INTRODUCTION TO MACHINE LEARNING

Introductions

Dr. | Prof | Professor
Dave | Kauchak

Pronouns: he/him/his

Why are you here?

What is Machine Learning?

Why are you taking this course?

What topics would you like to see covered?
Machine Learning is...

Machine learning is a subfield of computer science that evolved from the study of pattern recognition and computational learning theory in artificial intelligence.

--- Ethem Alpaydın

The goal of machine learning is to develop methods that can automatically detect patterns in data, and then to use the uncovered patterns to predict future data or other outcomes of interest.

--- Kevin P. Murphy

The field of pattern recognition is concerned with the automatic discovery of regularities in data through the use of computer algorithms and with the use of these regularities to take actions.

--- Christopher M. Bishop

Machine Learning is...

Machine learning is about predicting the future based on the past.

--- Hal Daume III
Machine Learning, aka

- data mining: data analysis, not prediction, though often involves some shared techniques
- inference and/or estimation in statistics
- pattern recognition in engineering
- signal processing in electrical engineering
- induction
- optimization

Goals of the course: learn about...

- Different machine learning problems
- Common techniques/tools used
  - theoretical understanding
  - practical implementation
- Proper experimentation and evaluation
- Dealing with large (huge) data sets
  - Parallelization frameworks
  - Programming tools

Goals of the course

- 90s: neural networks
- early 2000s: support vector machines
- after that: probabilistic models (aka graphical models)
- currently: neural networks, deep learning

Be able to laugh at these signs (or at least know why one might...)

Why wait till it’s too late!
Course expectations

Plan to stay busy!

Applied class, so lots of programming

Machine learning involves math

Where we've been!

Our ML suite:

- 29 classes
- 2951 lines of code

13

Where we've been!

Our ML suite:

- Supports 7 classifiers
  - Decision Tree
  - Perceptron
  - Average Perceptron
  - Gradient descent
  - 2 layer neural network
  - K-NN
  - Naive Bayes
- Supports two types of data normalization
  - Feature normalization
  - Example normalization
- Supports two types of meta-classifiers
  - OVA
  - AVA

Administrative

Course page: [http://www.cs.pomona.edu/classes/cs158/](http://www.cs.pomona.edu/classes/cs158/)

Assignments
- Weekly
- Mostly programming (Java, mostly)
- Some written/write-up
- Generally due Sunday evenings

Two "midterm" exams and one final (all time limited take home)

Late Policy

Collaboration
Other things to note

- Videos before class
- Lots of class participation!
- Read the book (it's good)

Machine learning problems

What high-level machine learning problems have you seen or heard of before?
Supervised learning: given labeled examples.
Supervised learning

Supervised learning: learn to predict new example

Supervised learning: classification

Classification: a finite set of labels

Classification Example

Classification Applications

Face recognition
Character recognition
Spam detection
Medical diagnosis: From symptoms to illnesses
Biometrics: Recognition/authentication using physical and/or behavioral characteristics: Face, iris, signature, etc.

...
Supervised learning: regression

Supervised learning: given labeled examples

Regression Example

Price of a used car

$x$: car attributes (e.g., mileage)

$y$: price

Regression Applications

Economics/Finance: predict the value of a stock

Epidemiology

Car/plane navigation: angle of the steering wheel, acceleration, ...

Temporal trends: weather over time

...
Ranking example

Given a query and a set of web pages, rank them according to relevance

Ranking Applications

- User preference, e.g. movie ranking
- iTunes
- Flight search (search in general)
- Reranking N-best output lists
  ...

Unsupervised learning

Unsupervised learning: given data, i.e. examples, but no labels

Unsupervised learning applications

- Learn clusters/groups without any label
- Customer segmentation (i.e. grouping)
- Image compression
- Bioinformatics: learn motifs
  ...

...
Reinforcement learning

- Left, right, straight, left, left, straight, left, right, straight, straight, straight
- Left, left, left, straight, left, right, straight, straight, straight

Given a sequence of examples/states and a reward after completing that sequence, learn to predict the action to take for an individual example/state.

Reinforcement learning example

Backgammon

- Left, right, straight, left, left, straight, left, right, straight, straight, straight
- Left, left, left, straight, left, right, straight, straight, straight

Given sequences of moves and whether or not the player won at the end, learn to make good moves.

Other learning variations

- What data is available:
  - Supervised, unsupervised, reinforcement learning
  - Semi-supervised, active learning, ...

- How are we getting the data:
  - Online vs. offline learning

- Type of model:
  - Generative vs. discriminative
  - Parametric vs. non-parametric
Representing examples

What is an example? How is it represented?

Features

Features are the questions we can ask about the examples

How our algorithms actually “view” the data

Classification revisited

During learning/training/induction, learn a model of what distinguishes apples and bananas based on the features

Examples

- red, round, leaf, 3oz, ...
- green, round, no leaf, 4oz, ...
- yellow, curved, no leaf, 8oz, ...
- green, curved, no leaf, 7oz, ...

Features

- $f_1$, $f_2$, $f_3$, ..., $f_n$

Label

- apple
- banana
Classification revisited

The model can then classify a new example based on the features.

Apple or banana?

Why?

Training data
- red, round, no leaf, 4oz, ...
- green, round, no leaf, 4oz, ...
- yellow, curved, no leaf, 4oz, ...
- green, curved, no leaf, 5oz, ...

Label
- apple
- banana

Examples
- red, round, leaf, 3oz, ...
- green, round, no leaf, 4oz, ...
- yellow, curved, no leaf, 4oz, ...
- green, curved, no leaf, 5oz, ...

Test set
- red, round, no leaf, 4oz, ...
- green, curved, no leaf, 4oz, ...
- yellow, curved, no leaf, 4oz, ...
- green, curved, no leaf, 5oz, ...

Learning is about generalizing from the training data.

What does this assume about the training and test set?
Past predicts future

Training data

Test set

Not always the case, but we’ll often assume it is!

More technically...

We are going to use the probabilistic model of learning.

There is some probability distribution over example/label pairs called the data generating distribution.

Both the training data and the test set are generated based on this distribution.

What is a probability distribution?
Probability distribution

Describes how likely (i.e. probable) certain events are
- Describes probabilities for all possible events
- Probabilities are between 0 and 1 (inclusive)
- Sum of probabilities over all events is 1

Probability distribution

- Training data
  - High probability
    - round apples
    - curved bananas
    - apples with leaves
  - Low probability
    - curved apples
    - red bananas
    - yellow apples

data generating distribution

- Training data
- Test set

- data generating distribution

- Training data
- Test set

- data generating distribution
data generating distribution

Training data  Test set