CS305 Machine Learning

Course Information

Instructor: Brian Tjaden
Pronouns: He, Him, His
Course Materials: http://cs.wellesley.edu/~cs305

Programming Environment

- We will both implement our own machine learning algorithms and use existing machine learning algorithm implementations
- There are many software libraries for studying and programming machine learning applications
- In this course, we will use Python and its libraries numpy, matplotlib, and sklearn
- For development, we will use Anaconda together with Jupyter notebooks

Aspects of the Course

Different machine learning (ML) courses emphasize different things

- Analyzing interesting data
- Using existing ML technologies
- Societal impact of ML
- Implementing ML algorithms
- Real world applications of ML
- Ethical considerations
- Theoretical foundations of ML
Overview

Supervised Learning
- Regression Models
- Linear Classifiers
- Non-Linear Classifiers
- Decision Trees
- kNN
- Support Vector Machines
- Collaborative Filtering

Parametric
- Linear Regression
- Perceptron
- Neural Networks
- Hidden Markov Models

Unsupervised Learning
- Hierarchical Clustering
- K-Means
- Gaussian Mixture Models
- Dimensionality Reduction

What is Learning?

How old do you need to be in order to vote?

There is a line that passes through (1,4) and (5,1). Does (3,3) also lie along this line?

In machine learning, computers learn with experience. They solve a problem based on data or they get better at a task as they have more examples.

Observe these photos of two people

Quick!

Is this Z or TS?
How would you write a program to recognize a particular person in a picture? Could this program be adapted easily to recognize some other person? Automatically learn from examples

Program should observe patterns in pictures of faces to recognize new ones

Is this how humans perform tasks?

- Do we pattern match from examples?

- Or do we have a deeper understanding / intuition for how to recognize faces, drive safely, and play chess?

Flavors of Machine Learning

1. **Supervised learning:**
   
given training examples with labels, learn to automatically label (aka classify) a new example

   - Z
   - TS
   - Z
   - TS
Flavors of Machine Learning

2. **Unsupervised learning:**
   given *data without labels*, extract hidden structure

3. **Reinforcement learning:**
   learn how to take actions to maximize total reward

---

Supervised Learning Phases: (1) Training

- **Training Data**
- **Classifier**
- **Testing Data**
- **Predicted Label**

**Are these equal?**
Training and Testing ML Algorithms

Data Representation with Features

- A **feature** is a question you ask of every example in your data (in training and testing)
- A **feature value** is the answer to that question for a particular example in the data
- Typically have LOTS of features
- Machine learning algorithms specify how to use these features to build classifiers

Common Danger: Overfitting

- TS’s shirt is black in all training examples
- Classifier may learn that “black shirt” is a good feature for recognizing her
- What happens if she is not wearing black clothing, or if someone else is wearing black clothing?

This Course

- Algorithms for learning supervised classifiers
- Some unsupervised learning
- Social and ethical issues
- Designing good features for different domains
- How to evaluate your classifier
- How to overcome overfitting
- Processing large amounts of data