Final Study Guide
COS 324 – Introduction to Machine Learning
Fall 2017, Princeton University

The following are study points. They serve as a reminder of concepts that you would need for the final exam.

Online learning and decision making.

- Be able to spell out the Consistent, Halving, and Weighted Majority algorithms for online learning.
- Given a labeled dataset, know how to repeatedly use the WM rule to make predictions.
- Understand how to apply online algorithms with the BOW model for text classification.

The power of randomness.

- Exercise the derivation of a lower bound for deterministic online learning.
- Spell out expected guarantee of RWM.
- Derive loss bound for different learning rates.

Learning from examples.

- Understand the ingredients of the statistical learning framework: distribution, labels, classifiers and their hypothesis classes, loss functions, training error, and test error.
- Understand the notion of overfitting: when generalization error greatly exceeds training error.

Learning theory and optimization.

- Definition of sample complexity of a finite hypothesis class.
- Computing sample complexity for simple cases as shown in class.
- The notion of linear separability for a labeled dataset.
- State and understand the Perceptron algorithm and its properties.

Convex optimization.

- Understand the definitions of convex functions and convex sets.
- Write down convex losses function that are relaxations of the $0-1$ loss.
- Describe a convex relaxation for linear classification, such as SVM.
- Compute the gradient of a given function.
- Be able to state the gradient descent algorithm and its guarantee for convex optimization over a constrained set.

- State the mathematical program corresponding to training a linear classifier.
- Describe an unbiased estimator for the gradient for a given convex loss function consisting of a sum of losses over a dataset. The expectation of the estimator is the true gradient. Be able to write this estimator in a simple setting given the loss function.
- State the stochastic gradient descent algorithm and be able to compute one step in a simple setting.
- Explain the advantage of SGD compared to GD.

Multi-class learning.

- Write down the logistic and max-margin multi-class loss functions.
- Derive the gradient of the logistic and max-margin multi-class losses.

Decision Trees.

- Spell out the maximum information gain (MIG) criterion for splitting a node in a decision tree.
- Given a simple dataset, know how to repeatedly apply the MIG rule to construct a decision tree.

Neural Netowork.

- Understand the notion of multi-layer neural network
- Understand inference in multi-layer neural network
- Understand gradient propagation for two-layer neural networks

Clustering.

- Understand the two steps clustering consists of.
- Rehearse the techniques studied *in class* and *homework* for associating an example with a cluster using a generalized distance $D(\cdot, \cdot)$
- Rehearse the techniques studied *in class* and *homework* for estimating a new center of cluster given the points associated with it.
- Revisit the clustering homework exercise in order to be able to find a center subject to constraints.

Boosting.

- Define the notion of weak learnability and contrast it with strong learnability.
- Define and understand the notion of edge of a weak learner.
- Give a bound on the number of weak learners required to construct a strong learner as a function of the edge of the weak-learners.