A Tiny Taste of Machine Learning

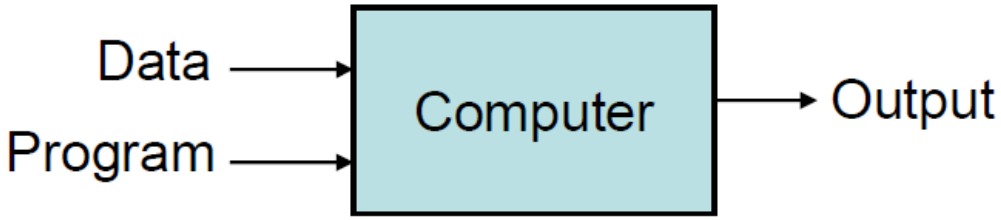
What Is Machine Learning

- Many useful programs learn something
- In the last two 6.00.2x lectures we used linear regression to learn models of data
- *"Field of study that gives computers the ability to learn without being explicitly programmed."* Arthur Samuel

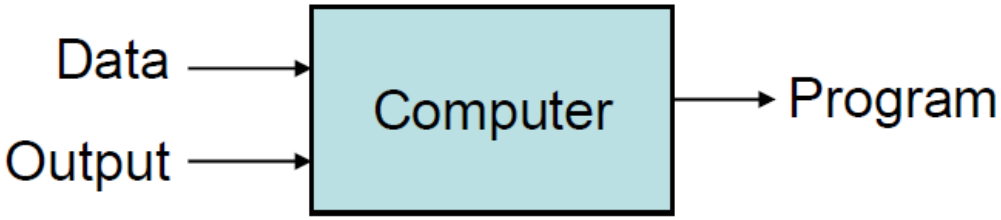
What Is Machine Learning?

- Modern statistics meets optimization

Traditional Programming



Machine Learning



How Are Things Learned?

- Memorization
  - Accumulation of individual facts
  - Limited by
    - Time to observe facts
    - Memory to store facts
- Generalization
  - Deduce new facts from old facts
  - Limited by accuracy of deduction process
    - Essentially a predictive activity
    - Assumes that the past predicts the future

Basic Paradigm

- Observe set of examples: training data
- Infer something about process that generated that data
- Use inference to make predictions about previously unseen data: test data

All ML Methods Require

- Representation of the features
- Distance metric for feature vectors
- Objective function and constraints
- Optimization method for learning the model
- Evaluation method

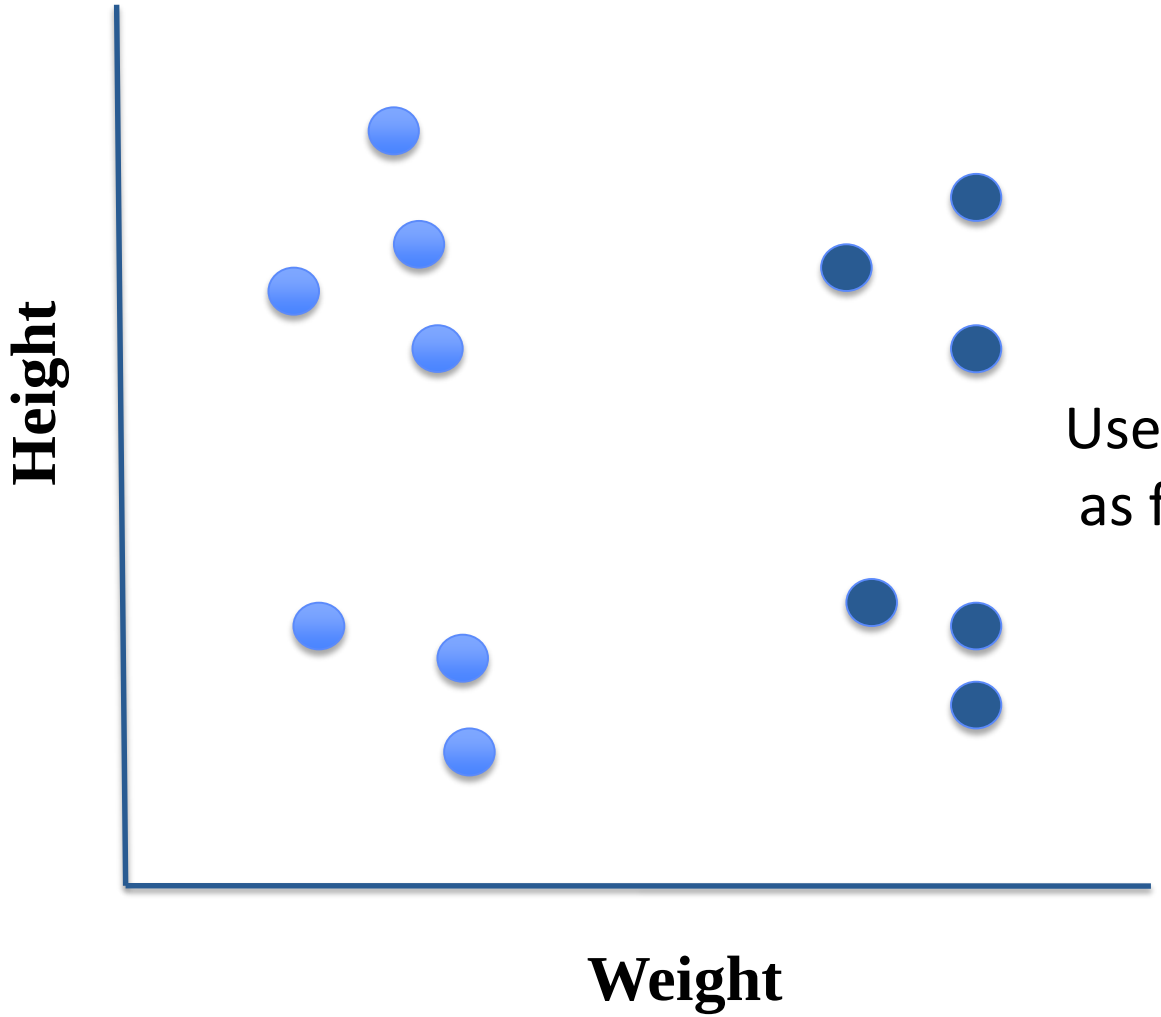
Supervised Learning

- Start with set of feature vector/value pairs
- Goal: find a model that predicts a value for a previously unseen feature vector
- Regression models predict a real
  - As with linear regression
- Classification models predict a label (chosen from a finite set of labels)

Unsupervised Learning

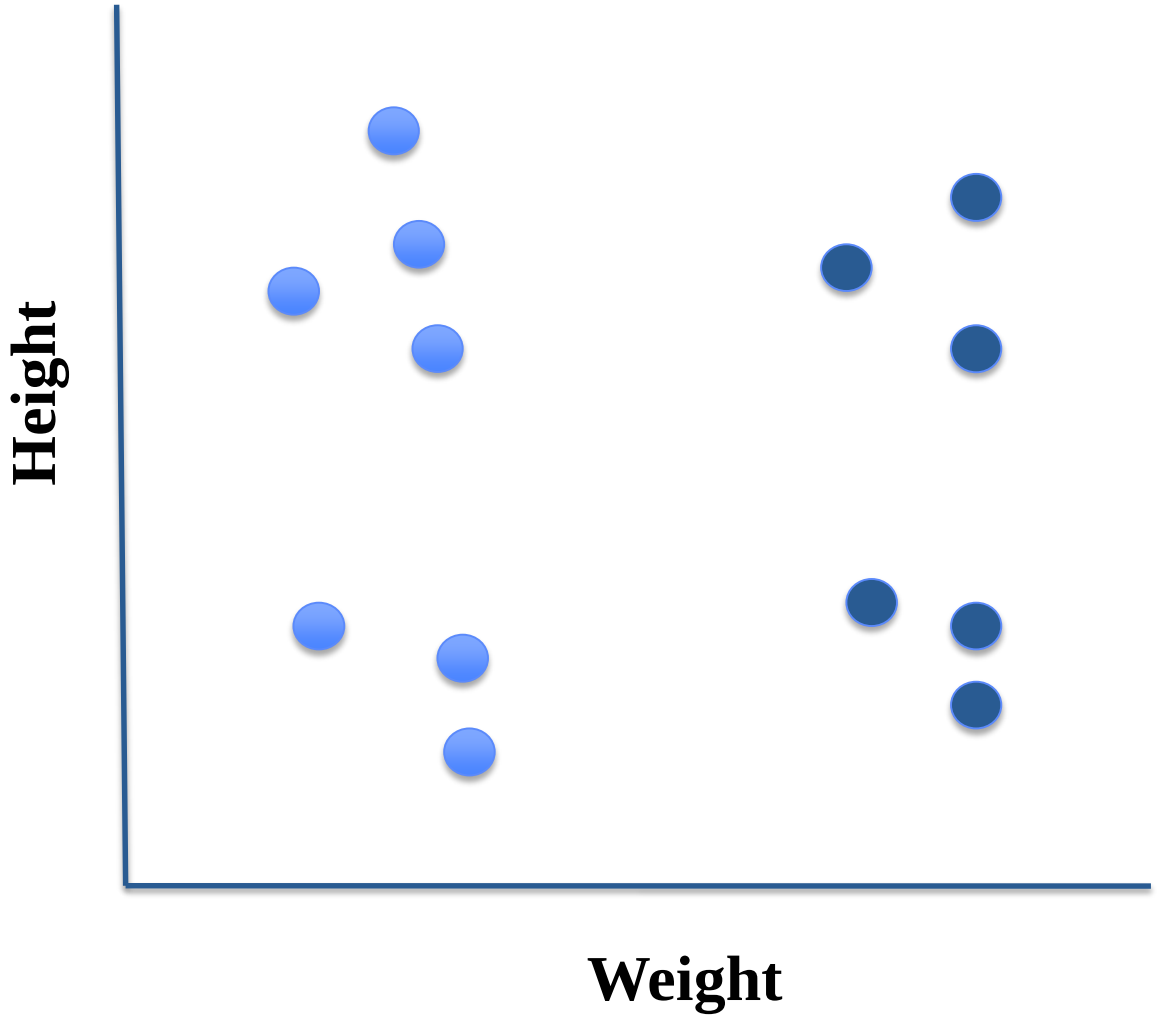
- Start with a set of feature vectors
- Goal: uncover some latent structure in the set of feature vectors
- Clustering the most common technique
  - Define some metric that captures how similar one feature vector is to another
  - Group examples based on this metric

Some Unlabeled 2D Data



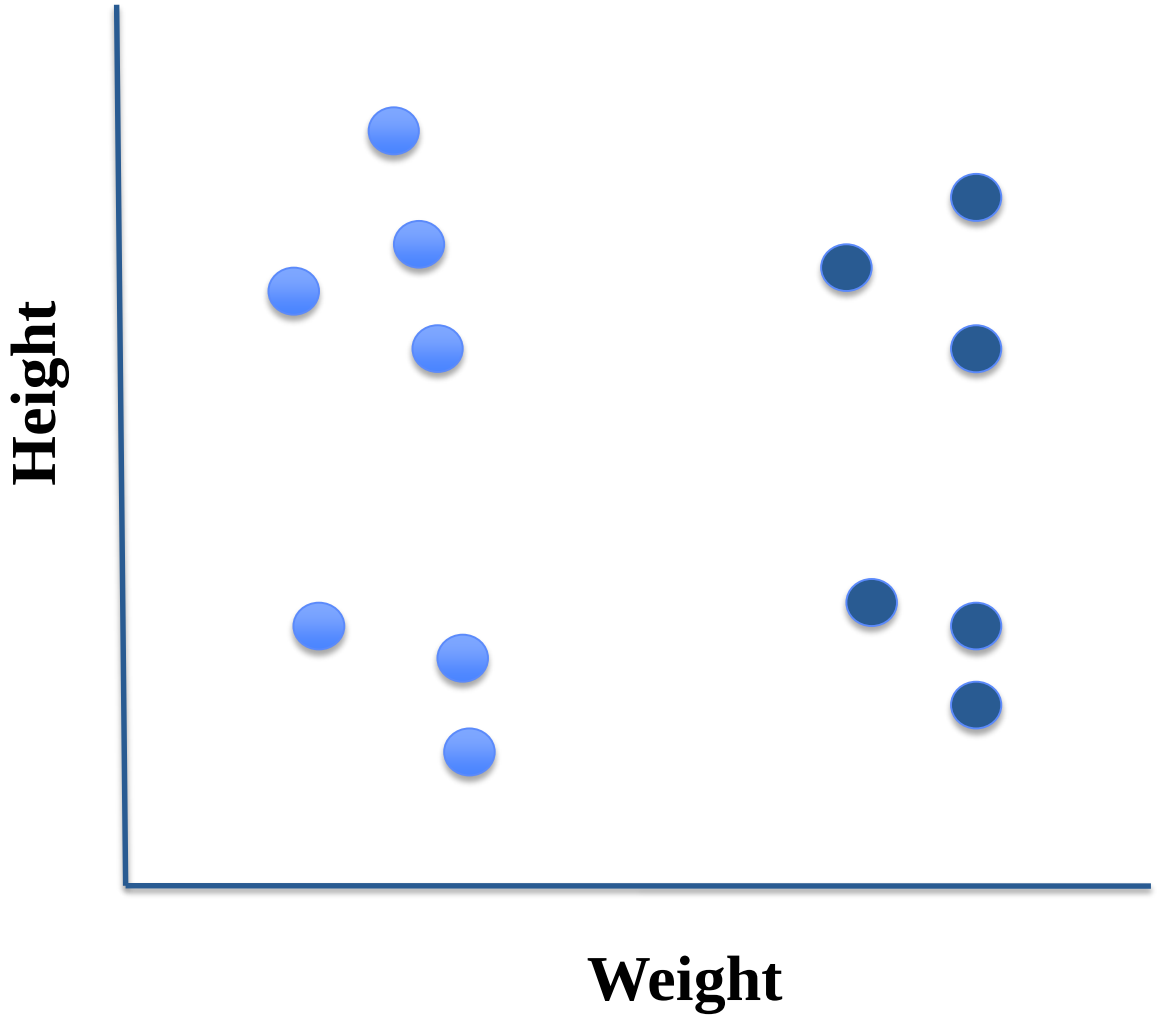
Use <weight, height>
as feature vector

Some Unlabeled 2D Data



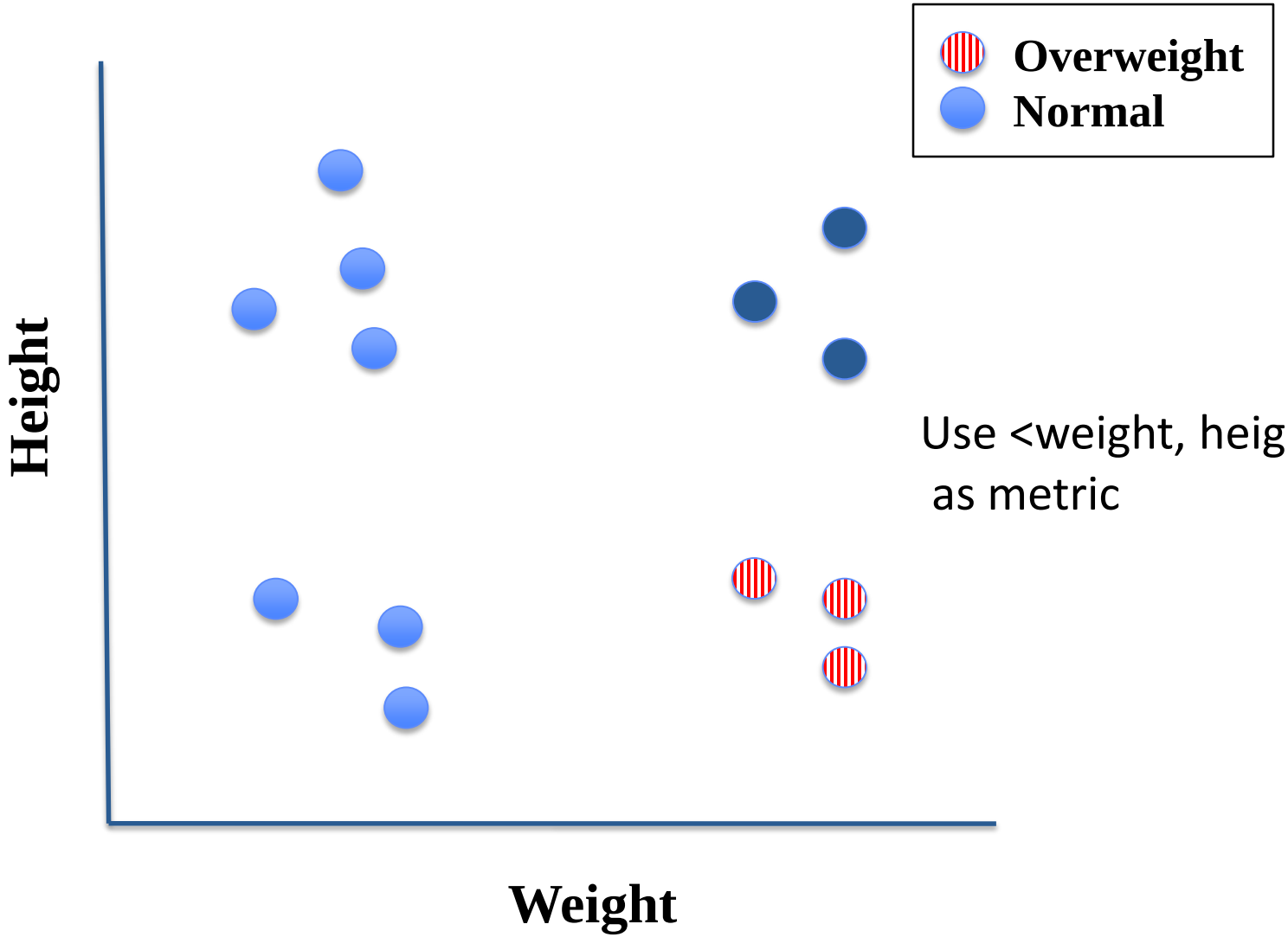
Use weight as feature vector

Some Unlabeled 2D Data



Use height as feature

Suppose Data Is Labeled



Use <weight, height>
as metric

Choosing Features

- Features never fully describe the situation
- Feature engineering
  - Represent examples by feature vectors that will facilitate generalization
  - Suppose I want to use 100 examples from past to predict, at the start of 6.00.2x, which students will pass the final exam
  - Some features surely helpful, e.g., their grade on the midterm, did they do the problem sets, etc.
  - Others might cause me to overfit, e.g., birth month
- Want to maximize ratio of useful input to irrelevant input
  - Signal-to-Noise Ratio (SNR)