Machine Learning for Data Science (CS4786)
Lecture 1

Tu-Th 10:10 to 11:25 AM
Hollister B14

Instructors: Lillian Lee and Karthik Sridharan
**Rough Details About the Course**

- Diagnostic assignment 0 is out: for our calibration. Hand in your assignments beginning of class on 29th Jan.

- We are thinking roughly three assignments

- (Approximately) 2 competition/challenges,
  - Clustering/data visualization challenge
  - Prediction challenge with focus on feature extraction/selection
Lets get started …
Each time you use your credit card: who purchased what, where and when

Netflix, Hulu, smart TV: what do different groups of people like to watch

Social networks like Facebook, Twitter, . . . : who is friends with who, what do these people post or tweet about

Millions of photos and videos, many tagged

Wikipedia, all the news websites: pretty much most of human knowledge
Guess?
Social Network of Marvel Comic Characters!

by Cesc Rosselló, Ricardo Alberich, and Joe Miro from the University of the Balearic Islands
What can we learn from all this data?
Use data to automatically learn to perform tasks better.

Close in spirit to T. Mitchell’s description
Pedestrian Detection
WHERE IS IT USED?

Market Predictions
WHERE IS IT USED?

Spam Classification

![Email inbox with spam messages highlighted]
More Applications

- Each time you use your search engine
- Autocomplete: Blame machine learning for bad spellings
- Biometrics: reason you shouldn’t smile
- Recommendation systems: what you may like to buy based on what your friends and their friends buy
- Computer vision: self driving cars, automatically tagging photos
- Topic modeling: Automatically categorizing documents/emails by topics or music by genre
- …
TOPICS WE HOPE TO COVER

1. Dimensionality Reduction:

2. Clustering and Mixture models:

3. Probabilistic Modeling & Graphical Models:

4. Some supervised learning: (if time permits)
Topics we hope to cover

1. Dimensionality Reduction:
   Principal Component Analysis (PCA), Canonical Component Analysis (CCA), Random projections, Compressed Sensing (CS), Independent Component Analysis (ICA), Information-Bottleneck, Linear Discriminant Analysis

2. Clustering and Mixture models:

3. Probabilistic Modeling & Graphical Models:

4. Some supervised learning: (if time permits)
Topics we hope to cover

1. Dimensionality Reduction:
   Principal Component Analysis (PCA), Canonical Component Analysis (CCA), Random projections, Compressed Sensing (CS), Independent Component Analysis (ICA), Information-Bottleneck, Linear Discriminant Analysis

2. Clustering and Mixture models:
   k-means clustering, gaussian mixture models, hierarchical clustering, link based clustering

3. Probabilistic Modeling & Graphical Models:

4. Some supervised learning: (if time permits)
Topics we hope to cover

1. Dimensionality Reduction:
   Principal Component Analysis (PCA), Canonical Component Analysis (CCA), Random projections, Compressed Sensing (CS), Independent Component Analysis (ICA), Information-Bottleneck, Linear Discriminant Analysis

2. Clustering and Mixture models:
   k-means clustering, gaussian mixture models, hierarchical clustering, link based clustering

3. Probabilistic Modeling & Graphical Models:
   Probabilistic modeling, MLE Vs MAP Vs Bayesian approaches, inference and learning in graphical models, Latent Dirichlet Allocation (LDA)

4. Some supervised learning: (if time permits)
Topics we hope to cover

1. Dimensionality Reduction:
   Principal Component Analysis (PCA), Canonical Component Analysis (CCA), Random projections, Compressed Sensing (CS), Independent Component Analysis (ICA), Information-Bottleneck, Linear Discriminant Analysis

2. Clustering and Mixture models:
   k-means clustering, gaussian mixture models, hierarchical clustering, link based clustering

3. Probabilistic Modeling & Graphical Models:
   Probabilistic modeling, MLE Vs MAP Vs Bayesian approaches, inference and learning in graphical models, Latent Dirichlet Allocation (LDA)

4. Some supervised learning: (if time permits)
   linear regression, logistic regression, Lasso, ridge regression, neural networks/deep learning, …
Topics we hope to cover

1. **Dimensionality Reduction:**
   - Principal Component Analysis (PCA), Canonical Component Analysis (CCA), Random projections, Compressed Sensing (CS), Independent Component Analysis (ICA), Information-Bottleneck, Linear Discriminant Analysis

2. **Clustering and Mixture models:**
   - k-means clustering, gaussian mixture models, hierarchical clustering, link based clustering

3. **Probabilistic Modeling & Graphical Models:**
   - Probabilistic modeling, MLE Vs MAP Vs Bayesian approaches, inference and learning in graphical models, Latent Dirichlet Allocation (LDA)

4. **Some supervised learning: (if time permits)**
   - linear regression, logistic regression, Lasso, ridge regression, neural networks/deep learning, …
Given (unlabeled) data, find useful information, pattern or structure

- Dimensionality reduction/compression: compress data set by removing redundancy and retaining only useful information

- Clustering: Find meaningful groupings in data

- Topic modeling: discover topics/groups with which we can tag data points
You are provided with $n$ data points each in $\mathbb{R}^d$

Goal: Compress data into $n$, points in $\mathbb{R}^K$ where $K << d$

- Retain as much information about the original data set
- Retain desired properties of the original data set

Eg. PCA, compressed sensing, …
Principal Component Analysis (PCA)

Eigen Face:

- Write down each data point as a linear combination of small number of basis vectors
- Data specific compression scheme
- One of the early successes: in face recognition: classification based on nearest neighbor in the reduced dimension space
Can we compress directly while receiving the input?
We now have cameras that directly sense/record compressed information ... and very fast!
Time spent only for reconstructing the compressed information
Especially useful for capturing high resolution MRI’s
You are at a cocktail party, people are speaking all around you.
But you are still able to follow conversation with your group?
Can a computer do this automatically?
Independent Component Analysis (ICA)

Blind Source Separation

- Can do this as long as the sources are independent
- Represent data points as linear (or non-linear) combination of independent sources
• Help visualize data (in relation to each other)
• Preserve relative distances among data-points (at least close by ones)
K-means clustering

- Given just the data points group them in natural clusters
- Roughly speaking
  - Points within a cluster must be close to each other
  - Points between clusters must be separated
- Helps bin data points, but generally hard to do
Probabilistic generative model for documents

Each document has a fixed distribution over topics, each topic is
has a fixed distribution over words belonging to it

Unlike clustering, groups are non-exclusive
Training data comes as input output pairs $(x, y)$

Based on this data we learn a mapping from input to output space

Goal: Given new input instance $x$, predict outcome $y$ accurately based on given training data

Classification, regression
Feature extraction is a problem/domain specific art, we won’t cover this in class

We won’t cover optimization methods for machine learning

Implementation tricks and details won’t be covered

There are literally thousands of methods, we will only cover a few!
How to think about a learning problem and formulate it

Well known methods and how and why they work

Hopefully we can give you an intuition on choice of methods/approach to try out on a given problem
Given data $x_1, \ldots, x_n \in \mathbb{R}^d$ compress the data points into a low dimensional representation $y_1, \ldots, y_n \in \mathbb{R}^K$ where $K << d$
Why dimensionality reduction?

- For computational ease
  - As input to supervised learning algorithm
  - Before clustering to remove redundant information and noise
- Data visualization
- Data compression
- Noise reduction
Desired properties:

1. Original data can be (approximately) reconstructed
2. Preserve distances between data points
3. “Relevant” information is preserved
4. Redundant information is removed
5. Models our prior knowledge about real world

Based on the choice of desired property and formalism we get different methods.
Sneak Peek

- Linear projections
- Principle component analysis