CS 7545: Machine Learning Theory **Fall 2018**

Lecture 1: Course Overview and Linear Algebra Review

Lecturer: Jacob Abernethy *Scribes: Kimberly Caras, Srijan Sood*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

1.1 Course Overview

Instructor: Jacob Abernethy
TAs: Bhuvesh Kumar, Jun-Kun Wang
Course Website: mltheory.github.io

This course focuses on the mathematical aspect of Machine Learning. Familiarity with two of the following core topics is recommended:

- Advanced Linear Algebra
- Convex Optimization and Analysis
- Probability and Statistics

Course Outline The course will be split into two main segments.

1. Online Learning: Adversarial Framework for Learning
   - Sequential treatment of observation $\rightarrow$ data is not assumed to be IID
   - More natural setting, similar to real-world applications
   - Guarantees for prediction and learning algorithms despite assumption that data was potentially generated by adversary
   - Examples: Solving zero-sum games, Differential privacy
2. Classical Machine Learning: Statistical Learning Theory
   - Assumption is that data is IID: Independent and Identically Distributed
   - Examples:
     - Vapnik-Chervonenkis Theory
     - Uniform Deviation Bounds
     - Generalization Guarantees
     - Sauer's Lemma

Grade Breakdown

- 50% - Homeworks: 5 HWs
- 40% - Final Exam
- 10% - Participation This will be based on scribing lectures. Scribes work in pairs to take notes of a lecture (written and verbal information) and then typesetting them in LaTeX.

Important Policies

- Use Piazza for general questions and discussion with other students
- Use Piazza for communication with instructors on general topics (if personal question, make private post)
- HWs to be typed in LaTeX

1.2 General Notation

- x is usually a vector in $\mathbb{R}^n$
- Uppercase letters are usually matrices i.e. $M \in \mathbb{R}^{nxm}$.
- x_i is the i_{th} element of vector x.

1.3 Positive Semi-Definite (PSD) and Positive Definite (PD)

A square matrix $M \in \mathbb{R}^{nxn}$ is

- Positive Semi-Definite ($M \succeq 0$): $x^T Mx \geq 0$ for all $x \in \mathbb{R}^n$
- Positive Definite ($M \succ 0$): $x^T Mx > 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$

1.4 Norms

A function : $\|\cdot\| : \mathbb{R}^n \to [0, \infty)$ is called a norm if it satifies the following properties:

- Identity of indiscernibles: $\|x\| = 0$ if and only if $x = 0$
- Absolute homogeneity: $\|\alpha x\| = |\alpha|\|x\|$ for all $x \in \mathbb{R}$
- Triangle Inequality: $\|x + y\| \leq \|x\| + \|y\|$

1.4.1 Examples of Norms

- 2-norm: $\|x\|_2 = \sqrt{\sum_{i=1}^{n}(x_i)^2}$
- 1-norm: $\|x\|_1 = \sum_{i=1}^{n}(|x_i|)$
- ∞-norm: $\|x\|_\infty = \max_{i=1}^{n}(|x_i|)$
- p-norm: $\|x\|_p = \sqrt[p]{\sum_{i=1}^{n} x_i^p}$
- M-norm: $M \succ 0$, $\|x\|_M = \sqrt{x^T Mx}$
- 0-norm: $\|x\|_0 =$ number of non-zeros in x. 0 -norm is not a real norm as it violates absolute homogeneity.

1.5 Dual Norms

Given any norm $\|\cdot\|$, its dual norm: $\|\cdot\|_*$

$$\|x\|_* = \sup_{y, \|y\| \leq 1} (y^T x)$$

Example: Dual norm of $\|\cdot\|_2$ is the $\|\cdot\|_2$ norm

- $\|z\|_{2*} = \sup_{v:\|v\|\leq 1} v^T z = \sup_{v:\|v\|\leq 1} \frac{v^T}{\|v\|_2} z = \frac{z^T z}{\|z\|_2} = \|z\|_2$

(Exercise) Dual norm of the p-norm is the q-norm when $\frac{1}{p} + \frac{1}{q} = 1$
(Exercise) Prove $\|x\|_*$ is a norm
(Exercise) Dual of M norm ($M \succ 0$) is M^{-1} norm

1.6 Young's inequaility

Lemma 1.1 *Jensen's Inequality for Concave functions:*
If f is a concave function on set $\mathcal{X}$, then for $x_1, x_2 \in \mathcal{X}$ and $\alpha \in [0, 1]$, we have

$$f(\alpha x_1 + (1-\alpha)x_2) \geq \alpha f(x_1) + (1-\alpha) f(x_2)$$

Lemma 1.2 *Young's Inequality.*
For all $a, b \geq 0$ and $\frac{1}{p} + \frac{1}{q} = 1$ we have,

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

Proof:
$log(ab) = log(a) + log(b)$
Note that log is a monotonic increasing concave function

$= \frac{p}{p} log(a) + \frac{q}{q} log(b)$

$= \frac{log(a^p)}{p} + \frac{log(b^q)}{q}$

$\leq log(\frac{a^p}{p} + \frac{b^q}{q})$ by Jensen's Inequality ■