CS698C 2021 August Final Exam

Anirudh Nanduri

TOTAL POINTS

275 / 300

QUESTION 1

1 Linear regression **80 / 100**

+ 80 Point adjustment

a) Excellent. b,c) All excellent. A word about why A^TA + C^TC is positive semi-definite would have been relevant. d)Adding ||Dx-f||^2 to the objective function is not correct. You make take the Lagrangian route and differentiate wrt to x and lambda. That will give an alternate solution to my solution.

QUESTION 2

2 Low rank Affine transformation minimization **95 / 100**

+ 95 Point adjustment

a) End of 1st page, startof 2nd, what happened to Frobenius norm square for Pythagoras' theorem to work. Cols of Y are in col space of A--beautiful! well stated. Excellent. b) tr (SC)^T((SD) <= ||SC|| ||SD|| as you have used will give a factor of 2. Instead take it as the sum of inner products and use inner product preservation with error as O(eps) times product of individual norms, and now proceed. You could have directly used affine embedding properties --this route takes you through its proof. But P has rank k, so r =O(k/eps^2+..) suffices, not d/eps^2.. (c) OK. Very nice mathematical writing. You are very good.

QUESTION 3

3 Countsketch random matrices **100 / 100**

+ 100 Point adjustment

Excellent. Very nicely written.

Anirudh
21111011

1. a) Given,

$A, C \in \mathbb{R}^{n\times d}$ and $b, d \in \mathbb{R}^{n}$

$$\min_{x \in \mathbb{R}^d} \|Ax-b\|_2^2 + \|Cx-d\|_2^2 \quad \text{--- (1)}$$

→ Let $S \in \mathbb{R}^{k\times n}$ dimension be a random matrix from Normal distribution $D(0, \frac{1}{k})$ where S_{ij} are iid normal random variables.

→ Dimensionality reduction technique used is

$$\min_{x \in \mathbb{R}^d} \|SAx - Sb\|_2^2 + \|SCx - Sd\|_2^2 \quad \text{--- (2)}$$

→ Let the minimum value be x^* of eqn (1).

→ To preserve minimum value x^*

$$\|Ax^* - b\|_2^2 + \|Cx^* - d\|_2^2 = (1+O(\epsilon))\left[\|SAx^* - Sb\|_2^2 + \|SCx^* - Sd\|_2^2\right] \quad \text{--- (3)}$$

→ Let $\hat{x}$ be the minimum value of eqn (2).

→ To preserve minimum value $\hat{x}$

$$\|SA\hat{x} - Sb\|_2^2 + \|SC\hat{x} - Sd\|_2^2 = (1+O(\epsilon))\left(\|A\hat{x} - b\|_2^2 + \|C\hat{x} - d\|_2^2\right) \quad \text{--- (4)}$$

Using JL lemma to preserve column space A and vector b $[A\ b]$.

$$\|SA\hat{x} - Sb\|_2^2 = (1 \pm O(\epsilon))\|A\hat{x} - b\|_2^2$$

Anirudh

2111011 (2)

~~tea~~

$\Rightarrow$

$$(1-\epsilon)\|A\hat{x}-b\|_2^2 \le \|SA\hat{x}-Sb\|_2^2 \le \|SAx^*-Sb\|_2^2 \le (1+\epsilon)\|Ax^*-b\|_2^2$$

(5)

$$\Rightarrow \|A\hat{x}-b\|^2 = (1+O(\epsilon))\|Ax^*-b\|^2$$

Similarly by preserving column space of C and vector d [C d].

$$\Rightarrow \|C\hat{x}-d\|_2^2 = (1+O(\epsilon))\|Cx^*-b\|^2 \quad -(6)$$

Here x^* is argmin $\|Ax-b\|^2 + \|Cx-d\|^2$.

Adding (5) and (6)

$$\|A\hat{x}-b\|_2^2 + \|C\hat{x}-d\|_2^2 = (1+O(\epsilon))\left[\|Ax^*-b\|^2 + \|Cx^*-d\|^2\right] \quad -(7)$$

For eqn (7) to hold we have to preserve $[A \; b]$ $[C \; d]$.

$\downarrow$ colspace A

Using γ net argument to preserve column space and taking $\gamma = 1/2$

We need to preserve $|M|^2$ vectors for a column space where $|M| = 5^d$.

$$\cancel{\to 1 - (|M|^2 + |M|^2)} \; \text{exp}$$

~~colspace A~~ colspace C.

Anirudh
21111011
3

Using Boole's inequality for eqn 7 and JL lemma above

$$P_r\left[\left(\|A\hat{x}-b\|_2^2 + \|C\hat{x}-d\|_2^2 \right) = (1+O(\epsilon)) \left[\|Ax^*-b\|^2 + \|Cx^*-d\|^2 \right] \right.$$

$$\geq 1 - \left(\underbrace{|M|^2}_{\text{column space of } A} + \underbrace{|M|^2}_{\text{column space of } C} \right) \exp\{-\epsilon^2 k\}$$

$$\geq 1 - (2|M|^2)\exp\{-\epsilon^2 k\}$$

$$= 1 - \exp\{-\epsilon^2 k + 2\log|2M|\}.$$

Taking $\gamma = 1/2$ $\quad m = \left(\dfrac{1+\frac{\gamma}{2}}{\frac{\gamma}{2}}\right)^d = 5^d$

$$= 1 - \exp\{-\epsilon^2 k + 2\log 2 + 2\log m\}$$

$$= 1 - \exp\{-\epsilon^2 k + 2d\log 5 + 2\log 2\}$$

(filtering constants)

$$= 1 - \exp\{O(-\epsilon^2 k) + d\log 5\}$$

Let the minimum value is preserved with probability δ

$$\Rightarrow 1 - \exp\{-O(-\epsilon^2 k) + d\} \leq 1 - \delta$$

$$\Rightarrow -\exp\{-O(-\epsilon^2 k) + d\} \leq -\delta$$

$$\Rightarrow -O(-\epsilon^2 k) + d \geq \log\delta$$

$$-\epsilon^2 k + d \leq \log\frac{1}{\delta}$$

$$k = \frac{O(1)}{\epsilon^2}\left(d + \log\frac{1}{\delta}\right).$$

Anirudh
21111011
4

1b) The problem

$$\min_{x \in \mathbb{R}^d} \|Ax - b\|^2 + \|Cx - d\|^2$$

can be solved using gradient descent. optimization technique

$$\min_{x \in \mathbb{R}^d} (Ax-b)^T (Ax-b) + ((Cx-d))^T ((Cx-d)$$

$$\cancel{x^T A^T A x}$$

$$= (x^T A^T - b^T)(Ax - b) + (x^T C^T - d^T)((Cx-d)$$

$$= x^T A^T A x - x^T A^T b - b^T A x + b^T b + x^T C^T C x - x^T C^T d - d^T C x + d^T d.$$

$$= x^T A^T A x - 2b^T A x + b^T b + x^T C^T C x - 2d^T C x + d^T d.$$

$x^T A^T b = (Ax)^T b$ & $b^T A x$ are same since scalars

Let $A^T A = P$, $C^T C = Q$, $\cancel{b^T A = b^*}$, $\cancel{d^T C = d^*}$, $A^T b = b^*$, $C^T d = d^*$

$$= x^T P x - 2x^T b^* + b^T b + x^T Q x - 2x^T d^* + d^T d \quad \text{— (1)}$$

$$\frac{\partial}{\partial x_i} x^T P x = \sum_{\substack{i=1 \\ i=j}}^{d} P_{ii} x_i^2 + \sum_{i \neq j} P_{ij} x_i x_j$$

$$= 2P_{ii} x_i + 2 \sum_{i \neq j}^{d} P_{ij} x_j$$

$$= 2 P_i' x \quad (i^{th} \text{ row of } P).$$

Anirudh

21111011 (5)

$$\frac{\partial}{\partial x} x^T P x = \begin{bmatrix} \frac{\partial}{\partial x_1} x^T P x \\ \vdots \\ \frac{\partial}{\partial x_d} x^T P x \end{bmatrix}$$

$$= 2 \begin{bmatrix} P_1' x \\ \vdots \\ P_d' x \end{bmatrix} = 2Px \quad \text{— (2)}$$

$b^* = A^T b$

$d^* = C^T d$

Similarly

$$\frac{\partial}{\partial x} x^T Q x = 2Qx \quad \text{— (3)}$$

Consider

$$\frac{\partial}{\partial x_i} x^T b^* = \frac{\partial}{\partial x_i} \sum_{j=1}^{j=d} x_j b_j^*$$

$$= b_i^*$$

$$\nabla 2x^T b^* = 2b^* \quad \text{— (4)}$$

Similarly

$$\frac{\partial}{\partial x_i} x^T d^* = d_i^*$$

$$\nabla 2x^T d^* = 2 \begin{bmatrix} d_1^* \\ \vdots \\ d_n^* \end{bmatrix} = 2d^* \quad \text{— (5)}$$

To find minima equating first derivative to zero.

Using (2), (3), (4), (5)

$$\frac{\partial}{\partial x}\left[x^T P x - \cancel{2A^Tb}\, 2x^T b^* + b^T b - 2x^T d^* + d^T d + x^T Q x \right]$$

$$= 2Px - 2b^* - 2d^* + 2Qx = 0$$

$$\Rightarrow 2Px - 2A^Tb - 2C^Td + 2Qx = 0.$$

$$(P+Q)x = A^Tb + C^Td$$

$$(A^TA + C^TC)\, x^* = A^Tb + C^Td.$$

Anirudh
2111011 (6)

Conditions for existance of minimum

1) Hessian matrix $\nabla(\nabla f(x))$ should be positive semidefinite

$\Rightarrow \nabla[Px - A^Tb + Qx - c^Td]$

$\nabla[P + Q]$

$A^TA + c^Tc$ matrix should be positive semi definite.

2) $A^Tb + c^Td$ should be in column space of matrix $A^TA + c^Tc$ as $(A^TA + c^Tc)x^* = A^Tb + c^Td$.

---

1 c) Algebraic solution to x^*.

$(A^TA + c^Tc)\, x^* = A^Tb + c^Td$.

$x^* = (A^TA + c^Tc)^{-1} (A^Tb + c^Td)$.

---

1 d) Given,

$D \in \mathbb{R}^{d \times e}$

$\min_{x \in \mathbb{R}^d} \|Ax - b\|^2 + \|Cx - d\|^2$

such that $x = \underset{w \in \mathbb{R}^e}{\mathrm{argmin}} \|Dw - f\|^2$

i) It is a constrained optimization problem and gradient descent can be applied to the expression.

Anirudh
2111011

7

$$\min_{x \in \mathbb{R}^d} \|Ax-b\|^2 + \|Cx-d\|^2 + \|Dx-f\|^2$$

$$(Ax-b)^T(Ax-b) + ((x-d)^T((x-d)) + (Dx-f)^T(Dx-f)$$

$$(x^TA^T-b^T)(Ax-b) + (x^TC^T-d^T)((x-d)) + (x^TD^T-f^T)(Dx-f)$$

$$f(x)= x^TA^TAx - 2x^TA^Tb + b^Tb + x^TC^TCx - 2x^TC^Td + d^Td$$
$$+ x^TD^TDx - 2x^TD^Tf + f^Tf$$

$$\frac{\partial}{\partial x} f(x) = 0$$

$$\Rightarrow 2A^TAx - A^Tb + 2C^TCx - C^Td + 2D^TDx - D^Tf = 0$$

$$\Rightarrow (A^TA + C^TC + D^TD)x^* = A^Tb + C^Td + D^Tf.$$

$$x^* = (A^TA + C^TC + D^TD)^{-}(A^Tb + C^Td + D^Tf).$$

d ii)

$$\|Ax^* - b\|^2 = (1+O(\epsilon))\|SAx^* - Sb\|^2$$

$$\|Cx^* - d\|^2 = (1+O(\epsilon))\|SCx^* - Sd\|^2$$

$$\|Dx^* - f\|^2 = (1+O(\epsilon))\|SDx^* - Sf\|^2$$

By preserving subspaces $[A \ b]$ $[C \ d]$, $[D \ f]$

column space of A

$$(\|A\hat{x}-b\|^2 + \|C\hat{x}-d\|^2 + \|D\hat{x}-f\|^2) =$$
$$(1\pm\epsilon)[\|Ax^*-b\|^2 + \|Cx^*-d\|^2 + \|Dx^*-f\|^2].$$

where $\hat{x} = ((SA)^T(SA) + (SC)^T(SC) + (SD)^T(SD))^{-}((SA)^TSb + (SC)^TSd + (SD)^TSf)$.

Aniruddh
21111011
6

Dimension of S matrix

$$k = \frac{O(1)}{\epsilon^2}\left(2d + \log\frac{1}{\delta}\right)$$

If we want to preserve r column spans

$$k = \frac{O(1)}{\epsilon^2}\left(rd + \log\frac{1}{\delta}\right)$$

---

2 a) $\min_{\text{rank}(X) \le k} \| AX - B \|_F$ — (1)

Projection of B onto Column space of A

$= UU^T B$ where $A = U\Sigma V^T$

Hence, B can be written as

$$B = UU^T(B) + (I - UU^T)B$$

Orthogonal projection on column space of A

Substituting in (1)

$$\min_{\text{rank}(X) \le k} \| AX - UU^T B + (I - UU^T) B \|_F$$

$UU^T = AA^-$

$$\min_{\text{rank}(X) \le k} \| AX - AA^- B + (I - AA^-) B \|_F$$

Anirudh
2111011
9

$= \| \underbrace{A(X - A^{-}B)}_{\text{colsp of } A} + \underbrace{(-I + AA^{-})B}_{\text{orthogonal to colsp}(A)} \|_F$

$= \| A(X - A^{-}B) \|_F^2 + \underbrace{\| (I - AA^{-}) B \|_F}$

This term can be ignored as it does not contain X

$= \min_{\text{rank}(X) \le k} \| AX - AA^{-}B \|_F$

Let $AX = Y$

$= \min_{\text{rank}(Y) \le k} \| Y - AA^{-}B \|_F$

Using Eckart Young theorem

$Y = [AA^{-}B]_k$

$AX = [AA^{-}B]_k$

As columns of Y are in colspace of A
projection of Y on A is AA^{-}

$AX^{*} = AA^{-}[AA^{-}B]_k$

Alternatively

$\min_{\text{rank}(Y \le k)} \| Y - AX^{*} \|_F$

$Y \neq AX^{*}$

2 b)

$\forall X \in \mathbb{R}^{n \times p}$

$\| SAA^{-}[AA^{-}B]_k X - SB \|_F \in (1 \pm \epsilon) \| AA^{-}[AA^{-}B]_k - B \|_F$

Taking

$AA^{-}[AA^{-}B]_k = P$

21/11/011

Anirudh

10

$\| SPX - SB \|_F \in (1 \pm \epsilon) \| PX - B \|$ — (1)

Here $X = I$

Using affine embedding ~~the~~ equation (1) can be preserved.

$X^* = P^{-}B$

$\hat{X} = (SP)^{-} SB$

~~for affine emb~~

$\| SPX - SB \|_F$

$= \| SPX - SPX^* + SPX^* - SB \|_F$

$= \| SPX - SPX^* \|_F^2 + \| SPX^* - SB \|_F^2$

$+ 2\, tr(SP(x_j - x_j^*))^T (SPx^* - SB)$

$= \| SPX - SPX^* \|_F^2 + \| SPX^* - SB \|_F^2$

$+ 2 \| SP(X - X^*) \|_F \| SPX - SB \|_F$ | $tr(A^T B) \le \|A\|_F \|B\|_F$

$= 2 \left(\| SPX - SPX^* \|_F^2 + \| SPX^* - SB \|_F^2 \right)$ Using $AM \geq GM$

$= (1 \pm \epsilon) \| PX - PX^* \|_F^2 + (1 \pm \epsilon) \| PX^* - B \|_F^2$

$= (1 \pm \epsilon) \| PX - B \|_F^2$

We need to preserve colspace of P (subset of colspace of A) and colums of B

Anirudh
2111011

(11)

$P = AA^{-}[AA^{-}B]_k$

$$r = O\left(\frac{1}{\epsilon^2}\left(\overset{\text{cols of A}}{n}\log\frac{1}{\delta} + \log\frac{\overset{\text{cols of B}}{r}}{\delta}\right)\right)$$

2c) $\hat{X} = \underset{X}{\arg\min} \| SAA^{-}[AA^{-}B]_k X - SAA^{-}B\|_F$.

$\hat{X} = \underset{X}{\arg\min} \| \underbrace{SAA^{-}[AA^{-}B]_k}_{P} X - SB\|_F$.

$\hat{X} = (SP)^{-}[(SP)(SP)^{-}B]_k$.

$= (SAA^{-}[AA^{-}B]_k)^{-}[(SAA^{-}[AA^{-}B]_k)(SAA^{-}[AA^{-}B]_k)^{-}B]$

as one of the terms has rank$\leq k$

$\Downarrow$

$[(SP)(SP)^{-}B]_k$

$\hat{X}$ has rank $\leq k$.

rank(AB) $\leq \min(\text{rank}(A), \text{rank}(B))$

As proved in 2b part.

$\| SAA^{-}(AA^{-}B]_k X - SB\|_F \in (1\pm\epsilon)\| AA^{-}[AA^{-}B]_k - B\|_F$.

$\| SPX - SB\|_F \in (1\pm\epsilon)\| P - B\|_F$ | $P = AA^{-}[AA^{-}B]_k$

$\Rightarrow \| SPX - SB\|_F \in (1\pm\epsilon)\| AA^{-}[AA^{-}B]_k - B\|_F$. — (1)

Anirudh
2111011

12

$$\min_{\operatorname{rank}(X)\le k} \|AX - B\|_F$$

$$= \|AA^{-}[A\bar{A}B]_k - B\| \quad — (2)$$

from (1) and (2).

$$\|A\hat{X} - B\|_F \le (1+\epsilon) \min_{\operatorname{rank}(X)\le k} \|AX - B\|_F$$

---

2 d) Consider the expression

$$\min_{\operatorname{rank}(X)\le k} \|AX - B\|_F$$

$$X^* = A^{-}[A\bar{A}B]_k \quad (\text{proved in 2a})$$

$$\min_{\operatorname{rank}(X)\le k} \|AX - B\|_F = \|AA^{-}[AA^{-}B]_k - B\|_F \quad — (1)$$

Similarly $\hat{X}^* = (SA)^{-}[SA(SA)^{-}SB]_k$.

$$\min_{\operatorname{rank}(X)\le k} \|SAX - SB\|_F = \|SA(SA)^{-}[SA(SA)^{-}SB]_k - SB\|_F$$

$$\|A\hat{X} - B\|_F = (1\pm\epsilon)\|AX^* - B\|_F \quad [\text{already proved in 2c}]$$

$$\|A(SA)^{-}[SA(SA)^{-}SB]_k - B\|_F \quad — (2)$$

$$= (1\pm\epsilon)\|AX^* - B\|$$

$$\min_{\operatorname{rank}(Y)\le k} \|Y(SA\bar{A}B) - A\bar{A}B\|_F = (1\pm\epsilon)\|A\hat{X} - B\|$$

Anirudh

21111011

2e)

$$\| Y S A A^{-} B R - B R \|_F^2$$

$$= \| Y S A A^{-} B R - B R (S A A^{-} B R)^{-} (S A A^{-} B R) + B R (S A A^{-} B R)^{-} (S A A^{-} B R) - B R \|_F^2$$

Adding and substracte

$$\cancel{= \| (Y S A A^{-} - B R (S A A^{-} B R)^{-} S A A^{-}) B R}$$

$$\cancel{+}$$

$$= \| [Y - B R (S A A^{-} B R)^{-}] (S A A^{-} B R) \|_F^2 \Rightarrow \text{row space of } S A A^{-} B R$$

$$+ \| B R ((S A A^{-} B R)^{-} (S A A^{-} B R) - I) \|_F^2 \Rightarrow \text{Orthogonal to row space of } B R.$$

~~This term can be ignored as it doesnot contain Y.~~

$$= \| Y S A A^{-} B R - B R (S A A^{-} B R)^{-} (S A A^{-} B R) \|_F^2$$

$$+ \| B R - B R (S A A^{-} B R)^{-} (S A A^{-} B R) \|_F^2.$$

Using $\| A - B \|_F^2 = \| B - A \|_F^2$

Anirudh
2111011 (14)

2f)

$$\hat{W} = \underset{rank(W) \le k}{\operatorname{argmin}} \| W - BR(SA\bar{A}BR)^{-}(SA\bar{A}BR) \|_F$$

Using Eckart Young Theorem

$$\hat{W} = [BR(SA\bar{A}BR)^{-}(SA\bar{A}BR)]_K$$

Required to complete

SVD of $BR(SA\bar{A}BR)^{-}(SA\bar{A}BR)$

Let $SA\bar{A}BR = U\Sigma V^T$

$$BR(U\Sigma V^T)^{-} U\Sigma V^T$$

$$BRV\Sigma^{-1}U^TU\Sigma V^T$$

$$[BRVV^T]_k$$

Hence SVD of $[BRVV^T]_k$ needs to be computed.

$$[BRVV^T]_k = U^* \Sigma^*_k V^{T*}$$

Anirudh
21111011
15

2g)

$$\hat{Y} = \underset{\text{rank}(Y) \le k}{\text{argmin}} \| Y SA A^{-} BR - BR (SA A^{-} BR)^{-} (SA A^{-} BR) \|_F$$

~~Let~~ $\cancel{Y SA A^{-} BR} = Z$.

~~and SVD of~~ and

$\cancel{Z - BR}$

Let SVD of $SA A^{-} BR = U \Sigma V^T$

$$= \underset{\text{rank}(Y) \le k}{\text{argmin}} \| Y U \Sigma V^T - BR (SA A^{-} BR)^{-} U \Sigma V^T \|_F$$

$$= \underset{\text{rank}(Y) \le k}{\text{argmin}} \| \underbrace{Y U \Sigma}_{Z} - BR (SA A^{-} BR)^{-} U \Sigma \|_F$$

$$Z^* = [BR (SA A^{-} BR)^{-} U \Sigma]_k$$

$$Y^* = [BR (SA A^{-} BR)^{-} U \Sigma]_k \Sigma^{-} U^T$$

2h)

Anirudh

2111011

16

$\hat{Y}$ is a rank $\leq k$ matrix

using Affine embedding

where $\hat{X}$ is soln of $\underset{\text{Rank } X \leq k}{} \| SAXR - SBR \|_F$

$$\| A\hat{X} - B \|_F \leq (1 \pm O(\epsilon)) \| AX^* - B \|_F \quad -① $$

$$\| A\hat{Y} - B \|_F \leq (1 \pm O\epsilon) \| A\hat{X} - B \|_F . \quad -② $$

Hence from ① and ②.

$$\| A\hat{Y} - B \|_F \leq (1 \pm O\epsilon)) \| AX^* - B \|_F .$$

with probability $1 - O(\delta)$

Anirudh
21111011

3a) Given,

Count sketch random matrix S of dimension $k \times n$.

Let randomized hash function be h

$h : \{1 \dots n\} \rightarrow \{1 \dots k\}$.

Sketch at row (bucket) i for matrix S is

$$[S(x)]_i = \sum_{j : h(j) = i} \epsilon_{ij} x_j$$

As given in the question ϵ_{ij} $i \in \{1 \dots k\}$ $j \in \{1 \dots n\}$ are iid redamacher variables

$\{b_{ij}\}$ are bernoulli variables

for $1 \leq i \leq k$, $1 \leq j \leq n$

$$b_{ij} = \begin{cases} 1 & \text{if } h(j) = i \\ 0 & \text{otherwise} \end{cases}$$

$$\boxed{[S(x)]_i = \sum_{j=1}^{n} b_{ij} \epsilon_{ij} x_j} \quad \text{--- (1)}$$

Sx

$\Downarrow$

$$\begin{bmatrix} b_{i1}\epsilon_{i1} & b_{i2}\epsilon_{i2} & \cdots & b_{in}\epsilon_{in} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$(Sx)_i = b_{i1}\epsilon_{i1}x_1 + b_{i2}\epsilon_{i2}x_2 + \dots + b_{in}\epsilon_{in}x_n$$

$$S_{ij} = b_{ij}\epsilon_{ij} = \begin{cases} \epsilon_{ij} & \text{if } b_{ij} = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$(Sx)_i^2 = \left(\sum_{j=1}^{n} b_{ij} \epsilon_{ij} x_j\right)^2 \quad \text{from eqn } (1)$$

Anirudh
21111011 18

= Sum of pure terms
+ sum of cross terms

Using
$(a_1 + a_2 + \dots a_n)^2$
$= \underbrace{\sum_{i=1}^{n} a_i^2}_{\text{pure terms}} + \underbrace{2\sum_{i,j} a_i a_j}_{\text{cross terms}}$

$$= \sum_{j=1}^{n} (b_{ij} \epsilon_{ij} x_j)^2$$

$$+ \sum_{1 \le j < j' \le n} 2\, b_{ij} b_{ij'} \epsilon_{ij} \epsilon_{ij'} x_j x_{j'}$$

Using $b_{ij}^2 = b_{ij}$ (square of bernoulli variable is bernoulli)

$\epsilon_{ij}^2 = 1$

$$= \sum_{j=1}^{n} b_{ij} x_j^2 + \sum_{1 \le j < j' \le n} 2\, b_{ij} b_{ij'} \epsilon_{ij} \epsilon_{ij'} x_j x_{j'}$$

Anirudh

211110 11

(19)

3 b)

From the resut in previous part

$$(Sx)_i^2 = \sum_{j=1}^{n} b_{ij} x_j^2 + \sum_{1 \le j < j' \le n} 2 b_{ij} b_{ij'} \epsilon_{ij} \epsilon_{ij'} x_j x_{j'}$$

k = No of rows of S

S is $k \times n$ matrix

$$\|Sx\|^2 = \sum_{i=1}^{k} (Sx)_i^2$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{n} b_{ij} x_j^2 + \sum_{i=1}^{k} \sum_{1 \le j < j' \le n} 2 b_{ij} b_{ij'} \epsilon_{ij} \epsilon_{ij'} x_j x_{j'}$$

Using $\sum_{i=1}^{r} \sum_{j=1}^{c} x_{ij}^2 = \sum_{j=1}^{c} \sum_{i=1}^{r} x_{ij}^2$

$$= \sum_{j=1}^{n} x_j^2 \sum_{i=1}^{k} b_{ij} + \sum_{i=1}^{k} \sum_{1 \le j < j' \le n} 2 b_{ij} b_{ij'} \epsilon_{ij} \epsilon_{ij'} x_j x_{j'}$$

x is a unit vector $\sum_{j=1}^{n} x_j^2 = 1$ $\quad \sum_{i=1}^{k} b_{ij} = 1$

$$= 1 + \sum_{i=1}^{k} \sum_{1 \le j < j' \le n} 2 b_{ij} b_{ij'} \epsilon_{ij} \epsilon_{ij'} x_j x_{j'}$$

$$\|Sx\|^2 - 1 = \sum_{i=1}^{k} \sum_{1 \le j < j' \le n} 2 b_{ij} b_{ij'} \epsilon_{ij} \epsilon_{ij'} x_j x_{j'}$$

Anirudh
2111011

20

3 c) From the previous part (3b)

$$\|Sx\|^2 - 1 = \sum_{i=1}^{k} \sum_{1 \le j < j' \le n} 2\, b_{ij} b_{ij'} \epsilon_{ij} \epsilon_{ij'} x_j x_{j'}$$

Taking Expectation on both sides

$$E\left[\|Sx\|^2 - 1\right] = E\left[\sum_{i=1}^{k} \sum_{1 \le j < j' \le n} 2\, b_{ij} b_{ij'} \epsilon_{ij} \epsilon_{ij'} x_j x_{j'}\right].$$

Consider the term

$$E\left[2\, b_{ij} b_{ij'} \epsilon_{ij} \epsilon_{ij'} x_j x_{j'}\right].$$

$$= E\left[2\, b_{ij} b_{ij'}\right] \underset{\substack{\Downarrow \\ 0}}{E[\epsilon_{ij}]}\, \underset{\substack{\Downarrow \\ 0}}{E[\epsilon_{ij'}]}\, x_j x_{j'}$$

independent.

$$= 0$$

Hence by linearity of expectation

$$E\left[\|Sx\|^2 - 1\right] = \sum_{i=1}^{k} \sum_{1 \le j < j' \le n} \underset{\substack{\Downarrow \\ 0}}{E\left[2\, b_{ij} b_{ij'} \epsilon_{ij} \epsilon_{ij'} x_j x_{j'}\right]}.$$

$$E\left[\|Sx\|^2 - 1\right] = 0.$$

Anirudh
2111011 21

4 a)

Given in question

$$\|Sx\|^2 - 1 = \sum_{i}^{K} \sum_{j<j'} T^{i}_{j,j'}$$

where $$T^{i}_{j,j'} = 2\, b_{ij}\, b_{ij'}\, \epsilon_{ij}\, \epsilon_{ij'}\, x_j\, x_{j'}$$

$$\left(\|Sx\|^2 - 1\right)^2 = \left(\sum_{k=1}^{K} \sum_{j<j'} T^{i}_{j,j'}\right)^2$$

$$= \sum_{\substack{i=i' \\ (j,j')\,(l,l')}} T^{i}_{j,j'}\, T^{i'}_{l,l'} \quad \text{(self term)}$$

$$+ \sum_{\substack{i\neq i' \\ (j,j')\,(l,l')}} T^{i}_{j,j'}\, T^{i'}_{l,l'} \quad \text{(cross term)}$$

Consider the cross term $i \neq i'$ in the above equation

$$\sum_{\substack{i\neq i' \\ (j,j'),(l,l')}} T^{i}_{j,j'}\, T^{i'}_{l,l'} = 4 \sum_{\substack{i\neq i' \\ (j,j')\,(l,l')}} \left(b_{ij}\, b_{ij'}\, \epsilon_{ij}\, \epsilon_{ij'}\, x_j\, x_{j'}\right) * \left(b_{i'l}\, b_{i'l'}\, \epsilon_{i'l}\, \epsilon_{i'l'}\, x_l\, x_{l'}\right)$$

Using graph terminology

ϵ_{ij} ———————— $\epsilon_{ij'}$

$\epsilon_{i'l}$ ———————— $\epsilon_{i'l'}$

Given rademacher variables are iid.

Hence if $i \neq i'$ irrespective of values of $(j,j')\ (l,l')$ they will <u>not</u> be paired

Anirudh

21111011

22

$$E\left[T^{i}_{j,j'}\, T^{i'}_{l,l'}\right] = 4 \sum_{i \neq i'} E\left[b_{ij}\, b_{ij'}\, \epsilon_{ij}\, \epsilon_{ij'}\, x_j\, x_{j'}\, * \; b_{i'l}\, \epsilon_{i'l'}\, \epsilon_{i'l}\, \epsilon_{i'l'}\, x_l\, x_{l'}\right].$$

$$= 0$$

~~pairwise~~ independent:

$$E[b_{ij}\, b_{ij'}]\, E[b_{i'l}\, b_{i'l'}]\, E[\epsilon_{ij}\, \epsilon_{ij'}\, \epsilon_{i'l}\, \epsilon_{i'l'}]$$

$$= E[\epsilon_{ij}]\, E[\epsilon_{ij'}]\, E[\epsilon_{i'l}]\, E[\epsilon_{i'l'}]$$

$$= 0$$

iid

redemacher

4b)

$$\left(\|Sx\|^2 - 1\right)^2 = \sum_{i,i',(j,j'),(l,l')} T^{i}_{j,j'}\, T^{i'}_{l,l'}$$

$$= \underbrace{\sum_{\substack{i=i' \\ (j,j')(l,l')}} T^{i}_{j,j'}\, T^{i}_{l,l'}}_{\text{Self terms}} + \underbrace{\sum_{\substack{i\neq i' \\ (j,j')(l,l')}} T^{i}_{j,j'}\, T^{i'}_{l,l'}}_{\text{Cross terms}}$$

$$E\left[\left(\|Sx\|^2 - 1\right)^2\right] = E\left[\sum_{\substack{i=i' \\ (j,j') \\ (l,l')}} T^{i}_{j,j'}\, T^{i'}_{l,l'}\right] + \underbrace{E\left[\sum_{\substack{i\neq i' \\ (j,j')(l,l')}} T^{i}_{j,j'}\, T^{i'}_{l,l'}\right]}_{\Downarrow\; 0}$$

proved in 4a)

$$= E\left[4 \sum_{\substack{i=i' \\ (j,j') \\ (l,l')}} b_{ij}\, b_{ij'}\, \epsilon_{ij}\, \epsilon_{ij'}\, b_{i'l}\, b_{i'l'}\, \epsilon_{i'l}\, \epsilon_{i'l'}\, x_j\, x_{j'}\, x_l\, x_{l'}\right]$$

~~Since $i = i'$~~

$$= 4 \sum_{i=i'(j,j')(l,l')} x_j\, x_{j'}\, x_l\, x_{l'}\, E\left[b_{ij}\, b_{ij'}\, b_{il}\, b_{il'}\, \epsilon_{ij}\, \epsilon_{ij'}\, \epsilon_{il}\, \epsilon_{il'}\right]$$

$$= 4 \sum_{\substack{i=i' \\ (j,j') \\ (l,l')}} x_j x_{j'} x_l x_{l'} \, E[b_{ij} b_{ij'}] \, E[b_{il} b_{il'}] \, E[\epsilon_{ij}\epsilon_{ij'}\epsilon_{il}\epsilon_{il'}]$$

$\uparrow$ self term $i=i'$ Independent

Anirudh
21111011
23

$$= 4 \sum_{\substack{i=i' \\ (j,j') \\ (l,l')}} x_j x_{j'} x_l x_{l'} \, E[b_{ij} b_{ij'}] \, E[b_{il} b_{il'}] \, E[\epsilon_{ij}^2 \epsilon_{ij'}^2]$$

If $i=i'$, $(j', j) = (l', l)$ then expectation is non zero

Since $\epsilon_{ij}^2 = 1$, $b_{ij}^2 = b_{ij}$

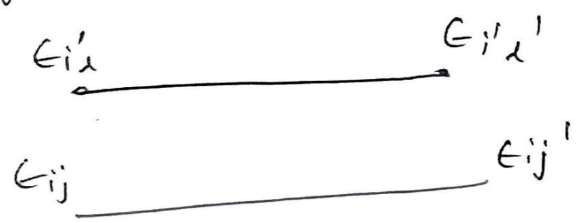
$$= 4 \sum_{\substack{i=i' \\ j,j'}} x_j^2 x_{j'}^2 \; E[b_{ij}] E[b_{ij'}]$$

---

4 c)

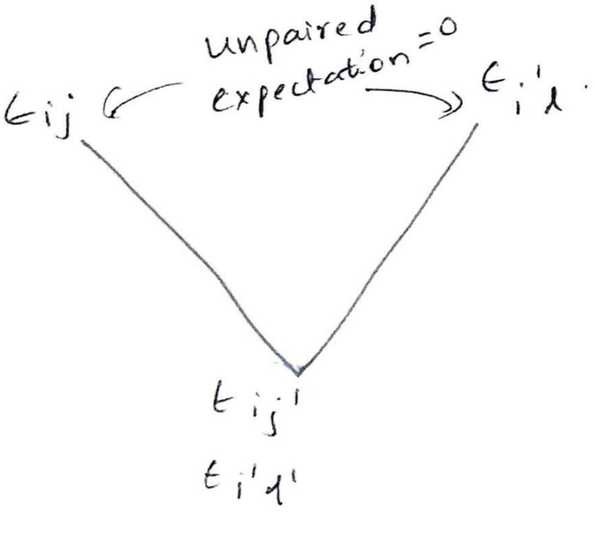
Graph terminology

If $i \neq i'$



unpaired expectation $= 0$.

If $i=i'$ then possible pairing combinations are



$i=i'$

If $j = l'$, $j' \neq l$	$j = l$, $j' \neq l'$	$j \neq l'$, $j' = l$	$j \neq l$, $j' = l'$

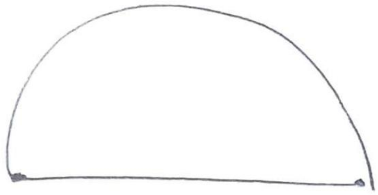
Anirudh
21111011

24

$$= \cancel{4 \sum_{i=i'} x_j x_{j'} x_l x_{l'} E[b_{ij} b_{ij'}]}$$

$\cancel{(j,j')}$
$\cancel{(l,l')}$

$(j, j') = (l, l')$



$\epsilon_{i'l}$ $\epsilon_{i'l'}$
ϵ_{ij} $\epsilon_{ij'}$

Hence

$$E[\epsilon_{i'l}\, \epsilon_{i'l'}\, \epsilon_{ij}\, \epsilon_{ij'}]$$
$$= E[\epsilon_{ij}^2\, \epsilon_{ij'}^2]$$
$$= 1 \quad \text{if paired twice.}$$

If $i \neq i'$

$$E[T^{i}_{j,j'} T^{i'}_{l,l'}] = 0$$

if $i \neq i'$

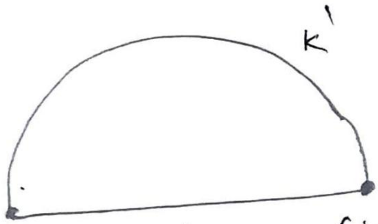
$$E[T^{i}_{j,j'} T^{i'}_{l,l'}] = 0$$

Expectation is non zero

$$= 4 \sum_{j,j'} x_j^2 x_{j'}^2 E[b_{ij}] E[b_{ij'}]$$

---

4 d). As mentioned above if $i = i'$ and $(j, j') = (l, l')$



K'
K
ϵ_{ij} $\epsilon_{ij'}$
j_1 j_2

we need to consider only a pair of edges (j, j') for same graph twice

$i = i'$ $(j, j') = (l, l')$

$j = l$ or $j = l'$
$j' = l'$ $\quad j' = l$.

Anirudh
2111101 (25)

4e)

$$T^{i}_{j,j'} \, T^{i}_{j,j'} = 4 b_{ij}^2 b_{ij'}^2 \epsilon_{ij}^2 \epsilon_{ij'}^2 x_j^2 x_{j'}^2$$

$$= 4 b_{ij} b_{ij'} \cdot 1 \cdot 1 \quad x_j^2 x_{j'}^2$$

$$= 4 b_{ij} b_{ij'} x_j^2 x_{j'}^2$$

$\epsilon^2 = \cancel{1}$
square of redemacher variable is 1

$b_{ij}^2 = b_{ij}$
square of bernoulli variable is bernoulli

p. T. O.

Anirudh
2111011
26

4f) Taking expectation on both sides

$$E[T^i_{j,j'} T^i_{j,j'}] = 4E[b_{ij} b_{ij'} x_j x_{j'}^2]$$

$$= 4 x_j x_{j'} \underbrace{E[b_{ij}] E[b_{ij'}]}_{\text{pairwise independence}}$$

$$= 4 x_j x_{j'}^2 \frac{1}{k} \cdot \frac{1}{k}$$

$$= \frac{4}{k^2} x_j x_{j'}^2$$

$E(b_{ij}) = P[b_{ij} = 1] = \frac{1}{k}$

probability of hashing to a row is $\frac{1}{k}$

$P[h(j) = i] = \frac{1}{k}$

4g) $$E[(\|Sx\|^2 - 1)^2] = E\left[\sum_i \sum_{i'} \sum_{j<j'} \sum_{l<l'} T^i_{jj'} T^{i'}_{ll'}\right]$$

$$= \sum_i \sum_{i'} \sum_{j<j'} \sum_{l<l'} E[T^i_{jj'} T^{i'}_{ll'}]$$

if $i \neq i'$
or $i = i'$ and $(j,j') \neq (l,$
$E[T^i_{jj'} T^{i'}_{ll'}] = 0$

proved in 4b

$$= \sum_i \sum_{j<j'} E[T^{i\,2}_{jj'}]$$

$$= \sum_i \sum_{j<j'} \frac{4}{k^2} x_j^2 x_{j'}^2$$

k rows ins $\Rightarrow$ $$= k \sum_{j<j'} \frac{4}{k^2} x_j^2 x_{j'}^2$$

$$= \frac{4}{k} \sum_{j<j'} x_j^2 x_{j'}^2$$

$$= \frac{4}{K} \sum_{j<j'} x_j^2 \, x_{j'}^2 = \frac{4}{K} \times \frac{1}{2}\left[\left(\sum_j x_j^2\right)^2 - \sum_j x_j^4 \right]$$

Anirudh

2111011

27

$$\leq \frac{4}{K}\left[\left(\sum_j x_j^2\right)^2 - \sum_j x_j^4 \right]$$

$$\leq \frac{4}{K}\left[\left(\sum_j x_j^2\right)^2 \right].$$

$$\leq \frac{4}{K}$$

$\sum_j x_j^2 = 1$

x is a unit vector

1) $(x_1^2 + \dots + x_n^2)^2$

$$= \sum_{i=1}^{n} x_i^4 + 2 \sum_{j=1}^{n} \sum_{\substack{j'=1 \\ j' \neq j}}^{n} x_j^2 x_{j'}^2$$

2) $\Rightarrow \left(\sum_{i=1}^{n} x_i\right)^2 = \sum_{i=1}^{n} x_i^4 + 2\sum_{j<j'} x_j^2 x_{j'}^2$

3. $\Rightarrow x_j^2 x_{j'}^2 = \frac{1}{2}\left[\left(\sum_{i=1}^{n} x_i^2\right)^2 - \sum_{i=1}^{n} x_i^4\right]$

4 $\cancel{\Rightarrow x_j^2 x_{j'}^2 \geq \left(\sum_{i=1}^{n} x_i^2\right)^2 - \sum_j x}$

---

4 h) from the previous parts we have

$$E\left[\left(\|Sx\|^2 - 1 \right)^2 \right] \leq \frac{4}{K} \quad \text{——— (4)}$$

Consider the expression

$$P\left[\left| \|Sx\|^2 - 1 \right| > \epsilon \right]$$

~~Using Markov~~

Anirudh
2111011 28

$$= P\left[\left(\|Sx\|^2 - 1\right)^2 > \epsilon^2\right]$$

Using Markov inequality $P[x \geq a] \leq \frac{E[x]}{a}$

$$P\left[\left(\|Sx\|^2 - 1\right)^2 > \epsilon^2\right] \leq \frac{E\left[\left(\|Sx\|^2 - 1\right)^2\right]}{\epsilon^2}$$

$$P\left[\left|\|Sx\|^2 - 1\right| > \epsilon\right] \leq \frac{4}{K\epsilon^2}$$

Using eqn 4
$E\left[\left(\|Sx\|^2 - 1\right)^2\right] \leq \frac{4}{K}$

Assuming

4 i)

$$P\left[\left|\|Sx\|^2 - 1\right| > \epsilon\right] \leq \frac{4}{K\epsilon^2} \quad \text{— (5)}$$

$$K \geq \frac{4}{\epsilon^2 \delta} \quad \text{(given in question)}$$

$$\delta \geq \frac{4}{K\epsilon^2}$$

$$\Rightarrow \frac{4}{K\epsilon^2} \leq \delta \quad \text{(Plugging in eqn 5)}$$

$$P\left[\left|\|Sx\|^2 - 1\right| > \epsilon\right] \leq \delta$$