Any programming language is allowed in this course.

* NumPy
* Java
* Matlab
* C Variants

50% homework grade (6 assignments)

1% off per hour

Midterm exam

We have a task (T), we want to define a performance (P) metric based on experience (E).

Machine learning:

* pattern recognition
* making decisions based off of prior intelligence

Machine learning falls into the group of AI

How we do it:

* What is the task/goal?
* How do we get experience?
  + Supervised learning (correct examples)
  + Unsupervised learning (we don't know if the answer is correct)
* What representation should be used for the experience of the task.

Unsupervised learning:

* Clustering:
* Find patterns through data.

Supervised learning:

* Regression
  + Predict outcome value
* Classification
  + Given data, can you find what category it belongs to.
  + Hotdog or Not hotdog

Course objectives:

* Regression
* Probabilities
* Classification
* Mixture models
* Sequential data

**Types of Algorithms**

* Instance based algorithms
  + In order for the system to do its job it needs all the historic data.
* Model based algorithms
  + The learning process/training process learns parameters that describe the models
* Parametric algorithms
  + Learn parameters that describe the model
* Non parametric algorithms
  + Not defined by parameters
* Linear
* Non-Linear

**Types of Data**

* Categorial/Enumerated
* Finite Discrete Valued
* Continuous Valued

There is no algorithm that works best for everything.

**Dealing with Data**

* Parsing text files
* Save files as a native object

**Feature Types**

* We can break features into three types:
  + Real valued
  + Enumerated
  + Categorical
* Depending on the algorithm we may need to:
  + Standardize
  + Convert enumerated/categorical into a set of binary features

**Preparing Data:**

1. Categorical -> Binary
2. Standardize Real-Valued features

**Standardizing Data**

Imagine if we have two sets of data about a person, their weight and their height. We need to make sure one of them doesn't affect the outcome more.

How we standardize:

* Zero mean
* Unit deviation

We treat each feature independently and:

* Center it (subtract the mean from all samples)
* Make them all have the same span (divide by standard deviation)

Standardizing *might* be better if we scale the values to the point to use integer variables.

**Data Dimensionality**

* Our data has a ton of information in it.
* A textbook has thousands of words.
* Computation cost?
* Statistical cost?
* Visualization

Having too much information is not great.

Dimensionality reduction:

Goal: Represent instances with fewer variables

* Try to preserve as much structure in the data as possible.
* If there is class information
  + Increase class separability

Benefits:

* Training is faster.
* Less data to cover the feature space.
* Easier visualization.

Approaches:

* Feature selection
  + Select a subset of the original features.
    - Using greedy heuristic to select the best feature.
      * This is not guaranteed to provide an optimal feature selection.
    - Greedy is slow, we can decrease the search space with information gain
      * Select a feature set based on the information it provides.