Introduction to machine learning

V. Kitov

Yandex School of Data Analysis

Imperial College London
Department of Physics

January 2015

Table of Contents

Course information

- 19 January - 28 January
- 10 lectures, 5 seminars
- provide introduction to machine learning

Recommended course materials

- Statistical Pattern Recognition, 3rd Edition, Andrew R. Webb, Keith D. Copsey, John Wiley & Sons Ltd., 2011.
- The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Trevor Hastie, Robert Tibshirani, Jerome Friedman, 2nd Edition, Springer, 2009. http://statweb.stanford.edu/~tibs/ElemStatLearn/
- Pattern Recognition and Machine Learning. Christopher Bishop. Springer. 2007.
- Additional sources - wikipedia, articles, tutorials.

Formal definitions of machine learning

- Machine learning is a field of study that gives computers the ability to learn without being explicitly programmed.
- 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 P at tasks in T improves with experience E.

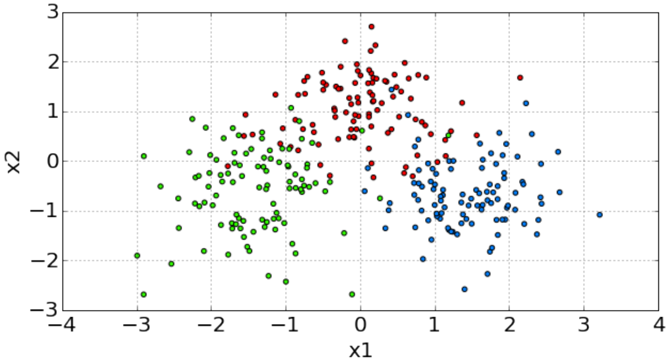
Supervised machine learning

- Find functional relationship between input variables x and output variables y based on expert knowledge and their common observations:

$$(x_1, y_1), (x_2, y_2), ...(x_N, y_N)$$

  - x as a vector is called object, pattern.
  - individual components of x are called features, regressors, inputs.
  - y is called output, target
- if $y \in \mathbb{R}$ => regression
- if $y \in \{\omega_1, \omega_2, ...\omega_C\}$ => classification / pattern recognition

Demonstration



Supervised learning: $x = (x_1, x_2)$, y specified by color

Examples

- Spam filtering
- Document classification
- Web-page ranking
- Sentimental analysis
- Intrusion detection
- Fraud detection
- Target detection / classification
- Handwriting recognition
- Part-of-speech tagging
- Credit scoring
- Particle classification

When it is used?

- making explicit rules is costly / unfeasible
  - dealing with huge datasets with many attributes (text categorization)
  - dealing with data having complex interdependency structure (computer vision)
  - dealing with dynamic environment, requiring quick adaptation (stock prices prediction)

Relationship to other fields

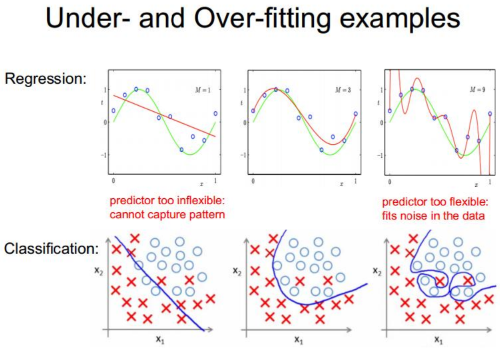
- Related fields:
  - data mining - finding unknown patterns from data with minimum prior knowledge
  - statistics - conclusions about data when statistical model is assumed true
- Machine learning stands in between
- Used fields in ML:
  - probability theory, statistics
  - optimization theory
  - theory of algorithms

Model specification

- **generative**: $p(x, y)$ is modelled
- **discriminative with probabilities**: $p(y|x)$ is modelled
- **discriminative without probabilities**: $y = f(x)$ is modelled

Complexities

- Limited sample
- Model bias (underfitting) versus variance (overfitting) tradeoff:



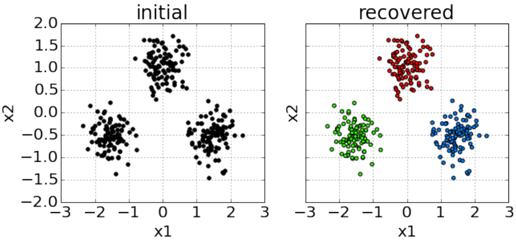
- Possible population drift

Unsupervised learning

- Find functional relationship between input variables x and output variables y based on expert knowledge and only x observations:

$$x_1, x_2, \ldots x_N$$

- Unsupervised learning is also known as clustering (for discrete output)



Unsupervised output recovery

Semi-supervised learning

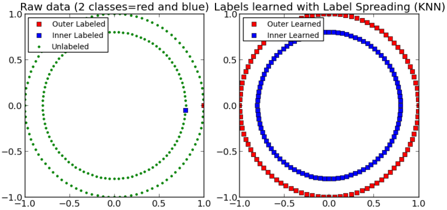
- A small number of joint observations is available:

$$(x_1, y_1), (x_2, y_2), ...(x_N, y_N)$$

- A bigger number of only input observations is also available:

$$x_1, x_2, ...x_M$$

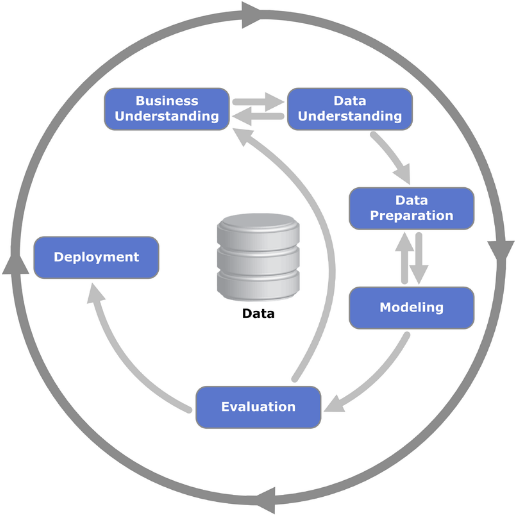
- Recover $x \to y$ relationship



Semi-supervised output recovery

Table of Contents

CrispDM workflow



CrispDM general comments

- Log each step
  - quantitative: procedures and results in report.
  - qualitative: explain why certain option was taken and alternative options ignored.

CrispDM - Business understanding

- Understand business goals and constraints
- State business objective in business terms
- State relevant data mining objective in technical terms
- State success criteria
- Produce plan of project

CrispDM - Data understanding

- Collect data
- Understand data
  - qualitative meaning (what and how was measured)
  - quantitative distribution (data type, range, variance, skewness)
- Explore data
  - basic dependencies
  - interesting subsets
  - statistical analysis
- Quality check
  - outliers
  - missing data
  - errors in measurements

CrispDM - Data preparation

usually takes most of the time

- Select data (select datasets, records, attributes)
- Clean data
  - missing values
  - outliers
  - erroneous values
  - inconsistent groups of attributes
- Construct data
  - derive attributes (normalization, aggregation, composition)
  - use background knowledge
  - fill missing values
- Integrate data together into connected structures (e.g. joined tables)
- Format data (uppercase/lowercase, encoding, etc.)

CrispDM - Modeling

- Select relevant models
  - depending on data mining objective
  - depending on data properties (possibly need to return to data preparation)
- Divide dataset into training/validation/test sets
- Build models
  - choose initial values for model parameters
  - choose parameter estimation techniques
  - estimate parameters
  - post-process results using domain knowledge

CrispDM - Evaluation

- evaluate model output quality using technical data mining criteria
  - compare to baseline
  - reliability of results (statistical significance, dependence on specific data assumptions)
  - check for systematic errors and interpret them (may be caused by missed factors/constraints)
- evaluate resulting models (interpretability, efficiency, scalability)
- analyze final business effect

CrispDM - Deployment

- plan deployment
- plan monitoring and maintaince
- produce final report
- review project experience
  - from project team
  - from customers

Table of Contents

Properties of machine learning algorithms

- Accuracy - achieved on observed data
- Generalization ability (robustness) - expected accuracy decrease on new data
- Online/offline - ability to adapt to new data without full recomputation
- Efficiency
    - computational complexity
        - to train the model
        - to apply the model to new observation
        - to adapt to new information (if online)
    - memory requirements
    - scalability

Properties of machine learning algorithms

- Data constraints
  - assumptions about data distribution
  - flexibility to adapt to data if data assumptions are violated
  - ability to operate well in multidimentional features space
  - possibility to work with discrete/continuous variables or both
- Usability
  - simplicity of maintaince
    - number of user-specified parameters
    - expertise needed to set user-specified parameters
    - simplicity of algorithm logic
  - interpretability of results
- Technical questions:
  - Availability of implicit feature selection
  - Necessity to normalize features before use

Table of Contents

Conclusion

- Course will be mainly devoted to
  - supervised learning
  - classification and regression problems
  - data preparation, modelling and model evaluation stages of CrispDM