END 1:13PM

E14161023891060

CS-UY 3943.00B	26 OCT WED	11:00 AM 12:53 PM (113) 1 Hr 53 Min

Moses Center for Student Accessibility - Alternative Testing - Instruction Form
Please Note: Check Student or Proctor Identification Before Handing Out Exam

ID: N15186861 **MIDTERM**

Student and Class Information

Student: **Amatya Pathak**
Pronoun: **he, him, his, his, himself**
Course: CS-UY 3943.00B - SELECTED TOPICS IN COMP SCI (CRN: 16469)
Advisor: Matt Orlewicz
Date & Time: 10/26/2022 at 11:00 AM
Location: 2 Metrotech Room 970
Proctor **Not Specified**
Instructor: **Christopher Paul Musco**
Email: cpm303@nyu.edu
Phone: **(401) 578 - 2541**

Class Meeting Time(s):

- Days: W Time: 11:00 AM - 01:30 PM Location: RGSH 325

Alternative Testing Agreement Note:

Not Specified

Approved Accommodation(s)

- Extra Time 1.50x

Exam Notes:

Not Specified

Actual Exam Time		Pick Up From:
Start	End	
		Delivered By:

Drop Off Received on (Date & Time): ________________

Name: ________________ Sign: ____________________

Exam Instructions Detail

1. **How are your exams going to be administered?**
   In-person
2. **How will Moses receive the exam?**
   I will email the exam to [Email: csaexams@nyu.edu]
3. **Are any supplemental materials allowed on the exam?**
   - One sheet of notes

   Note: Sheet of notes can be two sided piece of paper.
4. **Are students allowed scrap or lined paper?**
   Yes- for scrap use only , **Note:** If more space is needed on the exam paper, answers can be written on blank paper and attached.
5. **Will be you available for student questions?**
   I will be available for questions and here is who to contact and how to contact them (Include TAs, ofter professors if coordinated exam, etc). [R] , **Note:** I can be contacted at 401-578-2541 but there is a chance I am not available to take the call/busy with answering another students question. So I would strongly advise the student try to avoid asking questions and if they are confused about a definition, simply state their assumption/confusion in writing and move on.
6. **How do you want to receive the completed exam? Please specify any extra details below**
   The Moses Center will upload the exam to the Instructor Portal.
7. **Does your exam include audio or visual components?**
   - No, I do not

Amatya Porthale

NYU | TANDON SCHOOL OF ENGINEERING

New York University Tandon School of Engineering
Computer Science and Engineering

CS-GY 6763: Midterm Exam.
Wednesday Oct. 26th, 2021.
55 points total

You have 1 hour, 15 minutes to take the exam.
Show your work to receive full (and partial) credit.

1. Short answer. (**15pts – 3pts each**)

Indicate whether each of the following statements is always true, sometimes true, or never true. To receive full credit, provide a **SHORT JUSTIFICATION OR EXAMPLE** to explain your choice.

(a) For random variables X and Y, $\mathbb{E}[10X - 3Y] = 10\mathbb{E}[X] - 3\mathbb{E}[Y]$.

ALWAYS SOMETIMES NEVER

$E[X] = \int_0^\infty x f(x)\,dx$

$\int_0^\infty (x-Y) f(x)\,dx$

$\int_0^\infty x f(x) - Y f(x)\,dx$

$\int_0^\infty x f(x)\,dx - \int_0^\infty Y f(x)\,dx$

This is true if X and Y are known to be independent, since the integral of the sum of random variables is the sum of individual integrals, plus scalars can be extracted

(b) If random variables X, Y have the same variance σ^2, the average $Z = \frac{1}{2}(X+Y)$ has $\text{Var}[Z] \geq \frac{1}{2}\sigma^2$.

ALWAYS SOMETIMES NEVER

The true lower bound for variance would be $Var[Z] \geq \frac{1}{4}\sigma^2$, and so a higher bound misses the cases where $\frac{1}{4}\sigma^2 \leq Var(Z) \leq \frac{1}{2}\sigma^2$

(c) For a random variable Y that always takes positive integer values $1, 2, 3, \ldots$ we have $\Pr[Y \geq z] \leq \frac{\mathbb{E}[Y]}{z}$.

ALWAYS SOMETIMES NEVER

Markov's Inequality Definition

(d) Increasing the number of tables in a locality sensitive hashing scheme increases the expected number of false negatives.

ALWAYS SOMETIMES NEVER

Not always; a good uniform random hashing function can ~~uniformly~~ approximate a uniform distribution of hash values, and if the number of spots to hash to (n) is greater than the number of elements to be hashed, $E[\text{false negatives}] \xrightarrow{n\to\infty} 0$.

(e) Let $X_1, \ldots, X_n$ be random variables (not necessarily independent). $\Pr[\min_j X_j \leq z] \leq \sum_{i=1}^{n} \Pr[X_j \leq z]$.

$\leq n \cdot \frac{1}{n} = 1$

sum of

ALWAYS SOMETIMES NEVER

$P[\min_j \leq z] \geq \sum_{j=1}^{n} P[X_j \leq z]$

we are only sampling 1 X_j in both sides of the inequality, and the RHS is equal to $\frac{1}{n} \cdot n = 1$ ≤ 1

Since $\min_j X_j$ can be larger than X_j, there is a smaller ($\leq$ RHS) probability that $\min X_j \cdot X_j \leq Z$

3. Locality sensitive hash for hamming similarity. (**10pts**)

The giant internet company Popflix operates a movie streaming service that offers a total of d movies. They keep track of the movies viewed by each subscriber to the service by storing a length d binary vector with a 1 in all entries corresponding to movies that subscriber has watched. All other entries are set to 0. Popflix would like to efficiently find other subscribers with similar viewing habits using an LSH scheme. For two length d binary vectors $\mathbf{q}, \mathbf{y} \in \{0,1\}^d$, they model viewing similarity by the hamming similarity:

$$s(\mathbf{x}, \mathbf{y}) = 1 - \frac{\|\mathbf{q} - \mathbf{y}\|_0}{d}.$$

Above $\|\mathbf{q} - \mathbf{y}\|_0$ is the hamming distance, $\|\mathbf{q} - \mathbf{y}\|_0 = \sum_{i=1}^{d} |q_i - y_i|$. q_i and y_i denote the i^{th} entries of $\mathbf{q}$ and $\mathbf{y}$, respectively. For example, the hamming similarity between the follow two vectors is $2/3$.

$$\mathbf{q} = [1,0,0,1,1,0]$$
$$\mathbf{y} = [1,0,1,1,0,0]$$

differ in $\frac{2}{6} = \frac{1}{3}$

same in $2/3$

(a) (6pts) To implement their LSH scheme, Popflix first needs a locality sensitive hash function $h : \{0,1\}^d \to \{1,\ldots,n\}$ that maps user vectors into a table of size n. We construct h as follows: Let j be a uniform random integer in $\{1,\ldots,d\}$ and define the function $c : \{0,1\}^d \to \{0,1\}$ as $c(\mathbf{x}) = x_j$. Additionally, let g be a uniform random hash function from $\{0,1\} \to \{1,\ldots,n\}$. Finally, let:

$$h(\mathbf{x}) = g(c(\mathbf{x})).$$

Prove that h is a locality sensitive hash function for hamming similarity. **Hint:** Write down an expression for $\Pr[h(\mathbf{q}) == h(\mathbf{y})]$.

$P[h(q)==h(y)]$

$h(x) = g(c(x))$ $\{0,1\}^d \to \{0,1\}$ maps to same space, $P[\text{collision}]$ if elements >1 $\to 0$

$P[h(q)==h(y)] = \left(\frac{1}{2}\right)^n$, given binary vectors

$P[h(q)== h(y)) = \sum_{i=1}^{n} \mathbb{1}(h(q)==h(y))$

Locally sensitive

(b) (4pts) What are **two reasons** Popflix might **not want to use** locality sensitive hashing for similarity search, and might instead perform a linear scan.

When several movies hash together in the same bucket/table row may grow ineffient and O(n)-like, so it would be simpler to just use linear search in the first place. Another reason for linear scanning is if the order of the vector matters; perhaps an algorithm only needs to scan over a certain day to see how many people are watching similar movies

2. Mark-and-Recapture, The Missing Analysis (15pts)

Recall the mark-and-recapture estimator introduced in Lecture 1. We collect m items $d_1, \dots, d_m$, where each d_i is selected uniformly at random from the set $\{1, \dots, n\}$, where n is unknown. To estimate n we return

$$\tilde{n} = \frac{m(m-1)}{2D} \qquad \text{where} \qquad D = \sum_{i=1}^{n} \sum_{j<i} \mathbb{1}[d_i == d_j].$$

In class I claimed that if $m = O(\sqrt{n}/\epsilon)$, then with high probability,

$$(1-\epsilon)n \le \tilde{n} \le (1+\epsilon)n.$$

However, I never proved that fact. You will fill in the analysis here.

(a) (3pts) Show that $\mathbb{E}[D] = \frac{m(m-1)}{2n}$. We already proved this in class, so I'm asking you to reprove it.

$$E[D] = E\Big[\underbrace{\frac{m^2}{2\tilde{n}}}_{\text{constant}}\Big] - E\Big[\underbrace{\frac{m}{2\tilde{n}}}_{\text{constant}}\Big]$$

$$= \frac{m^2}{2n} - \frac{m}{2n} = \frac{m^2 - m}{2n} = \frac{m(m-1)}{2n}$$

(b) (4pts) Show that $\text{Var}[D] \le \frac{m(m-1)}{2n}$.

$$P\left[\left|D - \frac{m(m-1)}{2n}\right| \ge k\sigma\right] \le \frac{m(m-1)}{2n}$$

$$\le P[|D| \ge k\sigma] \le \frac{m(m-1)}{2n}$$

(c) (5pts) Prove that if $m = c\sqrt{n}/\epsilon$ for a large enough constant c, then

Chernoff Bond

$$\Pr[|D - \mathbb{E}[D]| \ge \epsilon\mathbb{E}[D]] \le 1/100.$$

$P[\text{Var}(D)] \ge$

$$P(|x-\mu| \ge \varepsilon\mu) \le 2e^{-\frac{\varepsilon^2\mu}{2+\varepsilon}}$$

minimize, $\varepsilon = 0$

(d) (3pts) Conclude that if $m = O(\sqrt{n}/\epsilon)$, then with probability 99/100,

$$(1-\epsilon)n \le \tilde{n} \le (1+\epsilon)n.$$

Hint: You may assume that $\epsilon \le 1/2$, and use the fact that in that case $\frac{1}{1-\epsilon} \le 1 + 2\epsilon$ and $\frac{1}{1+\epsilon} \ge 1 - \epsilon$.

4. Regularization (15pts)

Regularization is a popular technique in machine learning. It is often used to improve final test error, but can also help speed-up optimization methods like gradient descent by improving the condition number of the function being regularized.

(a) (5pts) Let $f(\mathbf{x})$ be a differentiable function mapping a length d vector $\mathbf{x}$ to a scalar value. Let g be the function with added Euclidean regularization:

$$g(\mathbf{x}) = f(\mathbf{x}) + \lambda\|\mathbf{x}\|_2^2$$

Above $\lambda < 0$ is a non-negative constant that controls the amount of regularization. Write down an expression for the gradient of g, $\nabla g(\mathbf{x})$, in terms of $\nabla f(\mathbf{x})$, λ, and $\mathbf{x}$.

$$\nabla g(x) = \frac{\partial}{\partial x}\left[f(x) + \lambda\|x\|_2^2\right]$$

$$= f'(x) + \lambda\frac{\partial}{\partial x}\|x\|_2^2$$

(b) (5pts) Prove that $h(\mathbf{x}) = \lambda\|\mathbf{x}\|_2^2$ is an α_1-strongly convex, β_1-smooth function with $\alpha_1 = \beta_1 = 2\lambda$.

For β_1 smoothness 2nd derivative must be bounded:

$h'(x) = 2\lambda\|x\|_2$

$h''(x) = 2\lambda(1)$ constant so it is bounded and β_1 smooth

For strong α_1-convexity, similar logic applies for the first derivative, which also has the outside constant equal to 2λ, and so $\alpha_1 = 2\lambda = \beta_1$

(c) (5pts) Assume that f is α_2-strongly convex and β_2-smooth. Using part (b), prove that the regularized function $g(\mathbf{x})$ has strictly better condition number than $f(\mathbf{x})$.

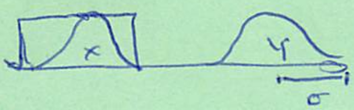
Hint: As a first step, show that when you add an α_1-strongly convex, β_1-smooth function to an α_2-strongly convex, β_2-smooth function, then the resulting function is $(\alpha_1 + \alpha_2)$-strongly-convex and $(\beta_1 + \beta_2)$-smooth.

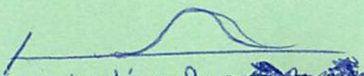
The $g(x)$ condition number is $\frac{2\lambda}{2\lambda} = 1$, so we need to prove $f(x)$'s condition number < 1, if $f(x)$ is α_2 strongly convex and added to the $\alpha_1\ \beta_1$ regularization term, we have an $(\alpha_1 + \alpha_2)$ and $(\beta_1 + \beta_2)$ result,

since $\alpha_1 \neq \alpha_2$ and $\beta_1 \neq \beta_2$ $\frac{\alpha_1 + \alpha_2}{\beta_1 + \beta_2} \neq 1$, thus $g(x)$ has a better condition number: 1.

1b: $\sigma^2 = E[(X-\mu_x)^2] = E[(Y-\mu_y)^2]$

$$Var[D] = E\left[D - \frac{m(m-1)}{2n}\right] \leq \frac{m(m-1)}{2n}$$





pen, boligrafo

$$E[X+Y] = \frac{X+Y}{2}$$

$$Var(X+Y) =$$

$$\int_0^\infty x f(x)\,dx$$

$$\int_0^\infty (x+Y) f(x)\,dx$$

$$\int_0^\infty x f(x) + Y f(x)\,dx$$

$$\tilde{n} = \frac{m(m-1)}{2D}$$

$$D = \frac{m(m-1)}{2\tilde{n}}$$

$$= \frac{m^2}{2\tilde{n}} - \frac{m}{2\tilde{n}}$$

$$E[D] = E\left[\frac{m^2}{2\tilde{n}}\right] - E\left[\frac{m}{2\tilde{n}}\right]$$

$$= \frac{m^2 - m}{2n}$$

$$= m(m-$$

$$E[D] = E\left[\sum_{i=1}^{n} \sum_{j<i} \mathbb{1}[d_i == d_j]\right]$$

$$n \cdot m \cdot \frac{1}{2}$$