Foundations of Machine Learning

Module 1: Introduction

Part c: Hypothesis Space and Inductive Bias

Sudeshna Sarkar

IIT Kharagpur

Inductive Learning or Prediction

- **Given** examples of a function *(X, F(X))*
  - **Predict** function *F(X)* for new examples *X*
- Classification

  F(X) = Discrete
- Regression

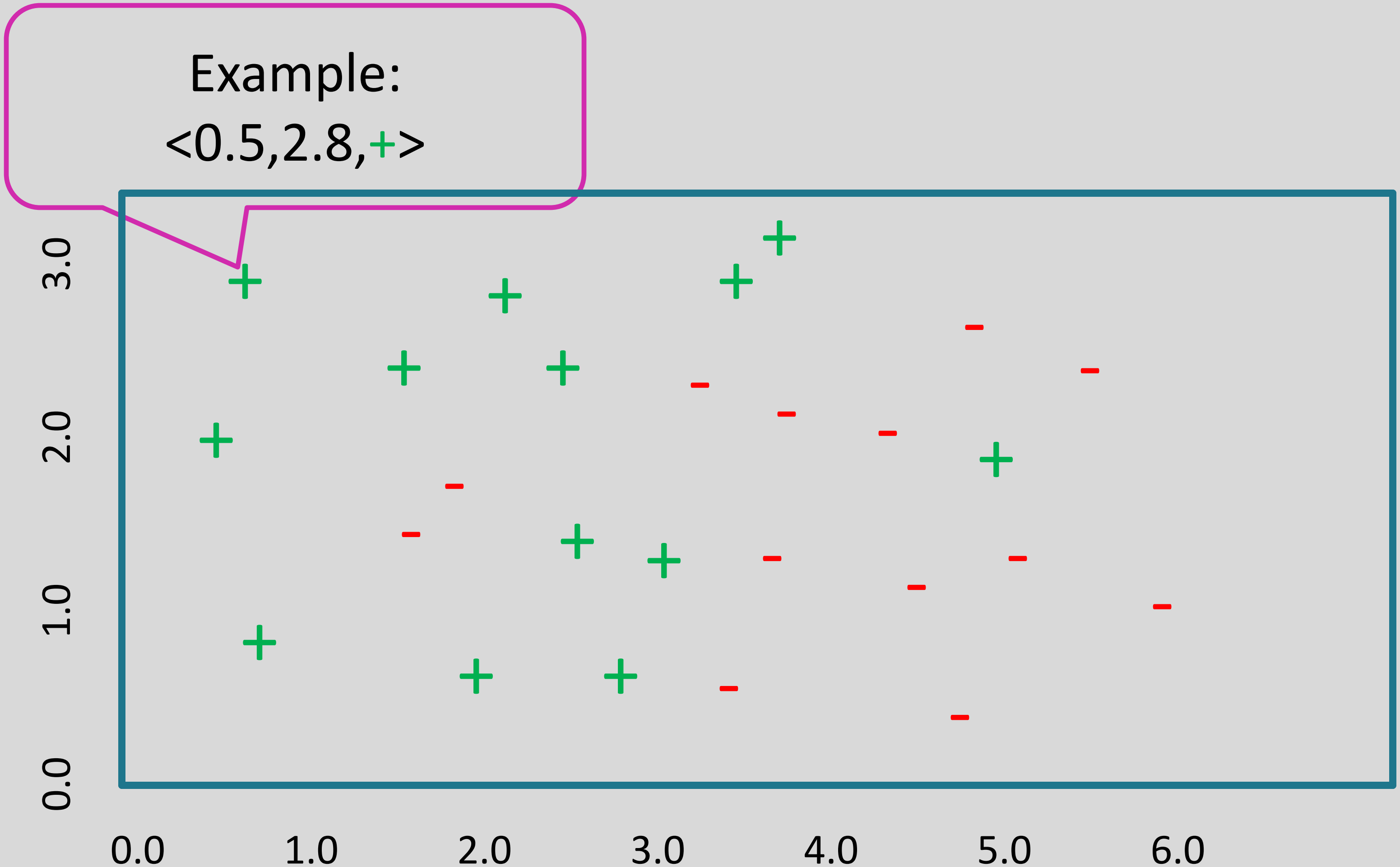
  F(X) = Continuous
- Probability estimation

  F(X) = Probability*(X):*

Features

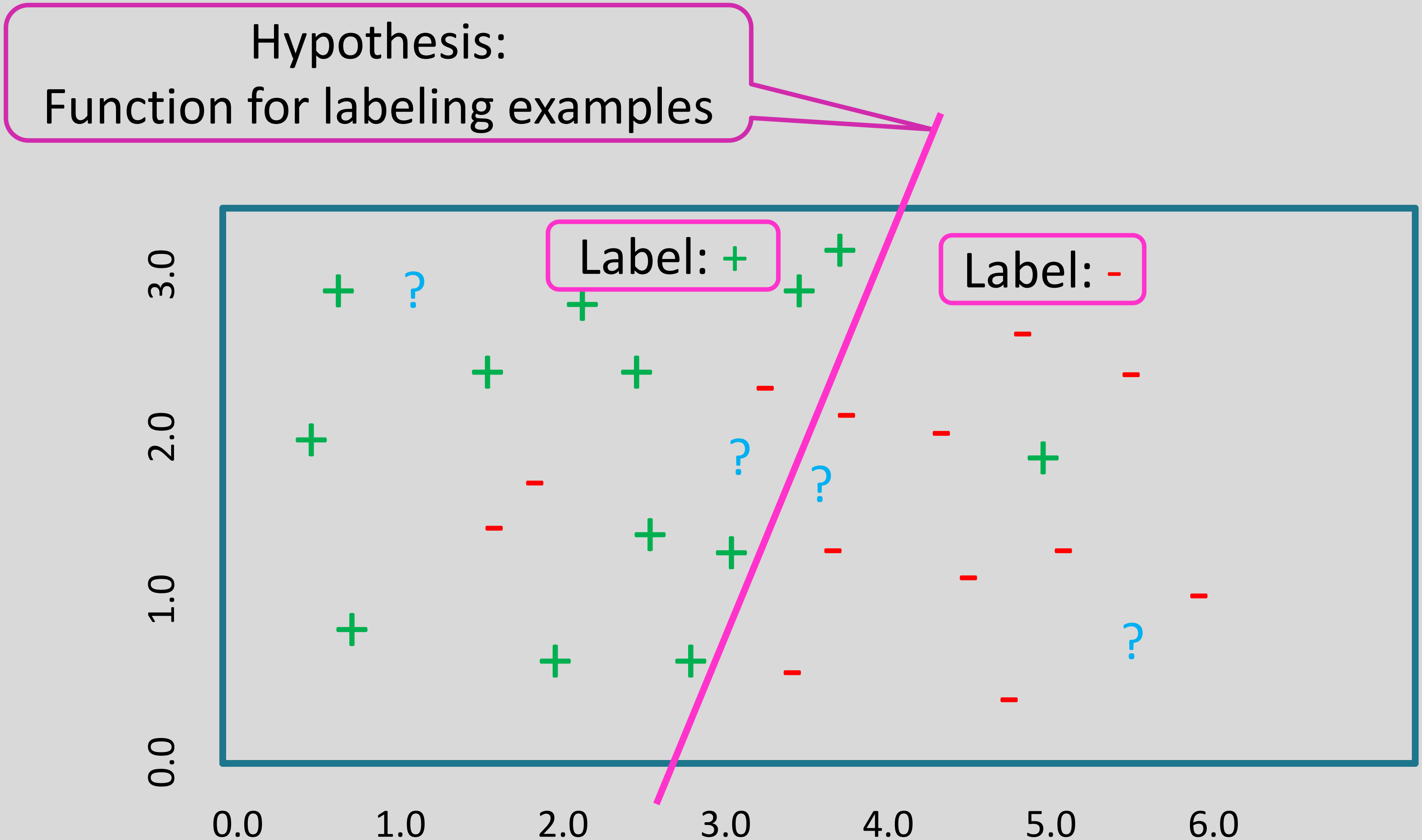
- **Features**: Properties that describe each instance in a quantitative manner.
- **Feature vector**: n-dimensional vector of features that represent some object

Feature Space



Slide by Jesse Davis: University of Washington

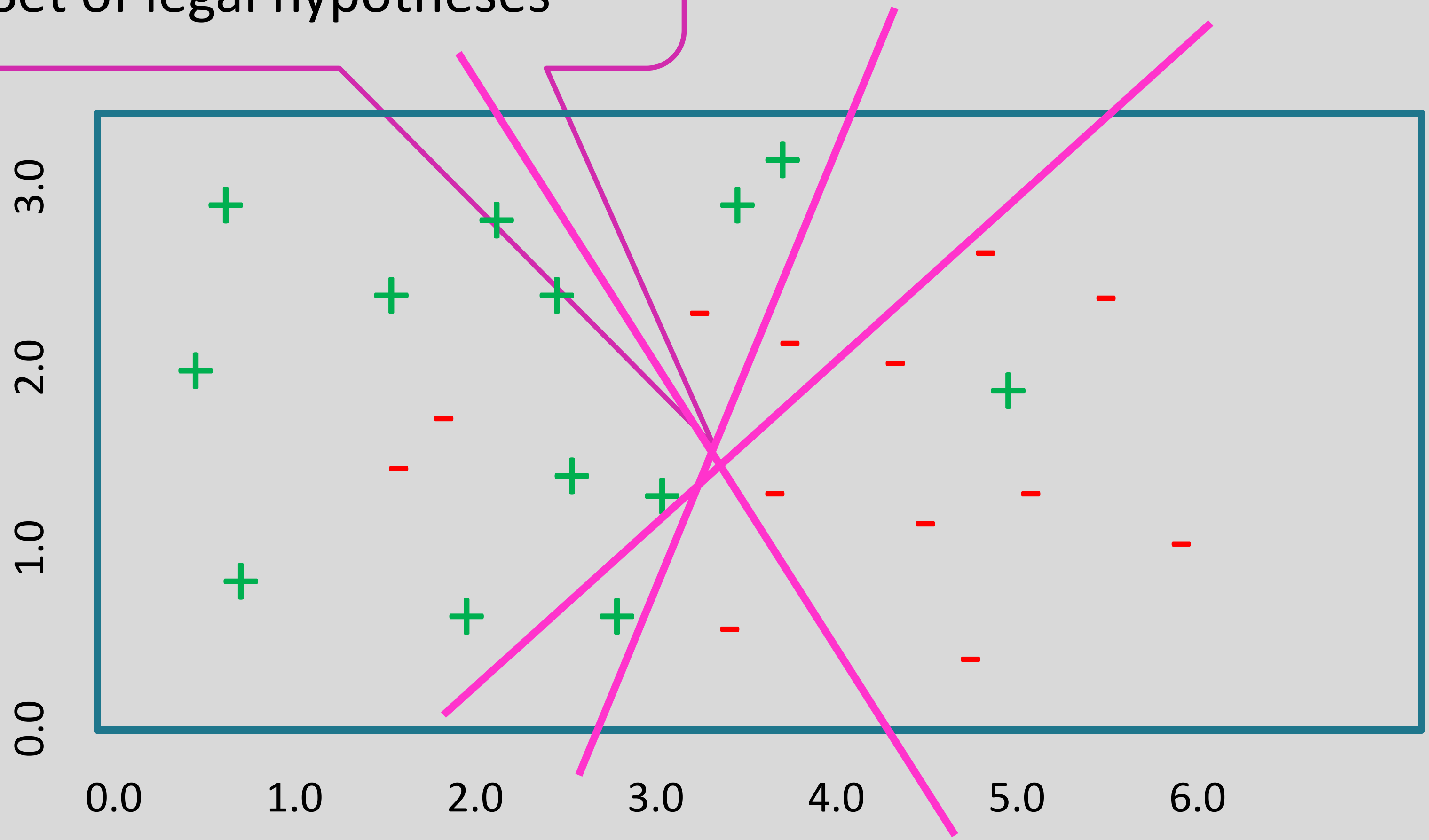
Terminology



Slide by Jesse Davis: University of Washington

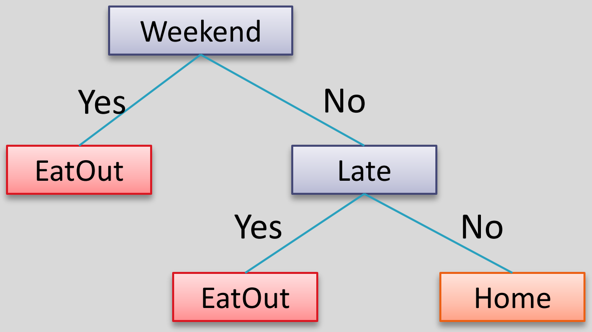
Terminology

Hypothesis Space:
Set of legal hypotheses

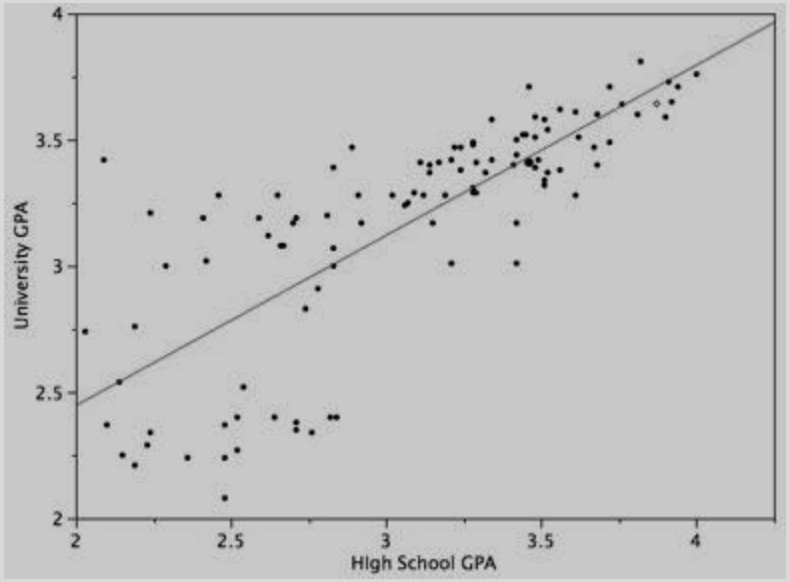


Representations

1. Decision Tree

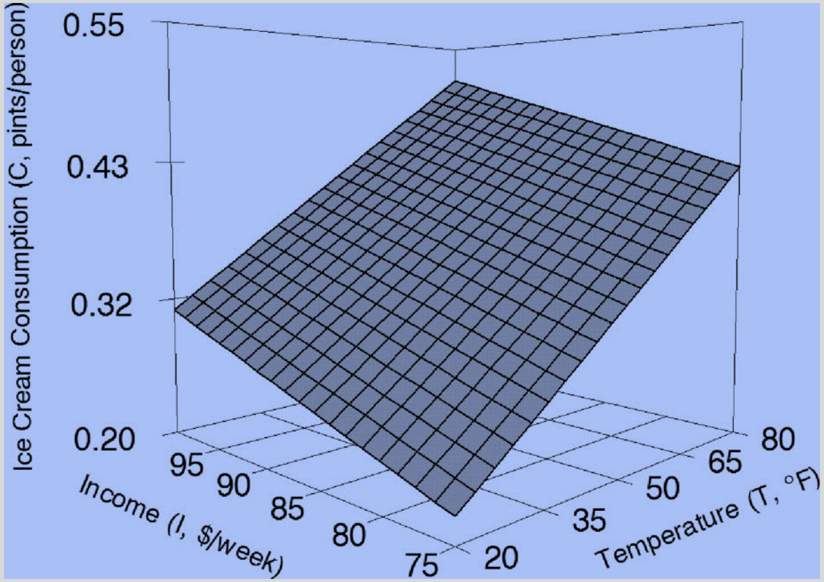


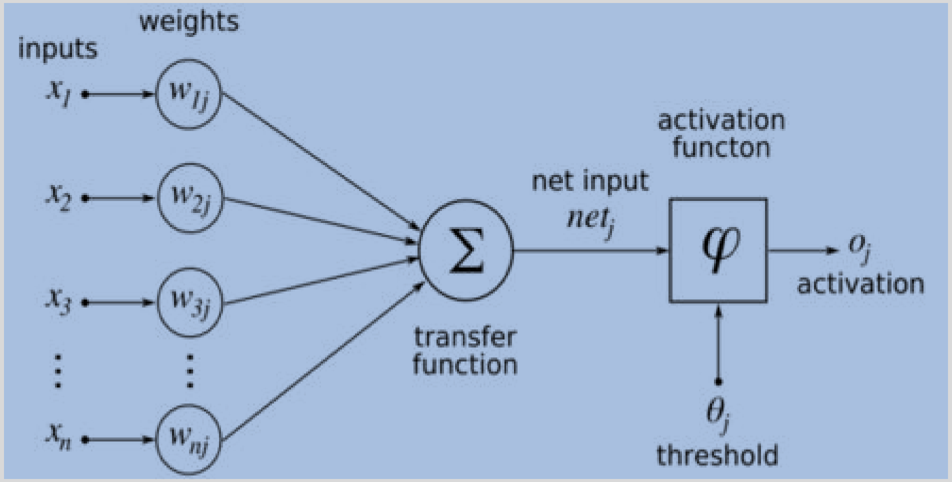
2. Linear function

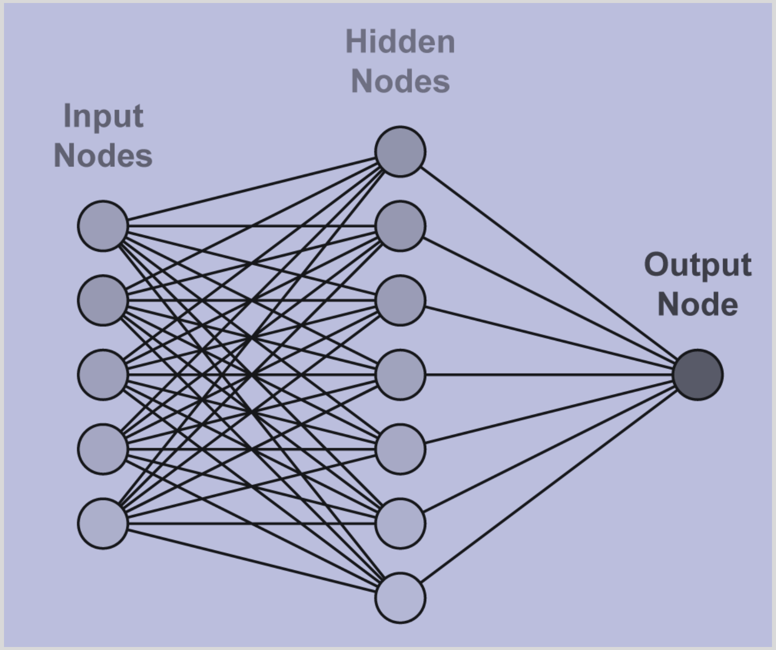


Representations

3. Multivariate linear function
4. Single layer perceptron
5. Multi-layer neural networks







Hypothesis Space

- The space of all hypotheses that can, in principle, be output by a learning algorithm.

- We can think about a supervised learning machine as a device that explores a "hypothesis space".
  - Each setting of the parameters in the machine is a different hypothesis about the function that maps input vectors to output vectors.

Terminology

- **Example (x,y):** Instance x with label y.
- **Training Data S:** Collection of examples observed by learning algorithm.
- **Instance Space X**: Set of all possible objects describable by features.
- **Concept c:** Subset of objects from X (c is unknown).
- **Target Function f:** Maps each instance $x \in X$ to target label $y \in Y$

Classifier

- Hypothesis h: Function that approximates f.
- Hypothesis Space $\mathcal{H}$: Set of functions we allow for approximating f.
- The set of hypotheses that can be produced, can be restricted further by specifying a language bias.
- Input: Training set $\mathcal{S} \subseteq X$
- Output: A hypothesis $h \in \mathcal{H}$

Hypothesis Spaces

- If there are 4 (N) input features, there are 2^{16} $\left(2^{2^N}\right)$ possible Boolean functions.
- We cannot figure out which one is correct unless we see every possible input-output pair $2^4(2^N)$

Example

Hypothesis language

1. may contain representations of all polynomial functions from X to Y if X = $\mathcal{R}^n$ and Y = $\mathcal{R}$,
2. may be able to represent all conjunctive concepts over X when $X = B^n$ and $Y = B$ (with B the set of booleans).

- Hypothesis language reflects an inductive bias that the learner has

Inductive Bias

- Need to make assumptions
  - Experience alone doesn't allow us to make conclusions about unseen data instances

- Two types of bias:
  - Restriction: Limit the hypothesis space

  - Preference: Impose ordering on hypothesis space

Inductive learning

- Inductive learning: Inducing a general function from training examples
  - Construct hypothesis *h* to agree with *c* on the training examples.
  - A hypothesis is consistent if it agrees with all training examples.
  - A hypothesis said to generalize well if it correctly predicts the value of *y* for novel example.
- *Inductive Learning is an Ill Posed Problem*:
  Unless we see all possible examples the data is not sufficient for an inductive learning algorithm to find a unique solution.

Inductive Learning Hypothesis

- Any hypothesis h found to approximate the target function c well over a sufficiently large set of training examples $\mathcal{D}$ will also approximate the target function well over other unobserved examples.

Learning as Refining the Hypothesis Space

- Concept learning is a task of searching an hypotheses space of possible representations looking for the representation(s) that best fits the data, given the bias.
- The tendency to prefer one hypothesis over another is called a **bias**.
- Given a representation, data, and a bias, the problem of learning can be reduced to one of search.

Occam's Razor

- A classical example of Inductive Bias

- the simplest consistent hypothesis about the target function is actually the best

Some more Types of Inductive Bias

- Minimum description length: when forming a hypothesis, attempt to minimize the length of the description of the hypothesis.

- Maximum margin: when drawing a boundary between two classes, attempt to maximize the width of the boundary (SVM)

Important issues in Machine Learning

- What are good hypothesis spaces?
- Algorithms that work with the hypothesis spaces
- How to optimize accuracy over future data points (overfitting)
- How can we have confidence in the result? (How much training data – statistical qs)
- Are some learning problems computationally intractable?

Generalization

- Components of generalization error
  - Bias: how much the average model over all training sets differ from the true model?
    - Error due to inaccurate assumptions/simplifications made by the model
  - Variance: how much models estimated from different training sets differ from each other

Underfitting and Overfitting

- Underfitting: model is too “simple” to represent all the relevant class characteristics
  - High bias and low variance
  - High training error and high test error
- Overfitting: model is too “complex” and fits irrelevant characteristics (noise) in the data
  - Low bias and high variance
  - Low training error and high test error