Midterm 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 midterm.

1. (Online learning and decision making.)
    - Be able to spell out the consistent, halving, and weighted majority algorithms for online learning.
    - Given a simple dataset, know how to repeatedly apply the WM rule to make a decision.
    - Understand how to apply online algorithms to the BOW model for text classification.
2. (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.
3. (Learning from examples.)
    - Know the components of the statistical learning framework (distribution, label, classifiers and their hypothesis classes, loss function, training error, test error).
    - Understand the notion of overfitting: when generalization error greatly exceeds training error.
4. (Learning theory and optimization.)
    - Definition of sample complexity of a finite hypothesis class.
    - Computing sample complexity for simple cases as done in class.
    - What it means for a labeled dataset of examples to be linearly separable.
    - State and understand the Perceptron algorithm.
5. (Convex optimization.)
    - Define a convex function and convex set.
    - Write down a convex loss function that is a relaxation 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.