# Machine Learning

Mahbub Majumdar

BRAC University
66 Mohakhali
Dhaka, Bangladesh

September 16, 2018

Inspiring Excellence

# Table of Contents

# Introduction

### Humans generate data

- Social media
- Online shopping
- Surfing

### Humans consume data

- e.g. deciding what to buy

### Third parties want to analyze the data

- Make predictions
  - e.g. what they can sell

# Introduction

To predict, classify behavior or data, we need *algorithms*

**Algorithm**

Set of instructions to map an input set/label set to an output set

**Sometimes, algorithms don't exist**

- customer behavior
- financial market prediction
- spam detection

Lack of knowledge is compensated by collecting lots of data and rules are inferred from the data.

# Introduction

### Key assumptions about data analysis

- the future $\simeq$ the past
- no new information

### Data Mining

Application of machine learning algorithms to large data sets

### Example

Using large amounts of data to construct simple models e.g. getting binary outputs during stock market prediction

# Applications of Machine Learning

**In credit scoring**

Banks use credit history and lots of other data to predict credit worthiness using simple models

**In fraud detection**

Fraudulent credit card transactions are automatically identified

**In manufacturing**

Optimization, control, troubleshooting

**In medicine**

Medical diagnosis, cancer detection

# Applications of Machine Learning

### In telecommunications

Call patterns are analyzed for network optimization

### In speech and vision recognition

- Facial recognition
  - humans recognize a few structural pieces of information
  - computers record $n \times n$ pixels and extract patterns from them

# Applications of Machine Learning

## In cosmology

- Analyze Cosmic Microwave Background (CMB) radiation to filter different inflationary models
- Analyzing astronomical large mapping data for signatures of possible cosmic strings and cosmic superstrings
- Some typical models of string compactification predict $10^{500}$ possible vacua
  - Machine learning is used to filter the vacua - the ones which are suitable for life

# Machine Learning

### Machine Learning is not just a *database problem*

- part of AI
- intelligent - "ability to adjust to a changing environment" by "learning"

### If a system can adjust to new conditions

- System designer does not need to foresee every possible state/situation
- No need to analyze and program for every possible case
- "label space" can be infinite

# Machine Learning

## General Remarks

- Machine Learning models may be
  - predictive ⇒ forecasting
  - descriptive ⇒ gain knowledge from data
- Machine Learning uses statistics for inference from data
  - distill the data into simple models
- What is the role of Computer Science in Machine Learning?
  - need efficient algorithms to solve optimization problems
  - need to store and process huge amounts of data
  - once a model is learned, need efficient algorithmic implementation (minimize space and time complexity)

# Learning Associations

**In retail: Basket analysis - find association between products**

- If buy $X$, then buy $Y$
- $X =$ iPhone
- $Y =$ lightning cable
- Target for cross-selling?

# Learning Associations

Association Rule: want to find $P(Y \mid X)$

- $Y =$ product that seller wants to sell
- $X =$ product already sold
- Suppose, $X =$ sit down at Sbarro for 1 slice of pizza
- Suppose, $Y =$ pizza to take home to parents
- If $P(Y \mid X) = 0.7$, 70% of non-takeaway customers buy pizza to take home
- If we want to differentiate customers, then we want to find $P(Y \mid X, D)$
- $D =$ feature data e.g. gender, age etc.

# Classification

## Credit Scoring

Banks want to predict risk associated with a future loan

$\Rightarrow$ default probability, P(Default—Credit Data)

Credit Scoring

$\Rightarrow$ assign a score (default probability) given information about past loan payments

$\Rightarrow$ building a credit history is important

$\Rightarrow$ ML fits a model to past data to calculate risk

## Classification

For example, we have to classes:

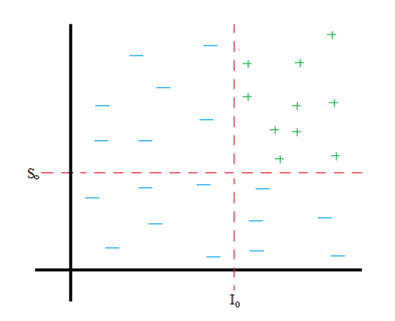
1 $\rightarrow$ low risk

2 $\rightarrow$ high risk

# Classification

After training on data, the classification rule:

IF income $> I_0$ AND savings $> S_0$ THEN low risk, ELSE high risk




**Discriminant Rule**

Function that separates examples into different classes

# Classification

## Pattern Recognition

- Recognizing character codes from images (OCR)
    - multiples classes - letters
    - recognizing handwritten letters
    - compound letters in Bangla
    - Does not have a formal description of "X" that covers all possible X
        - extract common characteristics
    - In text, redundancy exists
        - word is a sequence of characters
        - succesive characters are dependent and constrained
        - enables error correction e.g. "t?e" $\rightarrow$ "the"
    - Learning models exist that learn sequences and model these dependencies
- Facial recognition
    - more difficult than OCR
        - three dimensional, lighting, different poses
        - veiling of certain inputs e.g. glasses, makeup, facial hair

# Classification

## Pattern Recognition

- Medical diagnosis
  - need lots of test data to prevent ML algorithm from giving false positives
- Natural Language Processing (NLP)
  - acoustic data
  - temporal order, sequence of speech phonemes
  - different accents, gender, age

    pronounce same words differently
  - "best" language model
    - learning from large corpus of data
- Biometrics
  - integration of inputs from different modalities
  - For example,

    Physiological Characteristics → facial images, fingerprints, iris and palm scans

    Behavioral Characteristics → gait, keystroke, signature, voice
  - ML used in the separate recognizers and in the combinations of their

# Classification

## Knowledge Extraction

The discriminant, for example, teaches how to discriminate between high risk and low risk borrowers

## Compression

- Learning causes compression
  - e.g. learn $y = x^2, x \geq 0$
- requires less item to store and sometimes requires less computations to process

## Outlier detection/Novelty detection

Finding cases not following general rule e.g. fraud, cosmic strings

# Regression

## Pricing

- Let,

  y = price of a used car

  x = attributes of the car e.g. mileage

  then, $y = ax + b$
- Why linear?

  $\Rightarrow \Delta y \propto \Delta t$
- Log linear

  $\Rightarrow \Delta y \propto y \Delta x$

# Supervised Learning

- Both regression and classification are supervised learning
  - Task is to learn mapping from input to well defined output

$$y = g(x|\theta)$$

y = model

g = output of model/regression/classification

$\theta$ = parameters that need to be optimized such that estimates are as close as possible to the correct values

# Supervised Learning

## Regression example

Navigation of a mobile robot, autonomous driving

- $\Rightarrow$ output $\rightarrow$ angle to turn steering wheel
- $\Rightarrow$ input $\rightarrow$ GPS, video cameras

## Other examples

Expected goals, xg = $g(x_i|fitness_i)$ where i = 1,2,.....n

## Ranking

Output is a relative position/label instead of real numbers

- $\Rightarrow$ recommender system for movies, classes to take, teachers to follow

# Unsupervised Learning

- No supervision to tell the correct value of output
- There is only input data
- Goal
  - find regularities in the input data
  - look for pattern in the domain space

## Example of Density Estimation

- Clustering, finding groups of input
- Customer relationship management
  - Customer segmentation
    - Company wants to fit a profile to customers
    - Collects customer data (demographics, past transactions)
  - Once segmentation is found, company will use different strategies to target different groups
  - Allows identification of outliers
    - niche markets
    - unreliable customers

# Unsupervised Learning

## Application of Clustering

- Image compression
  - shades of a small number of colors
  - Pixels are 24 bits to represent 16 million colors
  - Pixels of only 64 shades of colors require 6 bits
  - e.g. both light blue and dark blue is considered blue
  - requires less space
- Ideally, would like to identify higher level regularities
  - texture
  - repeated image patterns
- Scanned images
  - bitmapped data
    $\Rightarrow$ 16 x 16 bitmap of "X" = 32 bytes
    $\Rightarrow$ map this to ASCII code $\rightarrow$ 1 byte

### Bioinformatics

- DNA - sequence of bases A, G, C, T
- RNA transcribed from DNA
- Proteins translated from RNA
- Protein is a sequence of amino acids
- Alignment
  - matching DNA/protein sequences
  - difficult because sequence may be very long, there are many template strings to match against, there maybe deletions, insertions and substitutions
  - Use clustering in learning motifs (sequence of amino acids)
    - motifs = like words
    - amino acids = words
    - proteins = sentences
  - read a string of protein/DNA/RNA

# Reinforcement Learning

- For some systems
  - output $\rightarrow$ sequence of actions
  - single action is not important
  - "policy" (sequence of correct actions) is important
  - no best action
  - action is good if it is part of a good policy
  - Machine learning

    $\Rightarrow$ assesses the goodness of policies

    $\Rightarrow$ learns from past good action sequences to generate other good policies
  - Examples:
    - Game playing (Chess, Checker, GO) $\rightarrow$ sequence of steps/moves
    - Robot walking

      $\rightarrow$ how to walk, run

      $\rightarrow$ robot can move in many directions

      $\rightarrow$ should learn correct sequence of actions to reach goal state

# Reinforcement Learning

## Reinforcement Learning Complex

- unreliable, incomplete sensory information
- state is partially observable
- e.g. robot may not know exact position in room
- some goals require multiple agents interacting in a sequence of actions e.g. robots playing football