New York University Tandon School of Engineering
Computer Science and Engineering

CS-GY 6763: Midterm Practice.

Logistics

- Exam will be held in class on **Tuesday, 10/26** starting at 2:05pm. Please arrive on time!
- Teal will give a lecture for the second half of class after a 15 minute break.
- You will have **one hour** to answer a variety of short answer and longer form questions.
- You can bring a one page sheet of paper (two-sided if you want) with notes, theorems, etc. written down for reference.
- I will be in the room to answer any questions.
- If you have permission from the department to be remote this semester, you will take the exam over Zoom. Arrive to the Zoom meeting at 2:00pm, download the exam from Gradescope, and then write your answer on whatever paper you wish. After the exam, you will have 15 minutes to scan and upload the solutions to Gradescope, remaining on Zoom the entire time. Please use a scanning app on your phone instead of just snapping a photo. I will have headphones on, so you can ask questions during the exam.

Concepts to Know

Random variables and concentration.

- Linearity of expectation and variance.
- Indicator random variables and how to use them.
- Markov's inequality, Chebyshev's inequality (should know from memory).
- Union bound (should know from memory).
- Chernoff and Bernstein bounds (don't need to memorize the exact bounds, but can apply if given).
- General idea of law of large numbers and central limit theorem.
- The probability that a normal random variables $\mathcal{N}(0, \sigma^2)$ falls further than $k\sigma$ away from its expectation is $\leq O(e^{-k^2/2})$.

Hashing, Dimensionality Reduction, High Dimensional Vectors

- Random hash functions.
- Random hashing for frequency estimation.
- Random hashing for distinct elements estimation.
- MinHash for Jaccard similarity estimation.
- Locality sensitive hash functions.
- MinHash and SimHash for Jaccard Similarity and Cosine Similarity.
- Adjusting false positive rate and false negative rate in an LSH scheme.
- Statement of Johnson-Lindenstrauss lemma (know from memory).
- Statement of *distributional* JL lemma and how it can be used to prove JL.

High dimensional geometry

- How to draw a random unit vector from the sphere in d dimensions (draw $\mathbf{x}$ with all entries i.i.d. $\mathcal{N}(0,1)$ and normalize it).
- How does $\|x - y\|_2^2$ relate to $\langle x, y\rangle$ if x and y are unit vectors?
- How many mutually orthogonal unit vectors are there in d dimensions?
- There are $2^{\theta(\epsilon^2 d)}$ nearly orthogonal unit vectors in d dimensions (with $\langle x, y\rangle \leq \epsilon$). Know roughly how prove this fact using the *probabilistic method*, which required a an exponential *concentration inequality* + *union bound*.
- Know how to prove that all but an $2^{\theta(-\epsilon d)}$ fraction of a balls volume in d dimensions lies in a spherical shell of width ϵ near its surface.
- The surface area/volume ratio *increases* in high dimensions.
- The cube volume/ball volume ratio *increases* in high dimensions.

Convex optimization

- Definition(s) of convex function.
- Definition of convex set.
- Gradient descent basic update rule.
- Definitions of G-Lipschitz, β-smooth, α-strongly convex.
- Definition of condition number.
- How much time does it take to multiply an $n \times d$ matrix by a $d \times m$ matrix?
- Be able to compute gradients of basic functions from $\mathbb{R}^d \to \mathbb{R}$.
- Definition(s) of convex function.

Practice Problems

Random variables and concentration.

1. Show that for any random variable X, $\mathbb{E}[X^2] \geq \mathbb{E}[X]^2$.
2. Show that for independent X and Y with $\mathbb{E}[X] = \mathbb{E}[Y] = 0$, $\text{var}[X \cdot Y] = \text{var}[X] \cdot \text{var}[Y]$.
3. Given a random variable X, can we conclude that $\mathbb{E}[1/X] = 1/E[X]$? If so, prove this. If not, give an example where the equality does not hold.
4. Indicate whether each of the following statements is **always** true, **sometimes** true, or **never** true. Provide a short justification for your choice.
   (a) $\Pr[X = s \text{ and } Y = t] > \Pr[X = s]$. ALWAYS SOMETIMES NEVER
   (b) $\Pr[X = s \text{ or } Y = t] \leq \Pr[X = s] + \Pr[Y = t]$. ALWAYS SOMETIMES NEVER
   (c) $\Pr[X = s \text{ and } Y = t] = \Pr[X = s] \cdot \Pr[Y = t]$. ALWAYS SOMETIMES NEVER
5. Assume there are 1000 registered users on your site $u_1, \ldots, u_{1000}$, and in a given day, each user visits the site with some probability p_i. The event that any user visits the site is independent of what the other users do. Assume that $\sum_{i=1}^{1000} p_i = 500$.
   (a) Let X be the number of users that visit the site on the given day. What is $E[X]$?

(b) Apply a Chernoff bound to show that $Pr[X \geq 600] \leq .01$.

(c) Apply Markov's inequality and Chebyshev's inequality to bound the same probability. How do they compare?

6. Give an example of a random variable and a deviation t where Markov's inequality gives a tighter upper bound than Chebyshev's inequality.

Hashing, Dimensionality Reduction, High Dimensional Vectors

1. Suppose there is some unknown vector $\boldsymbol{\mu} \in \mathbb{R}^d$. We receive noise perturbed random samples of the form $\mathbf{Y}_1 = \boldsymbol{\mu} + \mathbf{X}_1, \ldots, \mathbf{Y}_k = \boldsymbol{\mu} + \mathbf{X}_k$ where each $\mathbf{X}_i$ is a random vector with each of its entries distributed as an independent random normal $\mathcal{N}(0,1)$. From our samples $\mathbf{Y}_1, \ldots, \mathbf{Y}_k$ we hope to estimate $\boldsymbol{\mu}$ by $\tilde{\boldsymbol{\mu}} = \frac{1}{k}\sum_{i=1}^{k} \mathbf{Y}_i$.

   (a) How many samples k do we require so that $\max_{i=1,\ldots,d} |\boldsymbol{\mu}_i - \tilde{\boldsymbol{\mu}}_{\boldsymbol{i}}| \leq \epsilon$ with probability 9/10?

   (b) How many samples k do we require so that $\|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}\|_2 \leq \epsilon$ with probability 9/10?

2. Let $\boldsymbol{\Pi}$ be a random Johnson-Lindenstrauss matrix (e.g. scaled random Gaussians) with $O(\log(1/\delta)/\epsilon^2)$ rows. Prove that with probability $(1-\delta)$,

$$\min_{\mathbf{x}} \|\boldsymbol{\Pi}\mathbf{A}\mathbf{x} - \boldsymbol{\Pi}\mathbf{b}\|_2^2 \leq (1+\epsilon) \min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$$

   Is the following also true with high probability?

$$(1-\epsilon) \min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 \leq \min_{\mathbf{x}} \|\boldsymbol{\Pi}\mathbf{A}\mathbf{x} - \boldsymbol{\Pi}\mathbf{b}\|_2^2$$

Convex optimization

From *Convex Optimization* book (`https://web.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf`):

- **Exercises:** 3.7, 3.10, 3.11 (first part), 3.21 (lots of other problems if you want more practice, but many are on the harder side)

1. Let $f_1(x), \ldots, f_n(x)$ be β-smooth convex functions and let $g(x) = \frac{1}{n}\sum_{i=1}^{n} f_i(x)$ be their average. Show that g is β-smooth.

2. Let $f : \mathbb{R} \to \mathbb{R}$ be a β-smooth, α-strongly convex function. Let $g(x) = f(c \cdot x)$ for some constant $0 < c < 1$. How does g's smoothness and strong convexity compare to that of f? How about g's condition number?

3. Let $f(x) = x^4$. Is f G-Lipschitz for finite G? Is f β-smooth for finite B?