Lecture 1: Course Overview and Linear Algebra Review

Lecturer: Jacob Abernethy *Scribes: Chansoo Lee*

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

1.1 Important Information

- Homework must be typeset in LaTeX.
- No class on October 22nd (Thursday after Fall break) and November 24th (Tuesday before Thanksgiving).
- Mathematical maturity is required.
- Familiarity with at least two of the following topics is recommended: convex analysis, convex optimization, advanced statistics, probability theory, and machine learning.

1.2 Course overview

- Basics
  - Linear algebra
  - Convex analysis
  - Probability and statistics
- Batch Learning
  - PAC learning
  - Generalization error bounds
  - Rademacher complexity
  - VC Dimension
  - Uniform deviation bounds
  - Margin bounds
- Online Learning
  - Prediction with experts advice
  - Exponential Weights algorithm
  - Online convex optimization (OCO)
  - Applications in finance: Online portfolio section, option pricing, and gambling
  - Applications in differential privacy

1.3 Linear Algebra

We will use boldfaced lowercase letters to denote n-dimensional vectors, e.g. $\mathbf{x} \in \mathbb{R}^n$. The zero vector is denoted $\mathbf{0}$. The i-th coordinate of a vector $\mathbf{x}$ is denoted x_i. We use capital letters to denote matrices, e.g. $M \in \mathbb{R}^{n \times m}$.

Definition 1.1 (Norm) *A function* $\|\cdot\| : \mathbb{R}^n \to [0, \infty)$ *is called* norm*, if it satisfies the following properties:*

1. $\|\mathbf{x}\| = 0$ *if and only if* $\mathbf{x} = \mathbf{0}$
2. $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$
3. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

Definition 1.2 (PSD/PD) *A square matrix* $M \in \mathbb{R}^{n \times n}$ *is Positive Semi Definite (PSD), denoted* $M \succeq 0$*, if* $\mathbf{x}^\top M\mathbf{x} \geq 0$ *for all* $\mathbf{x} \in \mathbb{R}^n$*. A square matrix* $M \in \mathbb{R}^{n \times n}$ *is Positive Definite (PD), denoted* $M \succ 0$*, if* $\mathbf{x}^\top M\mathbf{x} > 0$ *for all* $\mathbf{x} \in \mathbb{R}^n$*,* $\mathbf{x} \neq \mathbf{0}$*.*

Examples

- 2-norm: $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$
- 1-norm: $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$
- ∞-norm: $\|\mathbf{x}\|_\infty = \max_{i=1}^n |x_i|$
- p-norm: $\|\mathbf{x}\|_p = (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}}$
- M-norm, for $M \succ 0$: $\sqrt{\mathbf{x}^\top M\mathbf{x}}$

Definition 1.3 (Dual norm) *Given any norm* $\|\cdot\|$*, its dual norm* $\|\cdot\|_*$ *is defined as*

$$\|\mathbf{x}\|_* = \sup_{\mathbf{y}:\|\mathbf{y}\|\leq 1} \mathbf{y}^\top\mathbf{x}.$$

Examples

- The dual of 2-norm is itself:
$$\sup_{v:\|v\|_2\leq 1} v^\top z = \frac{z^\top}{\|z\|_2} z = \|z\|_2.$$
- The dual of p-norm is q-norm, where $\frac{1}{p} + \frac{1}{q} = 1$. This includes the $p = 1$, $q = \infty$ pair.
- The dual of M-norm is M^{-1}-norm.

Lemma 1.4 (Young's inequality) *For all* $a, b \geq 0$*,* $ab \leq \dfrac{a^p}{p} + \dfrac{b^q}{q}$

Proof: By Jensen's inequality,

$$\log ab = \log a + \log b = \frac{1}{p}\log a^p + \frac{1}{q}\log b^q \leq \log\left(\frac{1}{p}a^p + \frac{1}{q}b^q\right)$$

■

Theorem 1.5 (Hölder's inequality) *For any* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$*,* $x^\top y \leq \|x\|_p\|y\|_q$ *where* $\frac{1}{p} + \frac{1}{q} = 1$

Proof: By Young's Inequality,

$$\frac{\mathbf{x}^\top \mathbf{y}}{\|\mathbf{x}\|_p \|\mathbf{y}\|_q} \leq \sum_{i=1}^{n} \frac{|x_i y_i|}{\|\mathbf{x}\|_p \|\mathbf{y}\|_q} \leq \sum_{i=1}^{n} \frac{1}{p} \frac{|x_i|^p}{\|\mathbf{x}\|_p^p} + \frac{1}{q} \frac{|y_i|^q}{\|\mathbf{x}\|_q^q} = \frac{1}{p} + \frac{1}{q} = 1.$$

■

Corollary 1.6 (Cauchy-Schwarz Inequality) *For any* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $x^\top y \leq \|x\|_2 \|y\|_2$.

Proof: It follows from Theorem 1.5. Alternatively, we can prove it by observing

$$\begin{aligned} 0 &\leq \|(\|\mathbf{x}\| y - \|\mathbf{y}\| \mathbf{x})\|^2 \\ &\leq 2\|\mathbf{x}\|^2 \|\mathbf{y}\|^2 - 2\|\mathbf{x}\| \|\mathbf{y}\| \mathbf{x}^\top \mathbf{y}. \end{aligned}$$

■

Theorem 1.7 (Generalized Hölder's inequality) *For any* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\mathbf{x}^\top \mathbf{y} \leq \|\mathbf{x}\| \|\mathbf{y}\|_\star$ *for any norm* $\|\cdot\|$. *Theorem 1.5 follows as a corollary.*

Proof: Using the fact that $\|\frac{\mathbf{x}}{\|\mathbf{x}\|}\| = 1$,

$$\mathbf{x}^\top \mathbf{y} = \|\mathbf{x}\| \left(\left(\frac{\mathbf{x}}{\|\mathbf{x}\|} \right)^\top \mathbf{y} \right) \leq \|\mathbf{x}\| \left(\sup_{\mathbf{z}, \|\mathbf{z}\| \leq 1} \mathbf{z}^\top \mathbf{y} \right) = \|\mathbf{x}\| \|\mathbf{y}\|_*$$

■