New York University Tandon School of Engineering
Computer Science and Engineering

CS-GY 9223D: Midterm Practice.

Logistics

- Exam will be released at 12am ET on **Monday, 10/26.**
- You can complete the exame during any **2 hour window** until midnight ET on **Friday, 10/30.**
- Solutions can be typed or handwritten, as long as they are converted to a PDF. Clearly mark solutions so I know what problem they are for.
- The exam will be administered on Gradescope or a similar platform (stay tuned for details). After you download it, you will be able to submit a PDF of your solutions until exactly 2 hours, 15 minutes after downloading, which allows 15 min. buffer time to scan and upload your work.
- You can use any resources from the class (your own notes, my notes, etc.) but no outside resources.
- Questions can be asked privately on Piazza, but I likely will not be able to respond to them all in time. So if something is not clear about a problem, very clearly state any assumptions you needed to make, and move forward with writing your solution. You let me know on Piazza after submitting if there was some confusion, so I can take this into account when grading.
- *Absolutely no discussion or collaboration* with other students is allowed. This midterm must represent your own work. Please do no take any risks here, as noticing overlap in solutions is very easy, and I don't want anyone to get in trouble with the school.

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 load balancing.
- Random hashing for distinct elements estimation.
- What's a streaming algorithm? What's a sketching algorithm?
- Locality sensitive hash functions.

- MinHash and SimHash for Jaccard Similarity and Cosine Similarity.
- S-curve tuning.
- Statement of Johnson-Lindenstrauss lemma (know from memory).
- Statement of *distributional* JL lemma and how it can be used to prove JL.

Convex optimization

- Gradient descent, stochastic gradient descent, and online gradient descent algorithms.
- Definition(s) of convex function.
- When can stochastic gradient descent be faster than gradient descent?
- Definitions of G-Lipschitz, β-smooth, α-strongly convex.
- How do these properties effect the convergence of gradient descent?
- Be able to compute gradients of basic functions from $\mathbb{R}^d \to \mathbb{R}$.
- How much time does it take to multiply an $n \times d$ matrix by a $d \times m$ matrix?

Practice Problems

Random variables and concentration.

From *Foundations of Data Science* book (https://www.cs.cornell.edu/jeh/book.pdf):

- **Exercises:** 2.1, 2.5, 2.9, 2.28, 2.29, 2.28, 2.36, 2.41, 6.6, 6.10.

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 $u_1, \ldots, u_1 000$, 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 $Pr[X 600].01$.
   (c) Apply Markovs inequality and Chebyshevs inequality to bound the same probability. How do they compare?
6. Give an example of a random variable and a deviation t where Markovs inequality gives a tighter bound than Chebyshevs inequality.

Hashing, Dimensionality Reduction, High Dimensional Vectors

1. SimHash can be applied to binary vectors $y, q \in \{0,1\}^d$. How does doing so compare to MinHash? How is the Jaccard similarity $J(y,q)$ related to the cosine similarity $\cos(\theta(y,q))$?

2. Suppose there is some unknown vector $\boldsymbol{\mu}$. 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 $\max |\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}| \leq \epsilon$ with probability $9/10$?

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

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

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

   Is the following also true with high probability?

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

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 si)

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?