CS-567 Machine Learning

Nazar Khan

PUCIT

Lecture 1
Introduction
Oct 3, 2016

Preliminaries

- The ability of biological brains to sense, perceive, analyse and recognise patterns can only be described as stunning.
- They also have the ability to learn from new examples with or without being taught.
- Mankind's understanding of how biological brains operate exactly is embarrassingly limited.
- However, there do exist numerous *practical* techniques that give machines the *appearance* of being intelligent.
- This is the domain of statistical pattern recognition and machine learning.

Preliminaries

- Instead of attempting to mimic the complex workings of a biological brain, this course
  - aims at explaining mathematically well-founded techniques for analysing patterns and learning from them, and is therefore
  - a *mathematically involved* introduction into the field of pattern recognition and machine learning.
- It will prepare you for further study/research in
  - Pattern Recognition
  - Machine Learning
  - Computer Vision
  - Big Data Analytics

  and others areas attempting to solve Artificial Intelligence (AI) type problems.

Prerequisites

- The course is designed to be self-contained.
- *Required mathematical details will be covered in the lectures*.
- However, this is a *math-heavy course*. Students are encouraged[1] to brush up on their knowledge of
  - probability and statistics
  - calculus (differentiation, partial derivatives)
  - linear algebra (vectors, matrices, dot-product, orthogonality, eigenvectors, SVD)
- The only way to benefit from this course is to be prepared to *spend lots of hours reading the text book and attempting its exercises* preferably alone or with a class-fellow.

[1] ordered

Administrative Stuff

Passing this course with at least a B-grade is necessary for students planning to undertake research in the CVML group.

Course web-page:
`http://faculty.pucit.edu.pk/nazarkhan/teaching/Fall2016/CS567/CS567.html`

Text:

- **Required**: *Pattern Recognition and Machine Learning* by Christopher M. Bishop (2006)
  If there is one book you buy, **this** should be it!
- **Recommended**: *Pattern Classification* by Duda, Hart and Stork (2001)

Lectures:
Monday and Wednesday, 8:15 am – 9:40 am in AlKhwarizmi lecture theater.

Administrative Stuff

Grading scheme:

Assignments	Quizzes	Mid-term	Final
20%	5%	35%	40%

- You may have to complete one or two projects in order to obtain complete marks in the mid-term and final exams.
- Theoretical assignments have to be submitted before the lecture on the due date.
- There will be no make-up for any missed quiz.
- Make-up for a mid-term or final exam will be allowed only under exceptional circumstances provided that the instructor has been notified beforehand.
- Instructor reserves the right to deny requests for any make-up quiz or exam.
- Worst score on quizzes will be dropped.
- Worst score on assignments will be dropped.

Programming Environment

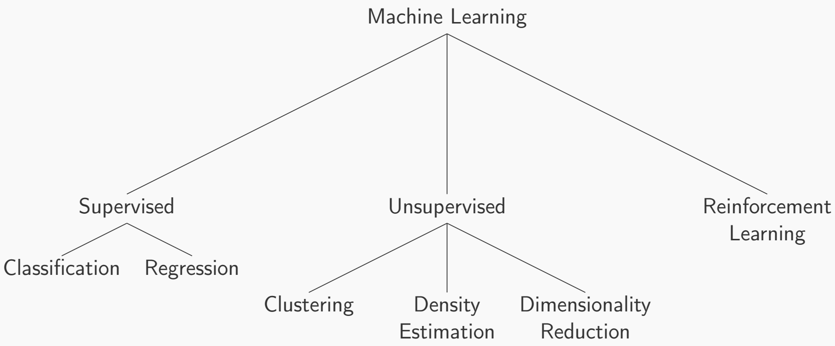
- We will be using `MATLAB` as our programming environment.
- Familiarise yourself with it as soon as possible.
- Resources (by Aykut Erdem):
  - Introduction to MATLAB, by Danilo Scepanovic
  - MATLAB Tutorial, Stefan Roth
  - MATLAB Primer, MathWorks
  - Code Vectorization Guide, MathWorks
  - Writing Fast MATLAB code, Pascal Getreuer
  - MATLAB array manipulation tips and tricks, Peter J. Acklam

Introduction

Machine Learning and Pattern Recognition are different names for essentialy the same thing.

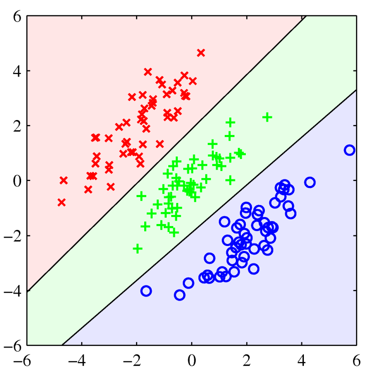
- Pattern Recognition arose out of Engineering.
- Machine Learning arose out of Computer Science.
- Both are concerned with automatic discovery of regularities in data.
- Regularity implies order. Learning implies exploiting order in order to make predictions.

Machine Learning

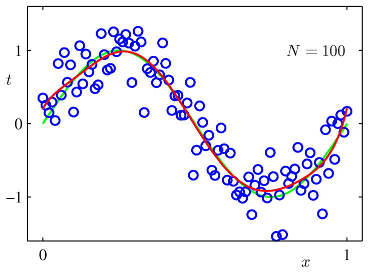


Supervised Learning

- **Classification**: Assign **x** to *discrete* categories.
  - Examples: Digit recognition, face recognition, *etc.*.
- **Regression**: Find *continuous* values for **x**.
  - Examples: Price prediction, profit prediction.



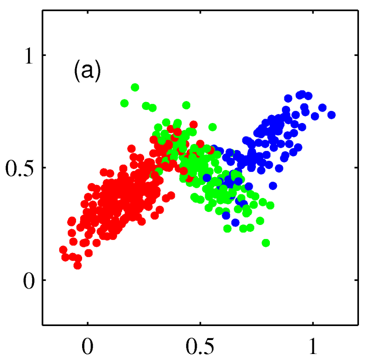
Classification



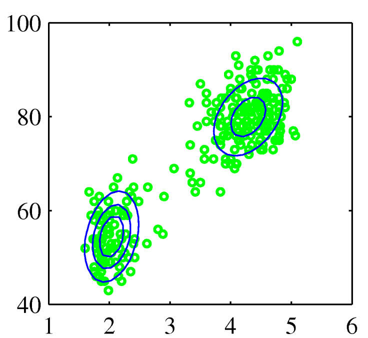
Regression

Unsupervised Learning

- **Clustering**: Discover groups of similar examples.
- **Density Estimation**: Determine probability distribution of data.
- **Dimensionality Reduction**: Map data to a lower dimensional space.



Clustering



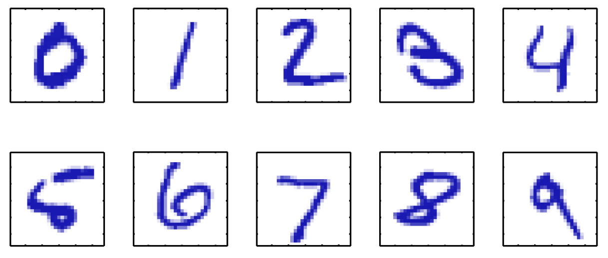
Density Estimation

Reinforcement Learning

- Find actions that maximise a reward. Example: chess playing program competing against a copy of itself.
- Active area of ML research.
- We will not be covering reinforcement learning in this course.

Classical Algorithms vs. Machine Learning

Problem: Given an image $\mathbf{x}$ of a digit, classify it between $0, 1, \ldots, 9$.



Non-trivial due to high variability in hand-writing.

Classical Algorithms vs. Machine Learning

Classical Approach: Make hand-crafted rules or heuristics for distinguishing digits based on shapes of strokes.
Problems:

- Need lots of rules.
- Exceptions to rules and so on.
- Almost always gives poor results.

Classical Algorithms vs. Machine Learning

ML Approach:

- Collect a large *training set* $\mathbf{x}_1, \ldots, \mathbf{x}_N$ of hand-written digits with known labels $t_1, \ldots, t_N$.
- Learn/tune the parameters of an *adaptive* model.
  - The model can adapt so as to reproduce correct labels for all the training set images.

Classical Algorithms vs. Machine Learning

- Every sample $\mathbf{x}$ is mapped to $f(\mathbf{x})$.
- ML determines the mapping f during the *training phase*. Also called the *learning phase*.
- Trained model f is then used to label a new *test image* $\mathbf{x}_{\text{test}}$ as $f(\mathbf{x}_{\text{test}})$.

Terminology

- *Generalization*: ability to correctly label *new* examples.
  - Very important because training data can only cover a tiny fraction of all possible examples in practical applications.
- *Pre-processing*: Transform data into a new space where solving the problem becomes
  - easier, and
  - faster.

  Also called *feature extraction*. The extracted features should
  - be quickly computable, and
  - preserve useful discriminatory information.

Essential Topics for ML

1. Probability theory – deals with uncertainty.
2. Decision theory – uses probabilistic representation of uncertainty to make optimal predictions.
3. Information theory