prediction is derived (*may be even more important*)
- Methods originate from artificial intelligence, statistics, information theory, etc.

Contact lenses: Recommendations

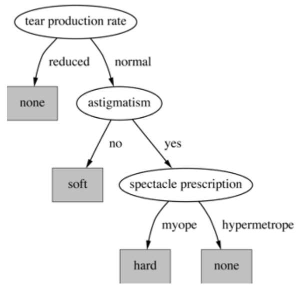
How to capture the essence of the problem?

Age	Spectacle prescription	Astigmatism	Tear production rate	Recommended lenses
Young	Myope	No	Reduced	None
Young	Myope	No	Normal	Soft
Young	Myope	Yes	Reduced	None
Young	Myope	Yes	Normal	Hard
Young	Hypermetrope	No	Reduced	None
Young	Hypermetrope	No	Normal	Soft
Young	Hypermetrope	Yes	Reduced	None
Young	Hypermetrope	Yes	Normal	hard
Pre-presbyopic	Myope	No	Reduced	None
Pre-presbyopic	Myope	No	Normal	Soft
Pre-presbyopic	Myope	Yes	Reduced	None
Pre-presbyopic	Myope	Yes	Normal	Hard
Pre-presbyopic	Hypermetrope	No	Reduced	None
Pre-presbyopic	Hypermetrope	No	Normal	Soft
Pre-presbyopic	Hypermetrope	Yes	Reduced	None
Pre-presbyopic	Hypermetrope	Yes	Normal	None
Presbyopic	Myope	No	Reduced	None
Presbyopic	Myope	No	Normal	None
Presbyopic	Myope	Yes	Reduced	None
Presbyopic	Myope	Yes	Normal	Hard
Presbyopic	Hypermetrope	No	Reduced	None
Presbyopic	Hypermetrope	No	Normal	Soft
Presbyopic	Hypermetrope	Yes	Reduced	None
Presbyopic	Hypermetrope	Yes	Normal	None

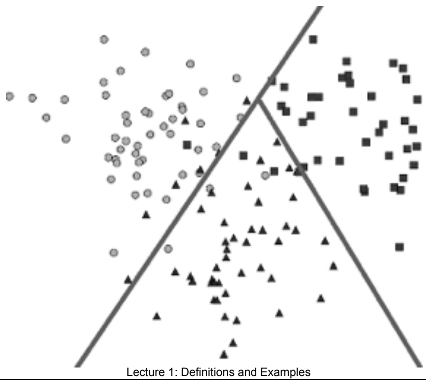
A complete and correct rule set

```
If tear production rate = reduced then recommendation = none
If age = young and astigmatic = no
   and tear production rate = normal then recommendation = soft
If age = pre-presbyopic and astigmatic = no
   and tear production rate = normal then recommendation = soft
If age = presbyopic and spectacle prescription = myope
   and astigmatic = no  then recommendation = none
If spectacle prescription = hypermetrope and astigmatic = no
   and tear production rate = normal then recommendation = soft
If spectacle prescription = myope and astigmatic = yes
   and tear production rate = normal then recommendation = hard
If age young and astigmatic = yes
   and tear production rate = normal then recommendation = hard
If age = pre-presbyopic
   and spectacle prescription = hypermetrope
   and astigmatic = yes then recommendation = none
If age = presbyopic and spectacle prescription = hypermetrope
   and astigmatic = yes then recommendation = none
```


A decision tree for this problem



A linear model for this problem



Definition (ex: contact lenses)

"A computer program is said to learn from experience **E** with respect to some class of tasks **T** and performance measure P, if its performance at tasks in **T**, as measured by P, improves with experience **E**."

Mitchell (1997: 2)

Contatc lenses recommendation problem:

Task **T**: Prescribing the right contact lenses

Training experience **E**: a set of correct diagnoses

Performance measure **P**: the number of correct diagnoses

Definition 3

”Machine learning is programming computers to optimize a performance criterion using example data or past experience. ”

Alpaydin (2014: 3)

Uses

” The model may be ***predictive*** to make predictions in the future, or ***descriptive*** to gain knowledge from data, or both.”

Alpaydin (2014: 3)

Many Types of Learning

- **Supervised Learning**
  - **Classification**
  - Regression
- **Unsupervised Learning**
  - **Clustering**
  - Association rules
- Semi-supervised Learning
- Reinforcement Learning
- etc.

Classification: Applications

Aka Pattern recognition

- Face recognition: Pose, lighting, occlusion (glasses, beard), make-up, hair style
- Character recognition: Different handwriting styles.
- Speech recognition: noisy recording, dialects...
- Medical diagnosis: From symptoms to illnesses
- Biometrics: Recognition/authentication using physical and/or behavioral characteristics: Face, iris, signature, etc
- etc.

Supervised Learning

Prediction of future cases: Use the rule to predict the output for future inputs

Examples:

classification: spam or non spam

regression: the price of a used car

Unsupervised Learning

Clustering: Grouping similar instances

Example applications

Customer segmentation in CRM

Image compression: Color quantization

Bioinformatics: Learning motifs

Reinforcement Learning

Learning a policy: A sequence of outputs
No supervised output but delayed reward
Ex
- Credit assignment problem
- Game playing
- Robot in a maze
- Multiple agents, partial observability, ...

Semi-Supervised Learning

Large amounts of input data and only some of the data is labelled.

It is expensive and time consuming to label data!

Some Canonical Learning Problems

- *Binary classification*
- *Multiclass classification*
- Structured Estimation
- Regression
- Ranking
- etc.

Machine Learning and Data Mining

- Machine Learning focuses on designing algorithms that can learn from and make predictions on the data.
- Data Mining is the application of machine learning methods to large databases.

Machine Learning & Language Technology

- Classical Applications:

① Document Categorization
② Part-Of-Speech Tagging
③ Word Sense Disambiguation
④ Named Entity Recognition
⑤ Syntactic Parsing
⑥ Machine Translation
⑦ etc.

In this course (1)

Theoretical part (online)
- Four supervised classification methods:
  - Decision Trees
  - K-Near Neighbours
  - Naive Bayes
  - Perceptron
- Two unsupervised methods
  - k-Means
  - Hierachical clustering

In this course (2)

Weka workbench (in-class): a general purpose software package of machine learning.

We will try to understand why the different methods issue different results for the same data.

We will learn how to evaluate the different methods on practical problems, such as spam filtering, sentiment classification etc.

Key point

Machine learning is all about finding patterns in data.

ML Methods and their Implementations

Theoretical modelling

Weka implementations

(you can also implement a ML method by yourself, if you wish…)

The end