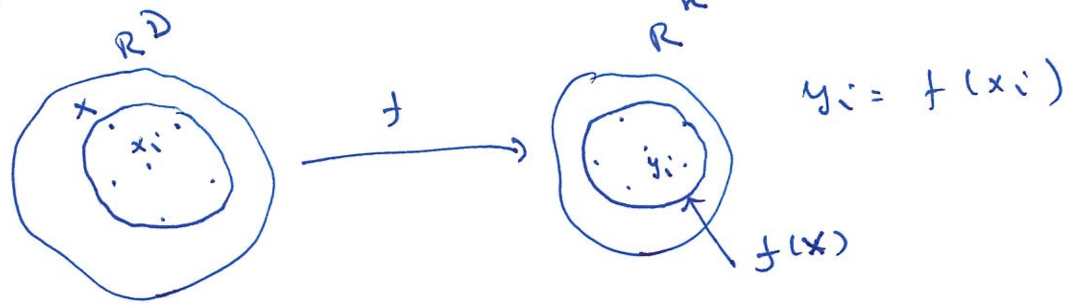
Topic 4. Dimension Reduction

- Deterministic → (P.C.A) Principal Component Analysis
- Randomized → Random Projections

1) Introduction

- Let X be finite subset of vectors in $\mathbb{R}^D$, D - large
- Goal is to find a mapping $f : X \subset \mathbb{R}^D \to \mathbb{R}^k$ ~~where $k \le D$~~ such that $f(X) =$ ~~[illegible]~~ $\{ f(x_i) \mid x_i \in X \}$ the set of all images of $x_i \in X$ in $\mathbb{R}^k$ have the desired property.

$\boxed{|X| = n}$



- One property of interest is the <u>preservation of pair-wise distances</u>. Consider Euclidean distances
- The map is said to be an <u>isometry</u> if for $u, v \in X$

$$\| f(u) - f(v) \| = \| u - v \| \longrightarrow (1)$$

- Isometric embedding is feasible for $k \geq D$. But, for $k \leq D$, the best we can hope for <u>near isometry</u> where distances are approximately preserved.

<u>Definition 1</u> $f: X \subset R^D \to R^k$ is called a <u>Lipschitz map</u> if for $u, v \in X$ and constants A and B the following holds:

$$A\|u-v\|^2 \le \|f(u)-f(v)\|^2 \le B\|u-v\|^2 \quad \to (2)$$

(ie) distances are approximately preserved.

- Methods to realize Lipschitz map is not easy and it is here random projections come into the picture.
- It is shown below that the map f that approximately preserve the distance can be realized as a linear map defined by random projection matrix

2) <u>Random Projection</u>: Let $X \subset R^D$ (ie) $X \in R^{D\times n}$ matrix where the n elements are the n columns of the given data matrix where D is large. k is the dimension of the target space.

- Let $R^T = [r_1, r_2, \ldots r_k] \in R^{D\times k}$ be the matrix of <u>linear transformation</u> where

$$Y = f(X) = R^* X \quad \text{(ie)} \quad \underset{k\times n}{Y} = \underset{k\times D}{R^*}\ \underset{D\times n}{X} \quad \to (3)$$

$$[Y_1\, Y_2 \cdots Y_n] = R^*[X_1\, X_2 \cdots X_n]$$

$Y_i \in R^k$, $R \in R^{k\times D}$, $X_i \in R^D$

(ie) Y's are samples in R^k

- Let $R \in \mathbb{R}^{k \times D}$ matrix of linear transformation of interest in realize the Lipschitz map that is very near isometry.
- Define $X = [x_1 \; x_2 \cdots x_n] \quad x_i \in \mathbb{R}^D$
  $Y = [y_1, y_2, \cdots y_n] \quad y_i \in \mathbb{R}^k$
- Define
$$\underset{k \times n}{Y} = \underset{k \times D}{R} \; \underset{D \times n}{X} \longrightarrow (3)$$

where $R = [r_{ij}] \quad 1 \le i \le k, \; 1 \le j \le D$

maps the vector $x_i \in \mathbb{R}^D$ to $y_i \in \mathbb{R}^k$ for $1 \le i \le n$ and $k << D$.

<u>Note:</u> . When we realize f as a linear map, it is computationally simpler to implement.

- Notice X is used to denote the set of n points in $\mathbb{R}^D$ as well as the matrix of size $D \times n$ where each vector is a column.
- X represents the given configuration of data in $\mathbb{R}^D$ and Y is the embedded configuration in $\mathbb{R}^k$
- Our goal is to preserve the local separation approximately.

- Two questions arise: How to pick the elements of the matrix $R \in \mathbb{R}^{k \times D}$? Prove that such a choice can indeed realize near isometric embedding of $X \subset \mathbb{R}^D$ in $\mathbb{R}^k$ for $k << D$.

- There are various choices for the elements of R.

  - Let r_{ij} be the typical element of R, $1 \le i \le k$, $1 \le j \le D$. There are three types

Types

1) $r_{ij} \sim N(0, 1)$ i.i.d samples

2) $r_{ij} \in \{-1, 1\}$ with probability $\frac{1}{2}$ each

3) $r_{ij} \in \{\pm\sqrt{3}, 0\}$: $r_{ij} = \pm\sqrt{3}$ with prob. $= 1/6$
$= 0$ with prob $= 2/3$

- In each case, verify

$$E[r_{ij}] = 0, \quad \mathrm{var}(r_{ij}) = 1$$

- The random matrix R is said to be of Type 1, 2, 3 if r_{ij} is of type 1, 2, 3 respectively.

- The action of R is defined as follows.

If $v = f(u)$, then

$$\boxed{v = f(u) = \frac{1}{\sqrt{k}} R u} \longrightarrow (4)$$

(ie) the D vectors in X are mapped to k vectors in Y.

- The set Y is called the random projection of X.

Property 1:- Let $R \in \mathbb{R}^{k \times D}$ matrix with its elements being i.i.d random variables with zero mean and unit variance. Then the map $f: \mathbb{R}^D \to \mathbb{R}^k$ defined by

$$f(a) = \frac{1}{\sqrt{k}} R a, \quad a \in \mathbb{R}^D. \to (5)$$

Then $f(a)$ is a random vector in $\mathbb{R}^k$ whose components are i.i.d random variables with mean zero and variance $\frac{1}{k}\|a\|^2$. That is

$$E\left[\|f(a)\|^2\right] = \|a\|^2 \longrightarrow (6).$$

Proof:- Let r^i be the ith row vector of R.

Then the inner product $\langle r^i, a \rangle = c_i$.

Then, $c = (c_1, c_2, \dots c_k)^T$ is given by

$$\mathbf{c} = f(a) = \frac{1}{\sqrt{k}} R a$$

- $$E(c_i) = \frac{1}{\sqrt{k}} \sum_{j=1}^{D} E[R_{ij} a_j]$$

$$= \frac{1}{\sqrt{k}} \sum_{j=1}^{D} a_j \, E[R_{ij}] = 0$$

- $$E[c_i c_j] = \frac{1}{k} \sum_{m=1}^{D} \sum_{j=1}^{D} a_m a_l E[r_{im} r_{jl}]$$

$$= 0 \text{ for } i \neq j$$

- $$E[c_i^2] = \frac{1}{k} E\left[\left(\sum_{j=1}^{D} a_j r_{ij}\right)^2\right]$$

$$= \frac{1}{k}\left[\sum_{j=1}^{D} a_j^2 E[r_{ij}^2] + 2\sum_{l \neq m} a_l a_m E[r_{il}] E[r_{jm}]\right]$$

$\boxed{r_{ij} \text{ are i.i.d} \quad E(r_{ij}) = 0 \quad E(r_{ij}^2) = 1}$

$$= \frac{1}{k}\sum_{j=1}^{D} a_j^2 = \frac{1}{k}\|a\|^2$$

$$\therefore \quad E[\|c\|^2] = E[\|f(a)\|^2] = E\left[\sum_{i=1}^{D} c_i^2\right]$$

$$= \sum_{i=1}^{D} E(c_i^2)$$

$$= k \cdot \frac{1}{k}\|a\|^2 = \|a\|^2.$$

- This result holds good for all three types of matrices described above.

<u>Property 2</u>:- Let f be a map as in Property 1. Let $a \in \mathbb{R}^D$ be a unit vector. Then $\sqrt{k}\, f(a)$ is a random vector in $\mathbb{R}^k$ whose entries are i.i.d r.v. with mean zero and unit variance

<u>Remarks</u>:-

1) The matrix $R \in \mathbb{R}^{D\times k}$ has non-trivial null space: $N(R) = \{u \in \mathbb{R}^D : Ru = 0\}$. Hence, the set X of points <u>cannot</u> be totally ~~arbitrary~~ arbitrary. So on a given sub-set we need to justify that Lipschitz embedding exits.

2) Johnson-Lindenstrauss first proved the following:

<u>J-L lemma (1981)</u>: Let $\varepsilon > 0$ and n, an integer be given. Then for all integers $k \geq k_0 = O(\varepsilon^{-2}\log n)$ and a set X of n points

randomly selected in R^D, there exists a $f: R^D \to R^k$ which satisfies

$$(1-\varepsilon)\|u-v\|^2 \le \|f(u)-f(v)\|^2 \le (1+\varepsilon)\|u-v\|^2 \quad \to (7)$$

for all $u, v \in X$.

- The f in J-L lemma is called J-L embedding.
- J-L did not give lower bound on k.
- It was made clear subsequently that f can be realized using random projection.
- Various lower bounds on k was subsequently established.

3) Random Projection based on Gaussian Variables

Theorem 1:- Dasgupta and Gupta (1999) For any $\varepsilon > 1$ and integer $n > 0$, let k be a positive integer such that

$$k \geq 4\left[\frac{\varepsilon^2}{2} - \frac{\varepsilon^3}{3}\right]^{-1} \log_e n. \quad \to (8)$$

$\log_e n = \ln n$
natural log

Then for any set $X \subset R^D$ of n-points, there is a linear map, $f: R^D \to R^k$ such that for all $u, v \in X$

$$(1-\varepsilon)\|u-v\|^2 \le \|f(u)-f(v)\|^2 \le (1+\varepsilon)\|u-v\|^2 \quad \to (9)$$

Furthermore, this map can be found in

random polynomial time.

Note:- An algorithm has randomized polynomial time if it runs in polynomial time (in input-size) and, if the correct answer is No, it always returns No, while the correct answer is Yes, it returns Yes with a positive probability.

- The proof of this theorem depends on a

Lemma 2: Let $R \in \mathbb{R}^{k \times D}$ random matrix of Type 1 and let $a \in \mathbb{R}^D$ be a unit vector. Let $y = Ra$ and $\beta > 1$. Then

$$P_r\left[\|y\|^2 \le \frac{k}{\beta} \right] < \exp\left[\frac{k}{2}\left(1 - \frac{1}{\beta} - \ln\beta\right)\right] \quad \rightarrow (9)$$

$\boxed{\ln = \text{natural log}}$

and

$$P\left[\|y\|^2 \geq k\beta \right] < \exp\left[\frac{k}{2}\left(1 - \beta + \ln\beta\right)\right] \quad \rightarrow (10)$$

Note:-

- These are related to concentration results.

- First let us take this Lemma for granted and use it to prove the theorem 1.

- We will come back later to prove this Lemma 2.

Proof of Theorem 1:- Recall R is a $k \times D$ matrix with i.i.d. elements from $N(0,1)$

. Let $u, v \in X$ be two distinct elements of X. and let $a = \dfrac{u-v}{\|u-v\|}$. Let $f(a) = \dfrac{1}{\sqrt{k}} R a$.

. define $z = f(a)$ and $y = \sqrt{k}\, z = Ra$

~~Then~~ $\cancel{(f(u) - f(v)) =}$ Then,

$$\|z\|^2 = \|f(a)\|^2 = \left\| f\left(\frac{u-v}{\|u-v\|}\right)\right\|^2$$

$$= \left\| \frac{1}{\sqrt{k}} R \left(\frac{u-v}{\|u-v\|}\right)\right\|^2$$

$$= \frac{1}{\|u-v\|^2 k} \|R(u-v)\|^2 \longrightarrow (11)$$

. Now $f(u) = \dfrac{1}{\sqrt{k}} R u$, $f(v) = \dfrac{1}{\sqrt{k}} R v$

$$f(u) - f(v) = \frac{1}{\sqrt{k}} R(u-v)$$

$$\Rightarrow \|f(u) - f(v)\|^2 = \frac{1}{k} \|R(u-v)\|^2 \longrightarrow (12)$$

. Combining (11) and (12):

$$\|z\|^2 = \frac{\|f(u) - f(v)\|^2}{\|u-v\|^2} \longrightarrow (13)$$

. Thus

$$P\left[\frac{\|f(u) - f(v)\|^2}{\|u-v\|^2} \le 1-\varepsilon\right] = P\left[\|z\|^2 \le 1-\varepsilon\right]$$

$$= P\left[\|y\|^2 \le k(1-\varepsilon)\right]$$

$$\longrightarrow (14)$$

. Likewise (please verify)

$$P\left[\frac{\|f(u)-f(v)\|^2}{\|u-v\|^2} \geq 1+\varepsilon\right] = P\left[\|y\|^2 \geq k(1+\varepsilon)\right] \rightarrow (15)$$

Part 1 : Now set $\boxed{\frac{1}{\beta} = 1-\varepsilon}$ in lemma 2 and apply to the right hand side of (14) to get (using (9))

$$P\left[\|y\|^2 \leq (1-\varepsilon)k\right] < \exp\left[+\frac{k}{2}\left[1-(1-\varepsilon)+\ln(1-\varepsilon)\right]\right] \rightarrow (16)$$

. Recall: $\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4}$ for $-1 \leq x \leq 1$

$$\Rightarrow \log(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} \ldots$$

$$\leq -x - \frac{x^2}{2} \longrightarrow (17)$$

. Apply to the exponent on the r.h.s. of (16)

$$1-(1-\varepsilon)+\ln(1-\varepsilon)$$

$$\leq \cancel{1} - \cancel{1} + \cancel{\varepsilon} - \cancel{\varepsilon} - \frac{\varepsilon^2}{2} = -\frac{\varepsilon^2}{2} \longrightarrow (18)$$

. combining (16) - (18) :

$$\boxed{P\left[\|y\|^2 \leq (1-\varepsilon)k\right] \leq \exp\left[-\frac{k\varepsilon^2}{2}\right]} \rightarrow (19)$$

Part 2 . Now set $\beta = 1+\varepsilon$ in Lemma 2 and apply it to the r.h.s. of (15) [using (10)]

$$P\left[\|y\|^2 \geq k(1+\varepsilon)\right] \leq \exp\left[+\frac{k}{2}\left(1-(1+\varepsilon)+\ln(1+\varepsilon)\right)\right] \rightarrow (20)$$

. The exponent becomes: $\log(1+\varepsilon) \leq \varepsilon - \frac{\varepsilon^2}{2} + \frac{x^3}{3} \rightarrow (21)$

$$\cancel{\frac{k}{2}\left[1-1-\varepsilon+\varepsilon-\frac{\varepsilon^2}{2}+\frac{x^3}{3}\right]-\frac{k}{}}$$

Then

$$\frac{k}{2}\left[1-(1+\varepsilon)+\ln(1+\varepsilon)\right]$$

$$\le \frac{k}{2}\left[1-1-\varepsilon+\varepsilon-\frac{\varepsilon^2}{2}+\frac{\varepsilon^3}{3}\right]$$

$$= \frac{k}{2}\left[-\frac{\varepsilon^2}{2}+\frac{\varepsilon^3}{3}\right] \longrightarrow (22)$$

. Combining (20) - (22) $\Longrightarrow$

$$\boxed{P\left[\|y\|^2 \geq k(1+\varepsilon)\right] \le e^{-\frac{k}{2}\left[\frac{\varepsilon^2}{2}-\frac{\varepsilon^3}{3}\right]}} \longrightarrow (23)$$

. Now consider the r. h. s. of (19) and (23):

. Recall: $$\frac{k}{2}\varepsilon^2 > \frac{k}{2}\left[\varepsilon^2/2-\frac{\varepsilon^3}{3}\right]$$

$$\Rightarrow \quad -\frac{k}{2}\varepsilon^2 \le -\frac{k}{2}\left[\frac{\varepsilon^2}{2}-\frac{\varepsilon^3}{3}\right]$$

$$\Rightarrow \quad \boxed{\exp\left[-\frac{k}{2}\varepsilon^2\right] \le \exp\left[-\frac{k}{2}\left(\frac{\varepsilon^2}{2}-\frac{\varepsilon^3}{3}\right)\right]} \longrightarrow (24)$$

. Hence, set

$$\exp\left[-\frac{k}{2}\left(\frac{\varepsilon^2}{2}-\frac{\varepsilon^3}{3}\right)\right] \le \exp\left[-2\ln n\right] \longrightarrow (25)$$

. This inequality (25) is true when

$$-\frac{k}{2}\left(\frac{\varepsilon^2}{2}-\frac{\varepsilon^3}{3}\right) \le 2\log n$$

$$\Rightarrow \quad \frac{k}{2}\left(\frac{\varepsilon^2}{2}-\frac{\varepsilon^3}{3}\right) \geq 2\log n$$

$$\Rightarrow \quad k \geq \frac{4\log n}{\left[\frac{\varepsilon^2}{2}-\frac{\varepsilon^3}{3}\right]} \longrightarrow (26)$$

. This applies to (20) and (23) simultaneously.

Thus, when k satisfies (26):

$$\exp[-2\log_e n] = e^{-2\ln n} = e^{-\log_e n^2}$$

$$= e^{\log\left(\frac{1}{n^2}\right)} = \frac{1}{n^2} \rightarrow (27)$$

~~Thus: For $u, v \in X$ and $u \neq v$:~~

$$\cancel{P\left[\frac{\|f(u)-f(v)\|^2}{\|u-v\|^2} \notin [1-\varepsilon, 1+\varepsilon]\right]}$$

~~Go To page 13~~

- So by (24) this bound applies to the r.h.s. of (19) and (23). Hence from (19) and (23) we get

$$P\left[\frac{\|y\|^2}{k} \leq (1-\varepsilon)\right] \leq \frac{1}{n^2}$$

and $$P\left[\frac{\|y\|^2}{k} \geq (1+\varepsilon)\right] \leq \frac{1}{n^2}$$

$$\boxed{\begin{aligned} &P[A \cup B] \\ &\leq P[A] + P(B) \end{aligned}}$$

- Then $$P\left[\frac{\|y\|^2}{k} \notin [1-\varepsilon, 1+\varepsilon]\right] \leq \frac{2}{n^2} \rightarrow (28)$$

- Hence, from

$$\frac{\|y\|^2}{k} = \frac{\|f(u)-f(v)\|^2}{\|u-v\|^2} \rightarrow (29)$$

we get that for each pair u, v and $u \neq v$,

$$(1-\varepsilon)\|u-v\|^2 \leq \|f(u)-f(v)\|^2 \leq (1+\varepsilon)\|u-v\|^2$$

holds with probability at least $\left(1-\frac{2}{n^2}\right)$

Lemma 2: Let $R \in \mathbb{R}^{k \times D}$ random matrix of Type 1 and let $a \in \mathbb{R}^D$ be a unit vector. Let $y = Ra$ and $\beta > 1$. Then

$$P\left[\|y\|^2 \leq \frac{k}{\beta}\right] < \exp\left[\frac{k}{2}\left(1 - \frac{1}{\beta} - \ln\beta\right)\right] \quad \to (9)$$

and

$$P\left[\|y\|^2 \geq k\beta\right] < \exp\left[\frac{k}{2}\left(1 - \beta + \ln\beta\right)\right] \quad \to (10).$$

Proof: Consider (9):

- $$P\left[\|y\|^2 \leq \frac{k}{\beta}\right] = P\left[\exp(-h\|y\|^2) \geq \exp\left(-\frac{kh}{\beta}\right)\right]$$

(Markov inequality) $$\leq \underbrace{E\left[\exp(-h\|y\|^2)\right]}_{I} \exp\left[\frac{kh}{\beta}\right] \quad \to (30)$$

Want to find the min. value of r.h.s of (30)

- Compute I:
  - $$\underset{k\times 1}{y} = \underset{k \times D}{R}\,\underset{D\times 1}{a} \Rightarrow y_i = \sum_{j=1}^{D} r_{ij} a_j$$ - Linear combination of r_{ij}
  - $r_{ij} \sim$ i i d $N(0,1)$, $\Rightarrow$ y_i is Normal
  - $E(y_i) = 0$ and $E(y_i^2) = Var(y_i)$
  $$= \sum E(r_{ij}^2)\, a_j^2$$
  $$= \sum a_j^2 = 1$$
  $\therefore \boxed{y_i \sim N(0,1)}$
  - $$E[y^T y] = E\left(\sum_{i=1}^{k} y_i^2\right) = \sum_{i=1}^{k} E(y_i^2) = k$$

(31)

$$\therefore E\left[\exp(-h y_i^2)\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-h y_i^2}\, e^{-\frac{y_i^2}{2}}\, dy_i \quad \to (31)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y_i^2\left[h + \frac{1}{2}\right]}\, dy_i$$

But $e^{-y_i^2 (h + \frac{1}{2})} = e^{-y_i^2 \left[\frac{2h+1}{2}\right]}$

$$\boxed{a^2 = \frac{1}{1+2h}} \qquad = e^{-\frac{y_i^2}{2\cdot\left(\frac{1}{1+2h}\right)}} = e^{-\frac{y_i^2}{2a^2}}$$

$$\therefore E\left[\exp(-h y_i^2)\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y_i^2}{2a^2}} \, dy_i$$

$$= a \cdot \frac{1}{\sqrt{2\pi}\, a} \int_{-\infty}^{\infty} e^{-\frac{y_i^2}{2a^2}} \, dy_i$$

$$= a = \frac{1}{\sqrt{1+2h}} \longrightarrow (32)$$

. Hence, from (30): <u>Combining (30), (31) and (32)</u>

$$E\left[\exp\left[-h \|y\|^2\right]\right] \exp\left[\frac{hk}{\beta}\right]$$

$= -h \sum_{i=1}^{k} y_i^2$

$$= \left(\frac{1}{\sqrt{1+2h}}\right)^k \exp\left[\frac{hk}{\beta}\right] = (1+2h)^{-\frac{k}{2}} \exp\left[\frac{hk}{\beta}\right] \longrightarrow (33)$$

. Define $g(h) = (1+2h)^{-k/2} \exp\left[\frac{hk}{\beta}\right]$ $\boxed{h > 0}$ $\longrightarrow (34)$

H.W. [. Verify that $g(h)$ attains a min value at $h^* = \frac{\beta - 1}{2}$. Then, from (34):

$$\Rightarrow \boxed{g(h^*) = \beta^{-\frac{k}{2}} \exp\left[\frac{k}{2}(1-\beta)\right]} \longrightarrow (35)$$

. Recall: $a = e^{\ln a}$ for $a > 0$ $\longrightarrow (36)$

Thus :

$$\beta^{-k/2} = e^{\ln \beta^{(-k/2)}}$$

$$= e^{-\frac{k}{2}\ln\beta} \longrightarrow (37)$$

- Substitute (37) in (35) :

~~min. val of $g(h)$~~ $\cancel{= g(h^*) = e^{-\frac{k}{2}\ln\beta}\, e^{\frac{hk}{\beta}}}$

$\cancel{= \exp\left[\frac{k}{2}\right.}$

- Then

$$g(h^*) = e^{-\frac{k}{2}\ln\beta}\, e^{\frac{k}{2}[1-\beta]}$$

$$= \exp\left[\frac{k}{2}(1-\beta-\ln\beta)\right] \longrightarrow (38)$$

(ii) The min. value of the r. h. s. of (30) is given by (38) which is the claim (9) in Lemma 2.

<u>Now, consider (10)</u>

$$P\left[\|y\|^2 \geq k\beta\right] = P\left[\exp\left(h\|y\|^2\right) > \exp(hk\beta)\right]$$

$$\boxed{\text{Markov inequality}} \quad \leq E\left[\exp\left(h\|y\|^2\right)\right]\exp[-hk\beta] \longrightarrow (39)$$

- Verify by the above argument that

$$\left.\begin{aligned} E\left[\exp\left(h\|y_i\|^2\right)\right] &= \frac{1}{\sqrt{1-2h}} \quad (0<h<\tfrac{1}{2}) \\ E\left[\exp\left(h\|y\|^2\right)\right] &= (1-2h)^{-k/2} \end{aligned}\right\} \longrightarrow (40)$$

$\therefore \quad P\left[\,\|y\|^2 \geq k\beta\,\right] \leq (1-2h)^{-\frac{k}{2}} \exp\left[-hk\beta\right]$

. Define $\quad g(h) = (1-2h)^{-k/2} \exp\left[-hk\beta\right]$

. $g(h)$ attains its min. at $h^* = \dfrac{\beta-1}{2\beta}$

and $\quad g(h^*) = \beta^{\frac{k}{2}} \exp\left[-\frac{k}{2}(\beta-1)\right]$

$$= \exp\left[-\frac{k}{2}(\beta-1) + \frac{k}{2}\ln\beta\right]$$

$$= \exp\left[\frac{k}{2}\left(1-\beta-\ln\beta\right)\right] \rightarrow (41)$$

which proves (40).

This concludes the proof of the theorem.