CS 5350/6350, DS 4350: Machine Learining Fall 2024

Mid-semester Review

February 19, 2024

What have we seen so far

This document summarizes the topics we have seen so far in the semester.

General supervised learning

- Supervised learning, instance spaces, label spaces, concept and hypothesis spaces
- Classification vs regression vs multi-class classification. What they are and how they are defined by the label space.
- Understanding why we need to restrict hypothesis spaces
- General issues in supervised learning: hypothesis spaces, representation (i.e features), learning algorithms

Decision trees

- What is a decision tree? What can they represent?
- How to predict with a decision tree
- Expressivity, counting the number of decision trees
- Dealing with continuous features
- Learning algorithm: The ID3 algorithm, entropy, information gain
- Overfitting (applicable not just to decision trees) and how to deal with it when training decision trees
- Variants of the entropy-based information gain measure.
- Dealing with missing features
- When to use decision trees

Linear classifiers

- What are linear classifiers? Why are they interesting?
- Linear separability. What can linear classifiers express? What can they not express? Examples of functions that are linearly separable and not linearly separable.
- Geometry of a linear classifier
- What is the role of the bias?
- Feature expansion to predict a broader set of functions

Mistake bound learning

- One way of asking how good is your classifier
- The general structure of an online learning algorithm: Repeated iterations of prediction followed by a possible update.
- Goal: Counting mistakes. What is a mistake bound algorithm?
- The CON and Halving algorithms. How many mistakes they make as a function of the size of the concept class and the ability to prove these results.
- Using the result from the Halving algorithm by counting the size of a concept class (especially simple Boolean functions)

Perceptron

- The original algorithm
- The geometry of the Perceptron update
- Variants of the Perceptron algorithm
- Margin of a classifier
- The Perceptron mistake bound theorem and its proof, and extending this proof to simple variants of the online Perceptron algorithm.
- Applying the Perceptron mistake bound to simple function classes like disjunctions by computing the values of R and γ for those classes.

Least mean squares regression

- What is linear regression? How is it different from classification?
- What is the least mean squares objective for regression?
- The idea of learning via minimizing a cost (or loss) function
- Deriving the gradient of the loss for least mean squares regression.
- Gradient and stochastic gradient descent for LMS. The difference between them.

Learning theory

- Assumption that train and test examples are drawn from the same distribution
- Batch learning and how it is different from mistake bound/online learning?
- PAC learning: How close will the approximation of the concept be (ϵ)? How sure are we that the learning algorithm will find a hypothesis with a good ϵ (δ)?
- Definition of PAC learning, efficient PAC learning
- Sample complexity and computational complexity
- Occam's razor for consistent learners (requires us to count sizes of hypothesis classes)