

CS 5350/6350: Machine Learning Fall 2015

Midterm Review

October 6, 2015

What you should know

General supervised learning

- Supervised learning, instance spaces, label spaces, concept and hypothesis spaces
- 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
- Dealing with missing features
- When to use decision trees

Nearest neighbors

- Instance based learning. How to predict? Importance of representation
- Different definitions of distance
- Dealing with symbolic features

- Choosing k for k -NN
- Practical aspects: Feature normalization could be important
- Advantages and disadvantages
- Voronoi diagrams
- Curse of dimensionality (applicable beyond nearest neighbors algorithm)

Linear classifiers

- What are they? Why are they interesting?
- What can they express? What can they not express?
- Geometry
- Feature expansion to predict a broader set of functions

Least mean squares regression

- What is it?
- The idea of learning via minimizing a cost function
- Gradient and stochastic gradient descent for LMS. Difference between them

Mistake bound learning

- One way of asking how good is your classifier
- The general structure of an online learning algorithm
- Goal: Counting mistakes. What is a mistake bound algorithm
- Halving algorithm
- Perceptron algorithm, geometry of the update, margin, Novikoff's theorem, variants
- Winnow algorithm, mistake bound, balanced winnow
- Perceptron vs. Winnow

Batch learning

- Assumption that train and test examples are drawn from the same distribution
- How it is different from mistake bound learning