CR12: Statistical Learning & Applications

Johnson-Lindenstrauss theory

Lecturer: Joseph Salmon *Scribes: Jordan Frecon and Thomas Sibut-Pinote*

1 Subgaussian random variables

In probability, Gaussian random variables are the easiest and most commonly used distribution encountered.

Definition. *Subgaussian*
Let X (random variable) is σ-subgaussian if there exist $\sigma > 0$ such as :

$$\forall t \in \mathbb{R},\ \mathbb{E}\left[\exp(tX)\right] \leq \exp(\frac{\sigma^2 t^2}{2})\,. \tag{1}$$

The quantity $\mathbb{E}\left[\exp(tX)\right]$ is called the **moment generating function** in by probabilists or the **Laplace transform** by analysts.

Proposition. *X σ-subgaussian Assume that X is σ-subgaussian, then the following statement are true:* $\mathbb{E}(X) = 0$ *and* $\mathbb{E}(X^2) = \mathrm{Var}(X) \leq \sigma^2$

Proof.

$$\begin{aligned}\mathbb{E}(\exp(tx)) &= \sum_{n\geq 0} t^n \frac{\mathbb{E}(X^n)}{n!} \leq \exp(\frac{t^2\sigma^2}{2}) \quad \text{(Fubini)}\\ &= \sum_{n\geq 0} \left(\frac{\sigma^2 t^2}{2}\right)^n \frac{1}{n!}.\end{aligned}$$

Up to order 2 and rearranging terms of order greater than 2 on the l.h.s:

$$1 + t\mathbb{E}(X) + \frac{t^2}{2}\mathbb{E}(X^2) \leq 1 + \frac{\sigma^2 t^2}{2} + g(t) \tag{2}$$

where $\frac{g(t)}{t^2} \to_{t\to 0} 0$. So by dividing both side by t and taking the limit when$t \to 0_+$ we show that $\mathbb{E}(X) \leq 0$.
With $t \to 0_-$ we prove that $\mathbb{E}(X) \geq 0$. So $\mathbb{E}(X) = 0$.
By dividing both side of (2) by t^2 and taking the limit we obtain $\mathbb{E}(X^2) \leq \sigma^2$. □

Example. *1. $\mathcal{N}(0,\sigma^2)$ is σ-subgaussian.*
Indeed, during previous courses, it has been checked that if $X \sim \mathcal{N}(0,1)$ then $\mathbb{E}(\exp(tX)) = \int_{-\infty}^{+\infty} \exp(tX)\exp(-\frac{x^2}{2})\frac{\mathrm{dx}}{\sqrt{2\pi}} = \int_{-\infty}^{+\infty} \frac{\exp(-\frac{1}{2}(x-t)^2)}{\sqrt{2\pi}} \exp(\frac{1}{2}t^2)\mathrm{dx} = \exp(\frac{t^2}{2})$. So if $X \sim \mathcal{N}(0,1)$ then X is 1-subgaussian. Now if $Y \sim \mathcal{N}(0,\sigma^2)$, then $\frac{Y}{\sigma} = X \sim \mathcal{N}(0,1)$ holds too, and so $\mathbb{E}(\exp(tY)) = \exp(\frac{\sigma^2 t^2}{2})$ and Y is σ-subgaussian

2. *Rademacher variable ($\varepsilon = +1$ or $\varepsilon = -1$ with probability 1/2) are 1-subgaussian.*

$$\begin{aligned}\mathbb{E}(\exp(tX)) &= \mathbb{P}(x=-1)\exp(-t) + \mathbb{P}(x=+1)\exp(t) = \frac{\exp(-t)+\exp(+t)}{2} \\ &= \cosh(t) = \sum_{n\geq 0} \frac{t^{2n}}{(2n)!} \\ &\leq \sum_{n\geq 0} \frac{(t^2)^n}{2^n n!} = \exp(\frac{t^2}{2}) \quad (\textit{using } 2^n n! \leq (2n)!, \textit{ see Appendix})\end{aligned}$$

3. *Uniform random variables over a compact interval $[-a,a]$ is a-subgaussian*

$$\begin{aligned}\mathbb{E}(\exp(tX)) &= \int_{-a}^{a} \exp(tx)\frac{\mathrm{dx}}{2a} = \frac{1}{2a}(\exp(ta) - \exp(-ta)) \\ &= \mathrm{sh}(at) = \sum_{n\geq 0} \frac{(at)^{2n}}{(2n+1)!} \quad (\textit{using now } 2^n n! \leq (2n+1)!) \\ &\leq \exp(\frac{a^2t^2}{2}).\end{aligned}$$

In this case, a^2 is an upper bound on the variance of X, since $\mathrm{Var}(X) = \int_{-a}^{a} x^2 \frac{\mathrm{dx}}{2a} = [a^3 + a^3]\frac{1}{6a} = \frac{a^2}{3}$. *Can the bound be made sharper?*

4. *X is a bounded and centered random variable, with $X \in [a,b]$. Then X is $\frac{b-a}{2}$-subgaussian. (cf. Hoeffding's inequality and McDiarmind's proof (lecture 3). Remark that here we do not need $a=-b$.*

Theorem. *Assume that X is σ-subgaussian and that $\alpha \in \mathbb{R}$, then αX is $(|\alpha|\sigma)$-subgaussian. Moreover Assume that X_1 is α_1-subgaussian and X_2 is α_2-subgaussian, then (X_1+X_2) is $\sigma_1+\sigma_2$-subgaussian.*

Proof. For the first part:

$$\begin{aligned}\mathbb{E}(\exp(t\alpha X)) &\leq \exp(t^2\alpha^2\frac{t}{2}) &(3)\\ &\leq \exp(|\alpha^2|\frac{\sigma}{2}t^2) &(4)\end{aligned}$$

For the second part compute:

$$\mathbb{E}\left(\exp(t(X_1+X_2))\right) = \mathbb{E}\left(\exp(tX_1)\exp(tX_2)\right)$$

Then, let us introduce $\frac{1}{p} + \frac{1}{q} = 1$ for some $p \geq 1$. It leads to

$$\begin{aligned}\mathbb{E}\left(\exp(t(X_1+X_2))\right) &= \mathbb{E}\left(\exp(tX_1p)\right)^{\frac{1}{p}}\mathbb{E}\left(\exp(tX_2q)\right)^{\frac{1}{q}} \\ &\leq \left(\exp(\frac{\sigma_1^2}{2}t^2p^2\right)^{\frac{1}{p}}\left(\exp(\frac{\sigma_2^2}{2}t^2q^2\right)^{\frac{1}{q}} = \exp\left(\frac{t^2}{2}(p\sigma_1^2 + q\sigma_2^2)\right).\end{aligned}$$

For example, if we choose $p = q = \frac{1}{2}$ we get $\frac{\sigma_1^2+\sigma_2^2}{4}$ (meaning that Cauchy-Schwartz is suboptimal in that case). The idea is to optimize this bound over $p \geq 1$. This gives the following choice:

$$p^* = \frac{\sigma_2}{\sigma_1} + 1$$

and thus leads to the bound $\mathbb{E}\left[\exp(t(X_1+X_2))\right] \leq \exp(\frac{t^2(\sigma_1+\sigma_2)^2}{2})$. □

Theorem. *Assume that X_1 is α_1-subgaussian and X_2 is α_2-subgaussian, and that moreover X_1 and X_2 are independent, then $(X_1 + X_2)$ is $\sqrt{\sigma_1^2 + \sigma_2^2}$-subgaussian.*

Proof.

$$\mathbb{E}[e^{t(X_1+X_2}] = \mathbb{E}[e^{t(X_1)}]\mathbb{E}[e^{t(X_2)}]{=}e^{\frac{\sigma_1^2}{2}t^2+\frac{\sigma_2^2}{2}t^2} = \exp(\frac{t^2(\sqrt{\sigma_1^2+\sigma_2^2})}{2})$$

where the first equality holds because X_1 and X_2 are independent. □

Theorem (Characterization of subgaussian variables)**.** *Let assume $\mathbb{E}(X) = 0$. Then the following propositions are equivalent*[1]*:*

1. $\exists c_1 > 0, \quad \forall \lambda \geq 0, \quad \mathbb{P}(|X| \geq \lambda) \leq 2\exp(-\lambda^2 c_1)$ *(tail)*
2. $\exists c_2 > 0, \quad \forall p \geq 1, \quad (\mathbb{E}|X|^p)^{\frac{1}{p}} \leq c_2\sqrt{p}$ *(Moment control)*
3. $\exists c_3 > 0, \quad \mathbb{E}\left(\exp(c_3 X^2)\right) \leq 2$ *(Laplace transform of X^2 is bounded)*
4. $\exists c_4 > 0, \quad \mathbb{E}(\exp(tX)) \leq \exp(c_4 \frac{t^2}{2})$ *(Laplace transform decay)*

Remark. *The number 2 in the third claim is arbitrary.*

Remark. *You can find articles/books, where the first proposition is taken as the definition for subgaussian.*

Proof. $1 \Rightarrow 2$

We can assume that $c_1 = 1$ (otherwise consider $\sqrt{c_1}X$ instead of X). Then use Fubini's theorem to show that $\mathbb{E}|X|^p = \int_{-\infty}^{+\infty} pt^{p-1}\mathbb{P}(|X| \geq t)\,\mathrm{d}t$.

Indeed, for $X \geq 0$, $X = \int_0^X \mathrm{d}t = \int_0^{+\infty} \mathbb{1}_{\{X\geq t\}}\mathrm{d}t$. By using Fubini, we can show that $\mathbb{E}(X) = \int_0^{+\infty} \mathbb{E}\left(\mathbb{1}_{\{X\geq t\}}\right)\mathrm{d}t$. [2] In the same manner, $\mathbb{E}(|X|^p) = \int_0^{|X|} pt^{p-1}\mathrm{d}t = \int_0^{+\infty} pt^{p-1}\mathbb{1}_{\{|X|\geq t\}}\mathrm{d}t$. Now,

$$\begin{aligned}
\mathbb{E}(|X|^p) &\leq p\int_0^{+\infty} 2t^{p-1}\exp(-t^2)\mathrm{d}t && (5)\\
&\leq p\int_0^{+\infty} 2\sqrt{u}^{p-1}\exp(-u)\frac{\mathrm{d}u}{2\sqrt{u}} \quad \text{(by using the change of variable } u = t^2\text{)} && (6)\\
&\leq \int_0^{+\infty} u^{\frac{p}{2}-1}\exp -u = 2\left(\frac{p}{2}\right)\Gamma\left(\frac{p}{2}\right) = 2\Gamma\left(\frac{p}{2}+1\right) && (7)\\
&\leq 2(\frac{p}{2})^{\frac{p}{2}} && (8)
\end{aligned}$$

where we have used the definition of the Γ function and the classical inequality $\Gamma(x+1) \leq x^x$ for any $x \geq 0$ (see Appendix).

And so, $\mathbb{E}(|X|^p)^{\frac{1}{p}} \leq 2^{\frac{1}{p}}(\frac{p}{2})^{\frac{1}{2}} \leq \sqrt{p}\underbrace{\frac{2}{\sqrt{2}}}_{c_2}$ (since $p \geq 2$).

$2 \Rightarrow 3$ Same remark: we start by assuming $c_2 = 1$, or then we can reduce the problem to that one by dividing X by c_2.

[1] and in particular are equivalent to being subgaussian

[2] Remark: This is a very simple equality, but it is very frequently used in probabilities, .

$$
\begin{aligned}
\mathbb{E}[\exp(aX^2)] &= 1 + \sum_{n\geq 1} \frac{\mathbb{E}[(aX^2)^n]}{n!} \\
&\leq 1 + \sum_{n\geq 1} a^n \frac{\mathbb{E}(X^{2n})}{n!} \\
&\leq 1 + \sum_{n\geq 1} a^n \frac{\sqrt{2n}^{2n}}{n!} \quad \text{(since } c_2 = 1\text{)} \\
&\leq 1 + \sum_{n\geq 1} a^n \frac{2^n n^n}{n!} \\
&\leq 1 + \sum_{n\geq 1} a^n (2e)^n \quad \left(\text{by using } n! \geq \left(\frac{n}{\mathrm{e}}\right)^n, \text{ see Appendix}\right) \\
&\leq 2 \qquad \left(\text{choosing } 2ae \leq \frac{1}{2}, \text{ and using } \sum_{n\geq 1} \left(\frac{1}{2}\right)^n = 1\right)
\end{aligned}
$$

$3 \Rightarrow 4$

$$
\begin{aligned}
\mathbb{E}\left(\exp(tX)\right) &= 1 + \int_0^1 (1-y)\mathbb{E}(t^2X^2\exp(ytX))\mathrm{d}y \quad (\text{Tailor expansion} + \mathbb{E}(X) = 0) \\
&\leq 1 + \int_0^1 (1-y)\mathbb{E}\left(X^2t^2\exp(t|x|)\right)\mathrm{d}y \\
&\leq 1 + \frac{t^2}{2}\mathbb{E}\left(X^2\exp(t|X|)\right) \\
&\leq 1 + \frac{t^2}{2}\mathbb{E}\left(X^2\exp(\frac{t^2}{2c_3} + \frac{X^2}{2}c_3)\right) \quad (\text{using } ab \leq a^2/2 + b^2/2) \\
&\leq 1 + \frac{t^2}{2}\exp(\frac{t^2}{2c_3}) \qquad \underbrace{\mathbb{E}\left(X^2\exp(\frac{X^2}{2}c_3)\right)}_{\leq \frac{2}{c_3}\mathbb{E}[\exp(X^2c_3)] \text{ using } X\leq\exp(X)} \\
&\leq \exp\left(\frac{5t^2}{2c_3}\right)
\end{aligned}
$$

$4 \Rightarrow 1$ (Chernoff-Bernstein)[3] $\forall\lambda \geq 0$, $\mathbb{P}\left(X \geq t\right) = \mathbb{P}\left(\exp(\lambda X) \geq \exp(\lambda t)\right)$ Markov [4] :

$$
\begin{aligned}
\mathbb{P}(X \geq t) &\leq \frac{\mathbb{E}(\exp(\lambda X))}{\exp(\lambda t)} && (9) \\
&\leq \exp\left(c_4\frac{\lambda^2}{2} - \lambda t\right) \quad (\text{Optimization w.r.t. } \lambda \to \lambda^* = \frac{t}{c_4}) && (10) \\
&\leq \exp\left(\frac{-t^2}{2c_4}\right) && (11)
\end{aligned}
$$

□

[3]It seems that Bernstein should be credited too for this method.

[4]Reminder of the Chebychev inequality: $\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X^2]}{t^2}$.

Lemma. *Assume X is subgaussian such that $\frac{(\mathbb{E}[|X|^p])^{\frac{1}{p}}}{\sqrt{p}} \leq K$ for some $K \geq 0$ and that $\begin{cases} \mathbb{E}X = 0 \\ \mathbb{E}X^2 = 1 \end{cases}$ then,*

$$\exists c > 0, \quad \mathbb{E}\exp\left(t(X^2-1)\right) \leq \exp(t^2 c) \quad \textit{for} \quad |t| \leq \left(\frac{1}{2\mathrm{e}K^2}\right). \tag{12}$$

Proof. Define $Y = X^2 - 1$. Then, as before we can write:

$$\begin{aligned}\mathbb{E}\left(\exp(tY)\right) &= 1 + \mathrm{E}(Y)t + \sum_{p\geq 2}\frac{t^p}{p}\mathbb{E}(Y^p) \\ &= 1 + \sum_{p\geq 2}\frac{t^p}{p}\mathbb{E}(Y^p).\end{aligned}$$

Reminding the Minkowski Inequality,

$$\left[\mathbb{E}\left(|X^2-1|^p\right)\right]^{\frac{1}{p}} \leq \left[\mathbb{E}\left(X^{2p}\right)\right]^{\frac{1}{p}} + 1 \leq K^2 p + 1.$$

one obtains

$$\begin{aligned}\mathbb{E}(\exp(tY)) &\leq 1 + \sum_{p\geq 2}\frac{|t|^p}{p!}\left(2^p p^p K^{2p} + 1\right) \\ &\leq 1 + \sum_{p\geq 2}\frac{|t|^p}{p!}\left(2^p p^p K^{2p} + 1\right) \\ &\leq 1 + \sum_{p\geq 2}\left[\left(2|t|\mathrm{e}K^2\right)^p + \frac{|t|^p}{p!}\right] \quad \text{(by using } p! \geq \left(\frac{p}{\mathrm{e}}\right)^p \text{, see Appendix)} \\ &\leq 1 + t^2\sum_{p\geq 0}\left[2\mathrm{e}K^2\left(2|t|\mathrm{e}K^2\right)^p + \frac{|t|^p}{(p+2)!}\right].\end{aligned}$$

For $|t| \leq \frac{1}{2\mathrm{e}K^2}$ there exist c such as :

$$\mathbb{E}(\exp(tY)) \leq 1 + ct^2 \leq \exp(ct^2).$$

□

Corollary. *Let us assume $X_i \overset{\text{iid}}{\sim} X$ for $i = 1, \cdots, k$ with X subgaussian such that $\frac{\mathbb{E}(|X|^p)^{\frac{1}{p}}}{\sqrt{p}} \leq K$. Then, $\exists c > 0$, $\mathbb{E}\left[\exp\left(\frac{t}{\sqrt{k}}\sum_{i=1}^k (X_i^2 - 1)\right)\right] \leq \exp(t^2 c)$ for $|t| \leq \frac{\sqrt{k}}{2\mathrm{e}K^2}$.*

Proof.

$$\begin{aligned}\mathbb{E}\left[\exp\left(\frac{t}{\sqrt{k}}\sum_{i=1}^k (X_i^2-1)\right)\right] &= \prod_{i=1}^k \mathbb{E}\left[\exp\left(\frac{t}{\sqrt{k}}(X_i^2-1)\right)\right] \\ &\leq \prod_{i=1}^k \exp(t^2 c/k) \quad (\text{for } |t| \leq \frac{\sqrt{k}}{2\mathrm{e}K^2}) \\ &\leq \exp(t^2 c).\end{aligned}$$

□

2 Random projections in high dimension

2.1 Theoretical results

Theorem (Johnson-Lindenstrauss's Lemma). *Let X be a subgaussian random variable such that $\frac{\mathbb{E}(|X|^p)^{\frac{1}{p}}}{\sqrt{p}} \leq K$, and*

$$\mathbb{E}(\exp(t(X^2 - 1))) \leq \exp(t^2 c)$$

for $|t| \leq \frac{1}{2eK^2}$. For any $\varepsilon \leq \frac{c}{eK^2}$ define $k = \frac{4c}{\varepsilon^2}\beta \log(d)$ for some $\beta > 0$, Then generate $R_{i,j} \overset{\text{iid}}{\sim} X$ where R is a $k \times d$ matrix. Introduce $T : \mathbb{R}^d \to \mathbb{R}^k$ such that for $x \in \mathbb{R}^d$, we have

$$(Tx)_i = \frac{1}{\sqrt{k}} \sum_{j=1}^{d} R_{i,j} x_j \, ,$$

for $i = 1, \cdots, k$ Then with probability $\geq 1 - 2\left(\frac{1}{d}\right)^{\beta}$, the following holds :

$$\{\forall x \in \mathbb{R}^d, (1-\varepsilon)\|x\|^2 \leq \|Tx\|^2 \leq (1+\varepsilon)\|x\|^2\} \tag{13}$$

or

$$\{\forall (x,y) \in \mathbb{R}^d \times \mathbb{R}^d, (1-\varepsilon)\|x-y\|^2 \leq \|Tx - Ty\|^2 \leq (1+\varepsilon)\|x-y\|^2\} \tag{14}$$

Proof. Let us denote $x \in \mathbb{R}^d$, $u = \frac{x}{\|x\|}$ and Y_i the column values of the output, i.e $Y_i = \sum_{j=1}^{d} R_{i,j} x_j$. Then,

$$\mathbb{E}(Y_i) = \mathbb{E}\left(\sum_{j=1}^{d} R_{i,j} u_j\right) = \sum_{j=1}^{d} \mathbb{E}(R_{i,j} u_j) = \sum_{j=1}^{d} u_j \mathbb{E}(R_{i,j}) = 0$$

$$\text{Var}(Y_i) = \text{Var}\left(\sum_{j=1}^{d} R_{i,j} u_j\right) = \mathbb{E}\left(\sum_{j=1}^{d} R_{i,j} u_j\right)^2 = \sum_{j=1}^{d} \text{Var}\left(R_{i,j} u_j\right) = \sum_{j=1}^{d} u_j^2 \text{Var}(R_{i,j}) = 1^5$$

So $(Y_i)_{i=1,\cdots,k}$ are independent and subgaussian thanks to Theorem 1 (same constant as X). Defining $Z = \frac{1}{\sqrt{k}}(Y_1^2 + \cdots + Y_k^2 - k)$, one can state the following bound:

$$\begin{aligned} \mathbb{P}\left(\|Tu\|^2 \geq 1 + \varepsilon\right) &= \mathbb{P}\left(Z \geq \varepsilon\sqrt{k}\right) \\ &\leq \exp\left(-\frac{\varepsilon^2 k}{4c}\right) \quad \text{(following lemma)} \end{aligned}$$

Remind that $k = \frac{4c}{\varepsilon^2}\beta \log d$, so

$$\mathbb{P}\left(\|Tu\|^2 \geq 1 + \varepsilon\right) \leq \exp\left(-\beta \log d\right) = \left(\frac{1}{d}\right)^{\beta}$$

The same kind of derivations leads to:

$$\mathbb{P}\left(\|Tu\|^2 \leq 1 - \varepsilon\right) \leq \exp\left(-\beta \log d\right) = \left(\frac{1}{d}\right)^{\beta}$$

□

Lemma. $Z = \frac{1}{\sqrt{k}}(Y_1^2 + \cdots + Y_k^2 - k)$ *satisfies* $\mathbb{P}(Z \geq \varepsilon k) \leq \exp(\frac{-\varepsilon^2 k}{4c})$ *for* $\varepsilon \leq \frac{c}{eK^2}$.

Proof.

$$\forall \lambda \geq 0, \mathbb{P}(Z \geq \varepsilon\sqrt{k}) \leq \frac{\mathbb{E}(\exp \lambda Z)}{\exp(\lambda\sqrt{k})} \tag{15}$$

$$\leq \exp\left(\lambda^2 c - \lambda\varepsilon\sqrt{k}\right) \quad \text{(Optimize w.r.t. } \lambda \to \lambda = \frac{\varepsilon\sqrt{k}}{2c}) \tag{16}$$

$$\leq \exp\left(-\frac{\varepsilon^2 k}{4c}\right) \tag{17}$$

□

Remark. *Here are a few comments on the previous result:*

- ε *is the precision needed.*
- β *is a confidence parameter governing the probability.*
- $k \asymp \frac{\log(d)}{\varepsilon^2}$

2.2 Historical remarks

- [Johnson and Lindenstrauss(1984)]: random space with dimension k. Technical tool:"concentration on the sphere“. The proof was not constructive.
- [Indyk and Motwani(1998)] and then [Dasgupta and Gupta(2003)]: the random space are generated in an explicit way: $R_{i,j} \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$.
- [Achlioptas(2003)] extended the property for computationally more tractable random spaces: $R_{i,j} \overset{\text{iid}}{\sim} \varepsilon$ Rademacher. Interesting features of this distribution being that they require only sums and subtractions operations.
- [Matoušek(2008)] generalized the proof for any subgaussian random variables for the elements of $R_{i,j}$.
- [Ailon and Chazelle(2009)] focused on a even faster implementation: $R = M_{k,d}H_dD_d$ where $M_{k,d}$ is random $k \times d$ sparse matrix (with probability $q \asymp \frac{\log^2 d}{d}$ that a term is non zero, and Gaussian), D has diagonal generated according to Rademacher distributions and H is the Hadamard matrix defined by $H_{2d} = \begin{pmatrix} H_d & H_d \\ H_d & -H_d \end{pmatrix}$ and $H_1 = (1)$. The later allows for fast computation of matrix/vector multiplications: one can use recursively only sums/subtractions, leading to $O(d \log d)$ operations (similar to the standard FFT).

2.3 Application: k-Nearest-Neighbors (k-nn)

Let us consider m points $(x_1, \cdots, x_m)$ in $\mathbb{R}^d$ and suppose that a new point x is coming. Imagine that one needs to find the closest point $x \in \mathbb{R}^d$ for simplicity (you can deal with the k-nn problem in a similar way), meaning the following problem needs to be solved:

$$\arg\min_{i=1}^{m} \underbrace{d^2(x_i, x)}_{\|x_i - x\|^2}$$

where $\|\cdot\|_2$ is the Euclidean norm in $\mathbb{R}^d$.

Computational cost for the naive way: $\mathcal{O}(md)$ operations are need, because one has to compute for the m points, the distance to x in $\mathbb{R}^d$. On the other hand, using the Jonhson-Lindenstrauss theory, and using random projections of the form $T : \mathbb{R}^d \to \mathbb{R}^k$. $x_i \to Tx_i$, one only needs to perform $\mathcal{O}(m \log(d))$ operations (note that this does take into account the projection step that can be done as preliminary treatment).

Techniques similar to J.L: you could do "randomized" SVD eigenvalue decomposition. You might not get a perfect eigenvalue decomposition, but with high probability you will get something that is "accurate enough".

Appendix:Standard inequalities

Cauchy-Schartz, Hölder, etc.

Simple ones:

$$2^n n! \leq (2n)! \tag{18}$$

Indeed it is true for $n = 0$, and for $n \geq 1$ then $(2n)! \geq 2n(2n-1)\dots(n+1)n!$ and then lower bound each of the first elements by 2.

$$n! \geq \left(\frac{n}{\mathrm{e}}\right)^n \tag{19}$$

$\mathrm{e}^n = \sum_{i=0}^{+\infty} \frac{n^i}{i!} \geq \frac{n^n}{n!}$, where the later holds by comparing lower bound the sum by the term corresponding to $i = n$.

References

[Achlioptas(2003)] D. Achlioptas. Database-friendly random projections: Johnson-lindenstrauss with binary coins. Journal of computer and System Sciences, 66(4):671–687, 2003.

[Ailon and Chazelle(2009)] N. Ailon and B. Chazelle. The fast Johnson-Lindenstrauss transform and approximate nearest neighbors. SIAM J. Comput., 39(1):302–322, 2009. ISSN 0097-5397.

[Dasgupta and Gupta(2003)] S. Dasgupta and A. Gupta. An elementary proof of a theorem of johnson and lindenstrauss. Random Structures & Algorithms, 22(1):60–65, 2003.

[Indyk and Motwani(1998)] P. Indyk and R. Motwani. Approximate nearest neighbors: towards removing the curse of dimensionality. In Proceedings of the thirtieth annual ACM symposium on Theory of computing, pages 604–613. ACM, 1998.

[Johnson and Lindenstrauss(1984)] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. Contemporary mathematics, 26(189-206):1, 1984.

[Matoušek(2008)] J. Matoušek. On variants of the Johnson–Lindenstrauss Lemma. Random Structures & Algorithms, 33(2):142–156, 2008.