CS789: Selected topics in Machine Learning and Neural Networks

Introduction

Jakramate Bootkrajang

Department of Computer Science
Chiang Mai University

"... theory is the first term in the Taylor series of practice"
– Thomas M. Cover, "1990 Shannon Lecture"

About the course

- CS789: Machine learning and Neural Networks
- Lecturer: Jakramate Bootkrajang
- Email: jakramate.b@cmu.ac.th
- Office hour: I am almost always at my desk, just walk in.
- Grading: 20% homework, 30% final.
- Programming language: MATLAB, Julia
- Background: Linear algebra, calculus, basic probability.

Course outline

- Maths refresher + introduction to Julia programming
- Linear methods for classification
  - Bayes classifier and Linear discriminant analysis
  - Logistic regression
  - Support Vector Machine
- Non-linear methods for classification
  - Kernel Fisher Discriminant
  - Kernel Logistic Regression
  - Support Vector Machine + Kernel trick
  - Ensemble methods: boosting, bagging

Machine learning

- What is machine learning ?
  - A study (design and analysis) of algorithms which gains expertise in some specific task using past experience (data).
  - Learning is a process of induction: “extrapolating” from examples.

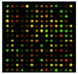
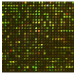
- Learning is useful when:
  - Humans are unable to explain their expertise (speech recognition)
  - Human expertise does not exist (Fraud detection)
  - Solution changes in time (online learning, objective function changes)

Data

Space of inputs: $\mathcal{X}$

- $x \in \mathcal{X} \subset \{0, 1, \ldots, 2^{16}\}^{64}$ a string of pixel intensities in and 8x8 image.
- $x \in \mathcal{X} \subset \mathbb{R}^{33000}$ is a set of gene expression levels.

$x_1 =$  $x_2 =$  $\ldots$

$x_1 =$  $x_2 =$  $\ldots$

$$x_1 = \begin{bmatrix} 1 \\ 2 \\ 16 \\ \vdots \end{bmatrix} \qquad x_2 = \begin{bmatrix} 0 \\ 5 \\ 19 \\ \vdots \end{bmatrix} \qquad \begin{bmatrix} \text{cokes / day} \\ \text{beers / day} \\ \text{BMI} \\ \vdots \end{bmatrix}$$

Data

- Sometimes the space $\mathcal{X}$ is uniquely defined for the problems

- In other cases, such as in vision/text/audio applications, many possibilities exist, and a good feature representation is key to obtaining good performance.

Data

Space of outputs $\mathcal{Y}$.

- $y \in \mathcal{Y} = \{0, 1\}$ is a binary label (1='tiger')
- $y \in \mathcal{Y} = [0, 200]$ is life expectancy

A pair (x, y) is a labelled example.

- (x, y) is an example of an image with a label $y = 1$, which stands for the presence of tiger in image x.

Dataset (or training data): examples $\{(x_1, y_1), \ldots, (x_n, y_n)\}$

Paradigms of machine learning

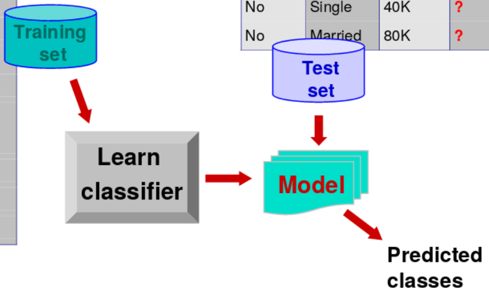
- Supervised learning: learning from labelled data, $\{(x_1, y_1), \ldots, (x_n, y_n)\}$
  - Real valued outputs: regression
  - Discrete valued outputs: classification
- Unsupervised learning: learning from unlabelled data, $\{x_1, \ldots, x_n\}$
- Semi-supervised learning: learning from labelled + unlabelled data.
- Reinforcement learning: learning from interactions with the world based on awards and penalty. Correct input/output pairs are never presented.

Classification illustrated

categorical categorical continuous class

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

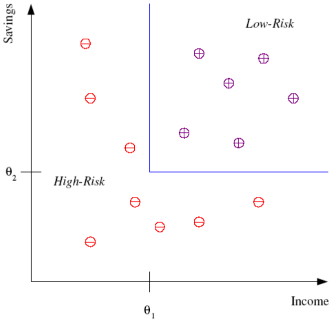
Refund	Marital Status	Taxable Income	Cheat
No	Single	75K	?
Yes	Married	50K	?
No	Married	150K	?
Yes	Divorced	90K	?
No	Single	40K	?
No	Married	80K	?



Courtesy: Jeff Howbert, Introduction to Machine Learning, Winter 2012, UWB.

Classification: applications

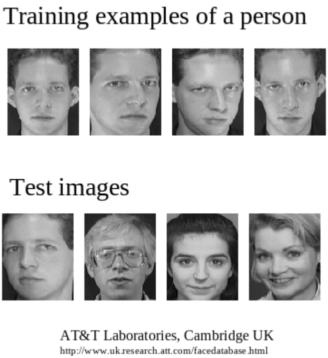
- Credit scoring



- Differentiating beween low-risk and high-risk customers from their *income* and *savings*.
- Discriminant IF *income* $> \theta_1$ AND *savings* $> \theta_2$ THEN low-risk ELSE high risk

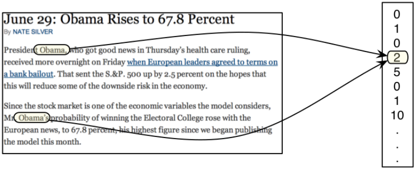
Classification: applications

- Face recognition: Pose, lighting, occlusion (glasses, beard), make-up, hair style



Classification: applications

- Text classification: Spam vs Normal, News topics classification
- This is a bag-of-words representation.

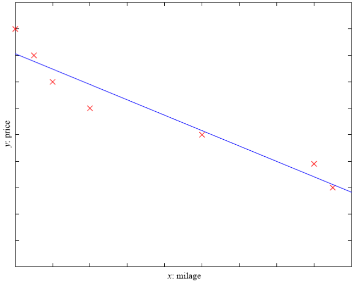


- If "1" stands for the category "politics", then this example can be represented as

$$\left(\boxed{\begin{array}{c} 0 \\ 1 \\ 0 \\ 2 \\ 5 \\ 0 \\ 1 \\ 10 \\ \vdots \end{array}}, 1 \right)$$

Regression

- Price of a used car
- $y = g(x|\theta)$, x: car's attributes, y: price.
- $g()$ is the model and θ is model's parameter.



- For navigating a car: Outputs angle of the steering wheel (CMU NavLab)

Unsupervised learning

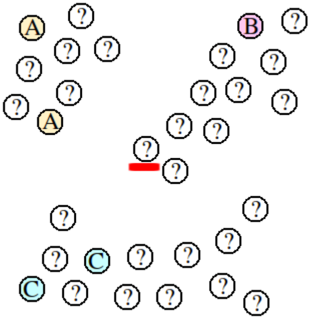
- Learning "what normally happens"
- Clustering: Grouping similar instances
- Example applications
  - Customer segmentation in CRM
  - Image compression: Color quantization

Semi-supervised learning

- Learning from small set of labelled data + large amount of unlabelled data.
- Focus on how to make use of unlabelled data to improve the performance.
- Possible applications are very similar to supervised learning.
- Example: Multiview learning, transductive learning.

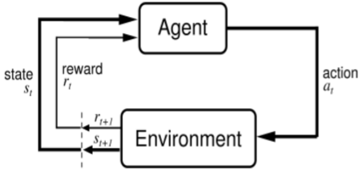
Semi-supervised learning

- Only 5 labelled examples.
- Supervised learning (e.g., k-nn) might misclassify a point in the middle of the figure.



Reinforcement learning

- Learning a policy: A mapping from situation (state) to action.
- Reward action which leads to good state.
- Punish action which leads to bad state.



  - Many control related applications
  - Game playing
  - Robot in a maze

Steps in machine learning

1. Know what you want to do. Understand your data. (Objective)
2. Build a model that is a good and useful approximation to the data. (Modelling)
3. Devise an algorithm to learn the model: how to adjust model's parameters. (Learning)
4. Test your model using existing data or new unseen data. (Performance measure)
5. Theoretically show that your model will work on any new data of the same kind.

Machine learning and related fields

- Machine learning: Focuses on modelling, learning algorithm and performance guarantee.
- Pattern recognition: Sub-field of machine learning focuses on classification tasks.
- Data mining: Focuses on making use of machine learning algorithm.
- Optimisation: Focuses on learning.
- Supporting fields: Mathematics, Statistics.

Challenges

- Structure of data is often unknown or poorly understood.
  - What is the best model for the data ?
- Multimodal data
  - combination of text, images, speech data
- Data abnomalities
  - Class imbalanceness
  - Noisy feature, noisy label or both
  - Missing or incomplete data
- High dimensionality: e.g., gene expression profiles
- Scaling of algorithms to massive data sets: e.g., face detection for every human on earth!
- Streaming data: data distribution keeps changing.

Resources: Books

- Foundation of machine learning, Mehryar Mohri, Afshin Rostamizadeh and Ameet Talwalkar, The MIT Press
- Pattern classification, Richard Duda , Peter Hart, David Storck Wiley-Interscience
- Understanding machine learning, Shai Shalev-Shwartz, Shai Ben-David, Cambridge University Press
- Machine learning, Tom Mitchell, McGraw-Hill Education
- Pattern recognition and machine learning, Christopher Bishop, Springer

Resources: Datasets

- UCI Repository: http://archive.ics.uci.edu/ml/
- UCI KDD Archive: https://kdd.ics.uci.edu/
- Delve: http://www.cs.toronto.edu/ delve/data/datasets.html
- Pascal Large Scale Learning Challenge: http://largescale.ml.tu-berlin.de/about/
- ImageNet: http://image-net.org/index
- MusicNet: http://homes.cs.washington.edu/ thickstn/musicnet.html

Prominant figures

- Michael I. Jordan (UC Berkeley) (Bayesian inference)
- Bernhard Schölkopf (TU Berlin) (Kernel methods)
- Chris Williams (University of Edinburgh) (Gaussian process)
- John Shawe-Taylor (UCL) (Theory)
- Andrew Ng (Stanford)
- Geoffrey Hinton (University of Toronto) (Deep learning)
- Carl Edward Rasmussen, Christopher M. Bishop, Zoubin Ghahramani, Terrence Sejnowski.

Journals

- Journal of Machine Learning Research (JMLR)
- Machine Learning
- Neural Computation
- Neural Networks
- IEEE Transactions on Neural Networks
- IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)
- Pattern Recognition
- Neurocomputing

Conferences

- International Conference on Machine Learning (ICML)
- European Conference on Machine Learning (ECML)
- Neural Information Processing Systems (NIPS)
- Uncertainty in Artificial Intelligence (UAI)
- Computational Learning Theory (COLT)
- International Joint Conference on Artificial Intelligence (IJCAI)