Υπογραμμικοί Αλγόριθμοι
Μάθημα 6 - 5/11/2019

Επιμέλεια Διαφανειών: Ερμής Σουμαλιάς

Ελάττωση Διάστασης

- Έχουμε ένα αντικείμενο $\mathcal{A}$
- Θέλουμε να βρούμε έναν τρόπο να σκιαγραφήσουμε το $\mathcal{A}$ και να πάρουμε ένα αντικείμενο σκ($\mathcal{A}$) έτσι ώστε το σκ($\mathcal{A}$) να έχει πολύ μικρότερο μέγεθος από το $\mathcal{A}$, και να μας επιτρέπει να λύσουμε το πρόβλημα που μας ενδιαφέρει.

Εφαρμογές

1. Αλγόριθμοι ροής ($\mathcal{A} \in \mathbb{R}^n$)
2. Κατανεμημένα Συστήματα
3. Προβλήματα Τεχνητής Μάθησης (Γραμμική Παλινδρόμηση)
4. Πολλαπλασιασμός Πινάκων
5. Βάσεις Δεδομένων

Λήμμα Johnson–Lindenstrauss

Έστω διανύσματα $x_1, x_2, x_3, .., x_N \in \mathbb{R}^n$.
Υπάρχει (γραμμικός) μετασχηματισμός $\Pi : \mathbb{R}^n \to \mathbb{R}^m, m = O(\epsilon^{-2} \log N)$, ώστε τα $y_i = \Pi x_i$ να ικανοποιούν $||y_i - y_j||_2 = (1 \pm \epsilon)||x_i - x_j||_2 \quad \forall i, j$[1].

Το συγκεκριμένο λήμμα έχει νόημα όταν $m << n$. Το m είναι το πλήθος των διαστάσεων του χώρου στον οποίο ζουν τα y_i. Με άλλα λόγια, αν $\mathcal{A} = \{x_1, \ldots, x_N\}$, το m είναι το πλήθος των διαστάσεων του σκ($\mathcal{A}$).

[1] Χρησιμοποιώντας το συμβολισμό $a = (1 \pm \epsilon)b$ εννοούμε ότι $a \in [(1 - \epsilon)b, (1 + \epsilon)b]$

Σχόλιο 1: Για παράδειγμα, αν κάποιος θέλει να λύσει ένα πρόβλημα η απάντηση του οποίου ε-ξαρτάται μόνο από τις σχετικές αποστάσεις κάποιων σημείων (πχ συσταδοποίηση/ clustering), μπορεί πρώτα να χρησιμοποιήσει το Johnson-Lindenstrauss για να κατέβει σε χαμηλότερη διάστα-ση, και να λύσει το πρόβλημά του εκεί, με την ελπίδα (ή την εγγύηση) ότι η λύση θα διατηρηθεί κι αυτή μερικώς. Για παράδειγμα, ένας εκθετικός αλγόριθμος στο πλήθος των διαστάσεων μπο-ρεί να παίρνει πλέον πολυωνυμικό χρόνο (στο πλήθος των σημείων)! Ελπίζουμε να δούμε μια τέτοια εφαρμογή αργότερα. Φυσικά, το παραπάνω υποθέτει μια αλγοριθμική εκδοχή του Λήμματος Johnson-Lindenstrauss η οποία υπολογίζει επαρκώς γρήγορα τα y_i.

Σχόλιο 2: Στο παραπάνω πήραμε το Π να είναι γραμμικός μετασχηματισμός. Αυτό δεν είναι απα-ραίτητο, αν και οι καλύτερες κατασκευές που ξέρουμε είναι γραμμικοί μετασχηματισμοί. Επίσης, οι γραμμικοί μετασχηματισμοί είναι επιθυμητοί για διάφορους λόγοςυ, ένας εκ των οποίων είναι γιατί βοηθούν στο να κατασκευάζουμε αλγόριθμους ροής με 'εύκολη/άμεση' ρουτίνα ανανέωσης (update).

Κάτω Φράγμα:

1. [Alon 2001]: Υπάρχει ένα σύνολο σημείων για το οποίο είναι απραίτητες $m = \Omega(\epsilon^{-2}\log N/\log(1/\epsilon))$ διαστάσεις.

2. [Larson, Nelson 2016]: Υπάρχει ένα σύνολο σημείων για το οποίο είναι απαραίτητες $m = \Omega(\epsilon^{-2}\log N)$ διαστάσεις, εφόσον η απεικόνιση είναι γραμμικός μετασχηματισμός.

3. [Larson, Nelson 2017]: Υπάρχει ένα σύνολο σημείων για το οποίο είναι απαραίτητες $m = \Omega(\epsilon^{-2}\log N)$ ακόμα και αν επιτρέπονται μη γραμμικοί μετασχηματισμοί.

Μη συνοχικοί πίνακες (Υπενθύμιση)

Ένας πίνακας $A \in R^{m\times n}$ λέγεται ε-(μη συνοχικός) αν $||A_i||_2 = 1$ και $|\langle A_i, A_j\rangle| \leq \varepsilon$

Γεωμετρική ερμηνεία

- $|\langle A_i, A_j\rangle| \leq \varepsilon$
- Δηλαδή οι στήμες του πίνακα είναι σχεδόν κάθετες ανά δύο, και με ℓ_2 νόρμα 1.
- Είχαμε: $m = \Theta(\epsilon^{-2}\log N) \implies n = 2^{\epsilon(\Theta^2 m)}$
- Για ε = 0 → Μπορώ να βρω $m = n$ κάθετα διανύσματα.
- Όταν αυξάνω το ε → Το πλήθος των διανύσματων που μπορώ να βρω αυξάνεται εκθετικά.
- Μία από τις κατασκευές που δείχνουν ότι τα πράγματα στις υψηλές διαστάσεις είναι πολύ διαφορετικά από αυτά που συμβαίνει στις μικρότερες.

Μια πρώτη Κατασκευή

$$\Pi = \frac{1}{\sqrt{m}} \begin{bmatrix} g_{1,1} & \cdots & \cdots & \cdots & g_{1.n} \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ g_{m,1} & \cdots & \cdots & \cdots & g_{m,n} \end{bmatrix}$$

- Το $\frac{1}{\sqrt{m}}$ αποτελεί όρο κανονικοποίησης.
- $g_{i,j} \sim \mathcal{N}(0,1) \quad \forall i,j \in [m] \times [n]$

Λήμμα Κατανομής Johnson–Lindenstrauss

Υπάρχει μία κατανομή $D_{\epsilon,\delta}$ πάνω σε πίνακες $\Pi \in \mathbb{R}^{n\times m}, m = \Theta(\epsilon^{-2}\log(\frac{1}{\delta}))$ έτσι ώστε

$$\forall x \in \mathbb{R}^n, \mathbb{P}\left[||\Pi x||_2^2 \neq (1 \pm \epsilon)||x||_2^2\right] < \delta.$$

Κάποιες παρατηρήσεις. Παρατηρούμε ότι $||\Pi x||_2 \sim ||x||_2 \implies ||\Pi cx||_2 = ||c(\Pi x)||_2 \sim c|||x||_2$, όπου με $\sim$ συμβολίζουμε ότι τα δύο μεγέθη είναι $(1 \pm \epsilon)$ του άλλου. Αυτό σημαίνει ότι μπορούμε να κανονικοποιήσουμε το διάνυσμά μας έτσι ώστε να έχει ℓ_2 νόρμα $= 1$.Έτσι αρκεί

$$\forall x \in \mathbb{R}^n \text{ with } \|x\|_2 = 1, \mathbb{P}\left[||\Pi x||_2^2 \notin (1 \pm \epsilon)\right] < \delta.$$

Το Λήμμα κατανομής Johnson–Lindenstrauss μπορεί απευθείας να αποδείξει το Λήμμα Johnson–Lindenstrauss για τον εξής λόγο. Θέλουμε να ισχύει η ιδιότητα της κατανομής για N(N-1)/2 διανύσματα (όλες τις διαφορές $x_i - x_j$). Από το φράγμα ένωσης έχουμε:

$$\mathbb{P}\left[\exists i,j \in [N], i \neq j : ||\Pi(x_i - x_j)||_2^2 \notin (1 \pm \epsilon)\|x_i - x_j\|_2\right] \leq \frac{N(N-1)}{2}\delta$$

Άρα για $\delta \leq \frac{4}{N(N-1)}$ παίρνουμε ότι η πιθανότητα να μην υπάρχει κατάλληλος πίνακας Π είναι μικρότερη από 1, αναγκαστικά πρέπει να υπάρχει (πιθανοτική μέθοδος), και ·έξει $m = \Theta(\epsilon^{-2}\log(\frac{1}{\delta})) = \Theta(\epsilon^{-2}\log(N^2)) = \Theta(\epsilon^{-2}\log(N))$ γραμμές. Στην πραγματικότητα, μπορούμε να βάλουμε $\delta = 1/N^c$, για οποιαδήποτε σταθερά c και να πάρουμε πιθανότητα αποτυχίας $1/N^{c-2}$. Αυτό θα δώσει μια πολλαπλαστιαστική σταθερά c στο πλήθος των γραμμών, επειδή το δ είναι μέσα στο λογάριθμο.

Από τα παραπάνω προκύπτει ότι αρκεί να αποδείξουμε το Λήμμα Κατανομής.

Απόδειξη του Λήμματος Κατανομής

Έστω $g_{i,j} \sim \mathcal{N}(0,1)$ για $i \in [m], j \in [n]$.

$$\Pi x = \frac{1}{\sqrt{m}} \begin{bmatrix} g_{1,1} & \dots & \dots & \dots & g_{1.n} \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ g_{m,1} & \dots & \dots & \dots & g_{m,n} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} g'_1 \\ g'_2 \\ \vdots \\ g'_m \end{bmatrix}$$

Πρώτα απ' όλα παρατηρούμε ότι κάθε $(\Phi x)_i$ είναι ένας γραμμικός συνδυασμός από τα x_i, και άρα λόγω της 2-στασιμότητας της κανονικής κατανομής έχουμε ότι $(\Phi x)_i \sim \mathcal{N}(0,1)$. Οπότε, $\|\Phi x\|_2^2 = \sum_{i=1}^{n} (g'_i)^2$, όπου $g'_i \sim \mathcal{N}(0,1)$. Και η κατανομή είναι ανεξάρτητη από το x.

Θα δείξουμε μόνο τις μίας από τις δύο ανισότητες του Λήμματος Κατανομής:

$$\begin{aligned} \mathbb{P}\left[\sum_{i=1}^{m} (g'_i)^2 \geq \lambda\right] = \mathbb{P}\left[s\sum_{i=1}^{m} (g'_i)^2 \geq s\lambda\right] &= \\ \mathbb{P}\left[e^{s\sum_{i=1}^{m} (g'_i)^2} \geq e^{s\lambda}\right] &\underset{Markov}{\leq} \\ \leq e^{-\lambda s}\mathbb{E}\left[e^{s\sum_{i=1}^{m} (g'_i)^2}\right] &= \\ e^{-\lambda s}\prod_{i=1}^{m}\mathbb{E}\left[e^{s(g'_i)^2}\right] &= \\ e^{-\lambda s}\left(\frac{1}{\sqrt{1-2s}}\right)^m &\underset{desirably}{<} \delta \end{aligned}$$

Μετά από τις πράξεις προκύπτει το ζητούμενο για $m = \Theta(\epsilon^{-2}\log(1/\delta))$ και βελτιστοποιώντας ως προς s.

Πρόβλημα της από πάνω κατασκευής: Ένα τυχαίο στοιχείο από μία Γκαουσιανή κατανομή χρειάζεται άπειρα δυφία (με πιθανότητα 1). Μπορείς κάποιος να πετάξει τα περισσότερα δυφία κάθε τυχαίας γκαουσιανής μεταβλητής και να δείξει ότι κρατώντας $\Theta(\log(\max_i\|x_i\|_2))$ ψηφία, το λάθος είναι πολύ μικρό, αλλά υπάρχει μια πολύ πιο φτηνή λύση, κρατώντας μόνο ένα διφύο.

Σκεπτικό Λύσης: Μπορώ αντί για $\sim \mathcal{N}(0,1)$ να βάλω στον πίνακα ± 1. Είναι η ίδια κατασκευή ουσιαστικά με το AMS σκιαγράφημα, αλλάζει όμως η ανάλυση και το πως ρυθμίζω τις παραμέτρους. Με AMS θα έβγαζα $m = \Theta(\epsilon^{-2}\frac{1}{\delta})$ αντί για $m = \Theta(\epsilon^{-2}\log(\frac{1}{\delta}))$.

$$\Pi x = \frac{1}{\sqrt{m}} \begin{bmatrix} \pm 1 & \pm 1 & \dots & \pm 1 & \pm 1 \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ \pm 1 & \pm 1 & \dots & \pm 1 & \pm 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Πριν δούμε την ακριβή ανάλυση, θα δούμε μια κλασική ανισότητα, και θα δούμε τη δύναμη των τυχαίων προσήμων: για το άνω φράγμα είναι εξίσου καλά (και καλύτερα στην πραγματικότητα) από τις γκαουσιανές.

Ανισότητα Khintchine

$$\underset{|x|=1}{\forall x}\ \mathbb{P}_{\sigma_i\in\{\pm1\}}\left[|\sum\sigma_i x_i|\geq\lambda\right]\lesssim e^{-\lambda^2},$$

όπου με $\lesssim$ κρύβω σταθερές τους εκθέτες του δεξιού μέλους, δηλαδή κανονικά θα έγραφα $\leq e^{-c\lambda^2}$, για κάποια σταθερά c.

Απόδειξη

Αρκεί να δείξω ότι[2] ότι

$$\left\|\sum\sigma_i x_i\right\|_p\leq||g||_p\leq c\sqrt{p}$$

,

όπου $g\sim\mathcal{N}(0,1)$ και $||Z||_p=\mathbb{E}[Z^p]^{1/p}$ για μια τυχαία μεταβλητή Z.

Έχουμε ότι

$$\mathbb{E}[g^p]=\begin{cases}0, & \text{p περιττός}\\ \leq p(p-1)(p-3)(p-5)\dots1<p^{p/2}=(\sqrt{p})^p, & \text{p αρτιος}\end{cases}$$

Έτσι η $\|g\|_p\leq c\sqrt{p}$ προκύπτει άμεσα.

Τώρα έχουμε ότι

$$\begin{aligned}||\sum\sigma_i x_i||_p^p=\\ \mathbb{E}[\Big(\sum\sigma_i x_i\Big)^p]=\\ \mathbb{E}[\sum_{i_1,i_2,\dots,i_p}\sigma_{i_1}\dots\sigma_{i_p}x_{i_1}\dots x_{i_p}]=\\ \sum_{i_1,i_2,\dots,i_p}\mathbb{E}[\sigma_{i_1}\dots\sigma_{i_p}]x_{i_1}\dots x_{i_p}\end{aligned}$$

Αν στον όρο $\sigma_{i_1}\dots\sigma_{i_p}x_{i_1}\dots x_{i_p}$ υπάρχει ένας δείκτης i που εμφανίζεται περιττό αριθμό φορών, έχουμε $\mathbb{E}[\sigma_{i_1}\dots\sigma_{i_p}]=0$. Επομένως επιβιώνουν μόνο οι όροι στους οποίους κάθε i εμφανίζεται άρτιο αριθμό φορών. Το γινόμενο $\mathbb{E}[\sigma_{i_1}\dots\sigma_{i_p}]$ σε εκείνη την περίπτωση είναι $1\leq\mathbb{E}[g_{i_1}\dots g_{i_p}]$ [3], όπου $g_i\sim\mathcal{N}(0,1)$ ανεξάρτητες. Και επειδή κάθε i εμφανίζεται άρτιο αριθμό φορών, το $x_{i_1}\dots x_{i_p}$ είναι πάντα μη αρνητικό (όλοι οι εκθέτες κάθε x_i είναι άρτιοι, και κατά συνέπεια

$$\mathbb{E}[\sigma_{i_1}\dots\sigma_{i_p}]x_1\dots x_p=\mathbb{E}[g_{i_1}\dots g_{i_p}]x_{i_1}\dots x_p.$$

Αυτό δείχνει ότι μπορούμε να φράξουμε από πάνω την p-οστή στιγμή του $\sum\sigma_i x_i$ με την p-οστή στιγμή της τυχαίας μεταβλητής όπου κάθε σ_i έχει αντικατασταθεί με g_i. Άρα προκύπτει ότι

[2]Μετά συνεχίζω όπως στην πρώτη άσκηση της πρώτης σειράς ασκήσεων.

[3]Διότι κάθε άρτια στιγμή μιας γκαουσιανής είναι τουλάχιστον 1.

$\|\sum \sigma_i x_i\|_p^p \leq \|g_i x_i\|_p^p = \|g\|_p^p$, από 2-στασιμότητα της κανονικής κατανομής και το γεγονός ότι το x είναι μοναδιαίο διάνυσμα.

Μετά από αυτό το ευχάριστο διάλλειμα, θα προχωρήσουμε τώρα να αποδείξουμε το Λήμμα κατανομής Johnson-Lindenstrauss. Βασικό εργαλείο θα είναι η ανισότητα των Hanson-Wright, η οποία χρησιμοποιείται για να δώσει φράγματα συγκέντρωσης σε τετραγωνικές μορφές τυχαίων προσήμων.

Ανισότητα Hanson-Wright

Έστω ένας συμμετρικός πίνακας $A \in \mathbb{R}^{n\times n}$, και έστω $\|A\|_F^2 = \sum_{i,j} A_{i,j}^2$ και $\|A\|_{op}$ η μέγιστη ιδιοτιμή του A. Έστω $\sigma^T = [\sigma_1, \sigma_2, \ldots, \sigma_n]$. Τότε

$$\mathbb{P}\left[|\sigma^T A\sigma - \mathbb{E}\left\{\sigma^T A\sigma\right\}| \geq \lambda\right] \lesssim e^{-\lambda^2/\|A\|_F^2} + e^{-\lambda/\|A\|_{op}}.$$

Σχόλια:

1. Το $\sigma^T A\sigma$ λέγεται τετραγωνική μορφή επειδή ισούται με $\sum_{i,j}\sigma_i\sigma_j \cdot A_{ij}$.
2. Η ποσότητα $\mathbb{E}\left\{\sigma^T A\sigma\right\}$ ισούται με το ίχνος (trace) του A.
3. Αναλύσαμε ακριβώς μια τετραγωνική μορφή (όπως θα δούμε παρακάτω) όταν κοιτούσαμε το AMS σκιαγράφημα, αλλά με διαφορετικά εργαλεία και με μόνο 4-ανεξάρτητα πρόσημα.
4. Η ανισότητα αυτή είναι μοιάζει με την ανισότητα του Khintchine (βάλε τα μη διαγώνια στοιχεία ίσα με 0 και τα διαγώνια ίσα με τις τιμές του διανύσματος x στο οποίο θες να εφαρμόσεις Khincthine), αλλά δεν υποννοεί την ανισότητα του Khintchine, γιατί έχει και μια εξάρτηση στη μέγιστη ιδιοτιμή του πίνακα, που στη συγκεριμένη περίπτωση μεταφράζεται σε εξάρτηση στο $\max_i |x_i|$. Ο λόγος που συμβαίνει αυτό είναι επειδή ζητάει κάτι πιο δυνατό από την ανισότητα του Khintchine, το κάτω φράγμα, το οποίο δε γίνεται να ισχύει λόγω του παρακάτω παραδείγματος: $x_1 = x_2 = \frac{1}{\sqrt{2}}$ και όλα τα άλλα $x_i = 0$.
5. Από το θεώρημα Courant-Fischer έχουμε ότι $\|A\|_{op} = \max_{\|x\|_2=1} x^T Ax$. Αν το γράψει κανείς ως $\max_{x\neq\vec{0}} \frac{x^T Ax}{x^T x} = \max_{x\neq\vec{0}} \frac{x^T Ax}{x^T Ix}$ είναι το πόσο μπορεί να μεγαλώσει μια τετραγωνική μορφή αν χρησιμοποιήσομε τον A ως τον πίνακα της σε σύγκριση με όταν χρησιμοποιήσουμε τον μοναδιαίο.

Προχωράμε να δείξουμε τώρα το Λήμμα Κατανομής Johnson-Lindenstrauss με την κατασκευή με τα τυχαία πρόσημα. Υπενθυμίζουμε ότι $\Pi_{i,j}$ είναι τυχαίο πρόσημο. Για $y = \Pi x$ έχουμε ότι

$$y_i = \frac{1}{\sqrt{m}} \sum_{j\in[n]} \Pi_{i,j} x_j, i \in [m],$$

και άρα
$$y_i^2 = \frac{1}{m}\left(\sum_{j\in[n]} \Pi_{i,j} x_j\right)^2 = \frac{1}{m}\sum_{j,j'\in[n]} \Pi_{i,j}\Pi_{i,j'} x_j x_j' = \Pi_i^T(\frac{1}{m}xx^T)\Pi_i,$$
όπου Π_i είναι η i-οστή γραμμή του Π. Προσέξτε ότι xx^T είναι ένας $n\times n$ πίνακας με το (ℓ,ℓ') στοιχείου του ίσο με $x_\ell \cdot x_{\ell'}$. Συχνά ονομάζεται και εξωτερικό γινόμενο. Άρα έχουμε ότι το y_i^2 είναι μια τετραγωνική μορφή με τυχαία πρόσημα (αυτά της γραμμής Π_i). Ας φτιάξουμε ένα διάνυσμα $\sigma\in\mathbb{R}^{nm}$, το οποίο είναι η κάθετη συνένωση των γραμμών $\Pi_1,\Pi_2,\ldots,\Pi_m$.

$$\sigma = \begin{bmatrix} \Pi_1 \\ \Pi_2 \\ \ldots \\ \Pi_{m-1} \\ \Pi_m \end{bmatrix} = \begin{bmatrix} \sigma_{1,1} \\ \sigma_{1,2} \\ \sigma_{1,3} \\ \ldots \\ \sigma_{2,1} \\ \sigma_{2,2} \\ \sigma_{2,3} \\ \ldots \\ \sigma_{m,n-2} \\ \sigma_{m,n-1} \\ \sigma_{m,n} \end{bmatrix} \in \mathbb{R}^{mn}$$

Τότε

$$y_1^2 + y_2^2 + \ldots + y_m^2 = \sum \Pi_i^T\left(\frac{1}{m}xx^T\right)\Pi_i = \sigma^T A\sigma,$$
όπου A ένας $(mn)\times(mn)$ πίνακας ο οποίος είναι χωρισμένος σε $n\times n$ μπλοκ (m^2 το πλήθος), με όλα τα διαγώνια μπλοκ να είναι ο πίνακας $\frac{1}{m}xx^T$ και όλα τα μη διαγώνια μπλοκ να είναι ο μηδενικός πίνακας (ας τον πούμε $0_{n\times n}$).

$$A = \begin{bmatrix} \frac{1}{m}xx^T & 0_{n\times n} & \ldots & 0_{n\times n} & 0_{n\times n} \\ 0_{n\times n} & \frac{1}{m}xx^T & \ldots & 0_{n\times n} & 0_{n\times n} \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ 0_{n\times n} & 0_{n\times n} & \ldots & 0_{n\times n} & \frac{1}{m}xx^T \end{bmatrix}$$
Είμαστε ακριβώς στην κατάσταση να εφαρμόσουμε την ανισότητα Hanson-Wright. Υπολογίζουμε τα $\|A\|_F, \|A\|_{op}$ ως εξής.

- $\|A\|_F^2 = m\cdot \|\frac{1}{m}xx^T\|_F^2 = m\cdot\frac{1}{m^2}\sum_{i,i'} x_i^2 x_{i'}^2 = \frac{1}{m}\left(\sum_i x_i^2\right)^2 = \frac{1}{m}.$

- Επειδή ο A είναι μπλοκ-διαγώνιος η μέγιστη ιδιοτιμή του είναι η μέγιστη ιδιοτιμή από τα μπλοκ του. Επειδή όλα είναι ίδια, έχουμε από Courant-Fischer

$$\|A\|_{op} = \max_{\|z\|_2=1} z^T \left(\frac{1}{m} x x^T\right) z = \frac{1}{m}\max_{\|z\|_2=1}(x^T z)\cdot(z^T x) = \frac{1}{m}\max_{\|z\|_2=1}\langle x, z\rangle\langle z, x\rangle = \frac{1}{m}\max_{\|z\|_2=1}\langle x, z\rangle^2 = \frac{1}{m},$$

  διότι για να μεγιστοποιείται το εσωτερικό γινόμενο δύο διανυσμάτων με σταθερό μέτρο πρέπει τα διανύσματα να έχουν την ίδια κατεύθυνση.

Με βάση τα παραπάνω, το δεξί μέλος της Hanson-Wright θα δώσει $e^{-\lambda^2/\frac{1}{m}} + e^{-\lambda/(1/m)} = e^{-m\lambda^2} + e^{-m\lambda}$. Εφόσον $\mathbb{E}\left\{\sigma^T A \sigma\right\} = \|x\|_2^2 = 1$ θέλουμε $\lambda = \epsilon$ και κατά συνέπεια για το Λήμμα Κατανομής πρέπει και αρκεί $e^{-m\epsilon^2} = O(\delta)$, από όπου παίρνουμε ότι $m = O(\epsilon^{-2}\log(1/\delta))$ αρκεί.

Γρήγορος Μετασχηματισμός Johnson-LindenStrauss

Έως τώρα έχουμε δείξει την πιθανοτική κατασκευή ενός πίνακας για το Λήμμα Κατανομής Johnson-Lindenstrauss με τις παρακάτω προδιαγραφές.

- $m = \Theta(\epsilon^{-2}\log(1/\delta))$.
- Χρόνος υπολογισμού Πx: $\mathcal{O}(nm) = \mathcal{O}(n\epsilon^{-2}\log(1/\delta))$.

Τώρα θα δείξουμε μια κατασκευή που ανταλλάσει χρόνο με πλήθος γραμμών.

- $m = O(\epsilon^{-2}\log^2(n/\delta))$.
- Χρόνος υπολογισμού Πx: $O(n\log n)$.

3 βήματα προς την επιτυχία:

- Αναποδογύρισε το πρόσημο κάθε x_i στην τύχη και πάρε το x' $\in \mathbb{R}^n$. $\rightarrow O(\text{n})$ χρόνος
- Επέλεξε $m = \Theta(\epsilon^{-2}\log^2(\delta/n))$ τυχαίους (με επανάληψη) συντελεστές Fourier του x'. $\rightarrow$ $\mathcal{O}(n\log n)$ χρόνος
- Κανονικοποίησε με $\sqrt{n/m}$, παίρνοντας το y.$\rightarrow \mathcal{O}(m)$ χρόνος

Ο Γρήγορος Μετασχηματισμός JL είναι Γραμμικός Μετασχηματισμός

Μπορούμε να γράψουμε

$$y = \sqrt{\frac{n}{m}} PFD$$

Όπου:

- **D:** Διαγώνιος πίνακας με τυχαία πρόσημα.
- **F:** Ο πίνακας του διακριτού μετασχηματισμού Φουριέ.
- **P:** Ένας πίνακας που έχει μόνο έναν άσσο σε κάθε γραμμμή τυχαία (πίνακας δειγματοληψίας).

<u>Σημαντικές Ιδιότητες του Πίνακα F:</u>

Όλα τα στοιχεία μικρά σε μέτρο/απόλυτη τιμή, της τάξης του $\frac{1}{\sqrt{n}}$

Ορθοκανονικός, δηλαδή $\|Fx\|_2 = \|x\|_2$ για κάθε x.

Πολλαπλασιασμός πίνακα με διάνυσμα σε $O(n \log n)$ χρόνο

Κάθε τέτοιος πίνακας επαρκεί.

Πρώτα δείχνουμε ότι ο FD 'επιπεδοποιεί' το x με καλή πιθανότητα. Δείχνουμε το επόμενο Λήμμα. Για μια απόλυτη σταθερά c ισχύει ότι

$$\mathbb{P}\left[\exists j \in [n] : |(FDx)_j| \geq \sqrt{\frac{c\log(n/\delta)}{n}}\right] \leq \frac{\delta}{100}$$

<u>Απόδειξη</u> Θα εφαρμόσουμε την ανισότητα του Khintchine. Στην πραγματικότητα η ανισότητα αυτή εφαρμόζεται μόνο για διανύσματα με πραγματικούς αριθμούς, αλλά εδώ θα το κάνουμε σαν να ισχύει το ίδιο και για μιγαδικά διανύσματα. Σπάζοντας σε πραγματικό και μιγαδικό μέρος και παίρνοντας τριγωνική ανισότητα, κανείς μπορεί να ανακατασκευάσει και για μιγαδικά διανύσματα το αποτέλεσμα.

$$\mathbb{P}[|\sum_{j\in[n]} F_{i,j}\sigma_j x_j| \geq \lambda] \underset{Khintchine}{\lesssim} e^{-\lambda^2 n^2} \leq \frac{\delta}{100n}$$

Άρα

$$\mathbb{P}\left[\exists j \in [n] : |(FDx)_j| \geq \sqrt{\frac{c\log(n/\delta)}{n}}\right] \leq n \cdot \frac{\delta}{100n} = \frac{\delta}{100},$$

από φράγμα ένωσης. Τέλος απόδειξης.

Επειδή $\|FDx\|_2 = \|Dx\|_2 = \|x\|_2$ μένει μόνο να δείξω ότι αν δειγματοληπτήσω το (αρκετά επίπεδο) διάνυσμα $x'' = FDx$ κατάλληλα, τότε το διάνυσμα που προκύπτει έχει μετά από την κανονικοποίηση περίπου την ίδια ℓ_2 μάζα.

Για το σκοπό αυτό θα μας φανεί χρήσιμη η <u>**Ανισότητα Bernstein:**</u>

Έστω ανεξάρτητες τυχαίες μεταβλητές $Z_1, Z_2, \ldots Z_m$ με $|Z_i| \leq$ K και $\sigma^2 = \sum_i Var(Z_i)$. Τότε

$$\mathbb{P}\left[\left|\sum_i Z_i - \mathbb{E}\left[\sum_i Z_i\right]\right| \geq \lambda\right] \lesssim e^{-\lambda/k} + e^{-\lambda^2/\sigma^2}$$

Μπορούμε να χρησιμοποιήσουμε την ανισότητα αυτή για να επιτύχουμε το ζητούμενο, θέτωντας $Z_i = (x_i'')^2$ αν διαλέξαμε το i, αλλιώς $Z_i = 0$. Θα δείτε στην τρίτη σειρά ασκήσεων την ανάλυση.