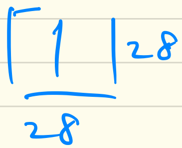
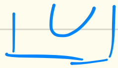
Random Projection

How to effective reduce dim?

dataset

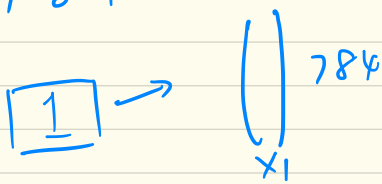




. . .

784 - dim

Query



Retrieve all imges with "1"

metric on x_1, x_2

$d(x_1, x_2) = \|x_1 - x_2\|_2$

$d \to 0 \quad x_1 \approx x_2$

$\to \infty \quad x_1 \neq x_2$

n: # img in database,

d: dim

$O(nd)$

① reduce n → hashing (not in this course)

② reduce d → dim reduction ✓

$d = 784$ $\qquad$ $k = 1$

10

100

$\square$

~~123~~

Given k

How to evaluate performance of the new data?

X_i in DB.

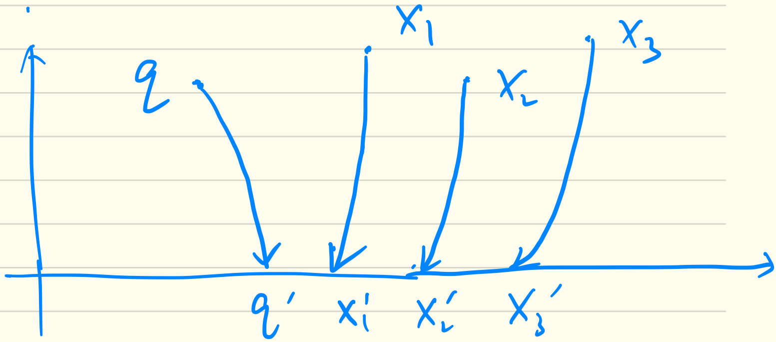
$$\| q - X_i \|_2$$

~~Ke~~ ~~Gra~~ Goal: $q', x_1' \cdots x_n' \in R^k$ on new data, Top 100 are same as Top 100 R^d

$$f: q \to q' = f(q) \quad R^d \to R^k.$$

$$x_i \to x_i' = f(x_i)$$

$$\| q' - x_i' \| = \| f(q) - f(x_i) \|$$



$$\boxed{\| f(q) - f(x_i) \|_2 \approx \| q - x_i \|_2}$$

$$f: \quad X \longrightarrow X'$$

$$x \Longrightarrow M \cdot x.$$

M: $k \times d$ matrix

$M_{ij} \sim N(0, \frac{1}{\;})$

random matrix

random projec.

$\forall\, q, \; x$

$$\| M \cdot q - M \cdot x \| \approx \| q - x \|$$

$$\Downarrow$$

$$\| M \cdot (q - x) \| \approx \| q - x \|$$

Goal: w.p. 0.99 $\underbrace{\| M \cdot x \|}_{\text{R.V.}} \approx \boxed{\begin{array}{c} \| x \| \\ \| \\ E[\| M \cdot x \|] \end{array}}$

$\| M \cdot x \| \approx E[\| M \cdot x \|]$

next week.

key msg from concentration inequalities:

w.h.p.

$$X \approx E[X]$$

Random Projection:

$\forall\, x \in \mathbb{R}^d$

$$\|M \cdot x\| \approx \|x\|$$

where M: $k \times d$ matrix

each $M_{ij} \sim N(0, 1/k)$

$\Rightarrow$ w.h.p.

$$1 - \epsilon \le \frac{\|M \cdot x\|}{\|x\|} \le 1 + \epsilon$$

$\epsilon \approx 0.01$

$$0.99 \le \quad \approx 1 \quad \le 1.01$$

Hope: $\boxed{\underbrace{\|M \cdot x\|}_{\text{R.V.}}} \approx \|x\|$

$\in \mathbb{R}$

Same

$\forall$ R.V. Z. $\quad Z \approx E[Z]$

(key msg from lect 1-2)

$\|M \cdot x\| \approx \boxed{E\left[\|M \cdot x\|\right]}$

Assume. $\underbrace{E[\|M \cdot x\|] = \|x\|}$

Then ~~to cent~~ concentration ineq. $\Rightarrow \underbrace{\|M \cdot x\| \approx \|x\|}$

Lemma: $E\left[\|M \cdot x\|^2\right] = \|x\|^2$

Proof: RHS $= \|x\|^2$

$= x^T x$

(x as column vector in $\mathbb{R}^d$)

$x = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}$

$\|x\|^2 = \sum_{i=1} x_i^2$

$x^T x = (x_1 \cdots x_d) \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}$

$= \sum x_i^2$

$\|M \cdot x\|^2 \approx \|x\|^2$

$\Rightarrow \|M \cdot x\| \approx \|x\|$

LHS

$$= E\left[\| \underbrace{M \cdot x}_{v} \|^2 \right]$$

$$= E\left[(M \cdot x)^T \cdot M \cdot x \right]$$

$$= E_M\left[x^T M^T M x \right]$$

$$= x^T \underbrace{\boxed{E[M^T M]}}_{= I_{d \times d}} x$$

$x \in \mathbb{R}$, $M \in \mathbb{R}^{k \times d}$

$x = 1$, $M = \cancel{x} \sim N(0,1)$

$$E[x \cdot M^T M y] = E[1 \cdot M^T M \cdot 1]$$

$$E[M] \triangleq \begin{bmatrix} E[M_{11}] & \cdots & E[M_{1d}] \\ \vdots & \ddots & \vdots \\ E[\ ] & \cdots & E[\ ] \end{bmatrix}$$

To prove: $E[M^T M] = I_{d\times d}$

$$M = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ m_1 & m_2 & \cdots & m_d \\ \downarrow & \downarrow & & \downarrow \end{pmatrix}_{k\times d}.$$

each $m_i \in \mathbb{R}^k$

$$M^T = \begin{pmatrix} \leftarrow & m_1^T & \rightarrow \\ \leftarrow & m_2^T & \rightarrow \\ & \vdots & \\ \leftarrow & m_d^T & \rightarrow \end{pmatrix}$$

For all $m_i \in \mathbb{R}$

$$M = (m_1 \;\; \cdots \;\; m_d)$$

$$M^T = \begin{pmatrix} m_1^T \\ \vdots \\ m_d^T \end{pmatrix}$$

$$E[M^T M]$$

$$= E\left[\begin{pmatrix} m_1^T \\ m_2^T \\ \vdots \\ m_d^T \end{pmatrix} \begin{pmatrix} m_1 & \cdots & m_d \end{pmatrix} \right]$$

$$= E\left[\begin{pmatrix} m_1^T m_1 & m_1^T m_2 & \cdots & m_1^T m_d \\ m_2^T m_1 & \cdots & \cdots & m_2^T m_d \\ \vdots & & & \\ m_d^T m_1 & \cdots & \cdots & m_d^T m_d \end{pmatrix} \right]$$

$$= \begin{pmatrix} E[m_1^T m_1^T] & E[\ ] & \cdots \\ \vdots & \vdots & \\ \vdots & E[\ ] & \\ \cdots & \cdots & E[m_d^T m_d] \end{pmatrix}_{d \times d}$$

$$E[m_i^T m_j], \quad 1 \le i \le d, \quad 1 \le j \le d \quad = I$$

Now: $\underline{E[m_i^T m_j]}$

$a \triangleq m_i$, $b \triangleq m_j$ $\in \mathbb{R}^k$

$\underline{E}[\underbrace{a^T b}]$

$= E[\sum_{i=1}^{k} \underline{a_i} \, \underline{b_i}]$

$= \sum_{i=1}^{k} \underbrace{E[a_i b_i]}$ $\nearrow$ I

Case 1: $m_i \neq m_j$ $(a \neq b)$

off-diagonal

a is indep of b.

$E[a_i b_i] = E[a_i] \cdot E[b_i]$ $=0$

Case 2: $m_i = m_j$ $(a = b)$

diagonal ele in $E[M^T M]$

$\boxed{E[a_i^2] = 1/k}$ $\Rightarrow E[a^T b] = 1$

$a_i \sim N(0, 1/k)$ $\quad var(a_i) = E a_i^2 - (E a_i)^2$

$1/k = E a_i^2 - 0$

In Hoeffding's

$$X = \sum_{i=1}^{n} X_i$$

Q1: X_i and ML

Q2: when $X = \sum_{i=1}^{n} X_i$

Model: (x, y)

$$y \simeq w \cdot x$$

collect $\{(x_i, y_i)\}_{i=1}^{n}$

Assume: $y \simeq w \cdot x$

$$\Rightarrow y \simeq f_w(x)$$

find w ~~f~~

$$\text{Split} \begin{cases} \text{Train } \{(x_i, y_i)\} \\ \text{Val } \{(x_i', y_i')\} \end{cases}$$

$\cancel{\emptyset}$

$$\min_w \sum_{i=1}^{n} \boxed{(y_i - w \cdot x_i)^2} + \lambda \cdot \|w\|_2^2$$

R.V.

λ	Val loss
0.01	10
0.1	100
0.2	0.1
⋮	⋮

test ≈ 0.1

all $(x, y) \sim D$

Fundamental assumption in learning theory.

Justify Val process:

Goal: minimize test loss

$$\text{Test loss} \overset{\Delta}{=} \boxed{\frac{1}{N}\sum_{i=1}^{N}(w \cdot \underline{x_i} - \underline{y_i})^2}$$

Train data

$N \to \infty$ $\qquad (\bar{x}_i, \bar{y}_i)$ test data

$$\underline{\text{Test loss}} \overset{\Delta}{=} \frac{1}{n}\sum_{i=1}^{n}(\bar{y}_i - \underline{w} \cdot \bar{x}_i)^2$$

$n \to \infty$

Q: $n = ?$

$n = 100$

100 piece data / day

day $1 \to \infty$

$\to 0$

$N(0, 1)$.

X_1, ... X_n.

$$\frac{1}{n}\sum X_i \xrightarrow[n\to\infty]{} 0 = E[X]$$

Central limit theorem / law of large numbers / ...

$$\frac{1}{n}\sum (y_i - w\cdot x_i)^2$$

$$\to \mathop{E}_{(x,y)\sim D}\left[(y - w\cdot x)^2\right]$$

$$\mathbb{1}\{E\} = \begin{cases} 1 & E \text{ happens} \\ 0 & \text{not happen} \end{cases}$$

loss of binary ~~a~~ classification

on (x, y)

$$\triangleq \mathbb{1}\{y \neq \text{sign}(\underline{w} \cdot x)\}$$

On $\{(x_i, y_i)\}_{i=1}^{n}$

$$\text{loss} = \frac{1}{n} \underbrace{\sum_{i=1}^{n} \mathbb{1}\{y_i \neq \text{Sign}(w \cdot x_i)\}}_{\#\text{ mistakes made by "}w\text{"}}$$

$$\text{Test loss} \rightarrow \mathop{E}_{(x,y)\sim D}\left[\mathbb{1}\{y \neq \text{sign}(w \cdot x)\}\right]$$

Claim: small val loss

$\Rightarrow$ small test loss

$$\underbrace{E[\,1\{y \neq \text{sign}(w\cdot x)\}\,]}_{E[f_w(x,y)]}$$

Val: $\frac{1}{m}\sum_{i=1}^{m} \underbrace{1\{y_i \neq \text{sign}(w\cdot x_i)\}}_{f_w(x_i,y_i)}$

Proof:

$|f_w(x,y)| \le 1$ a random draw of $f_w(x,y)$

H:

$$\Pr\left(\left|\sum f_w(x_i,y_i) - E[\Sigma]\right| \ge t\right) \le e^{-\frac{t^2}{n}}$$

$\Leftrightarrow$ w.p. $1 - e^{-t^2/n}$

$$\left|\underbrace{\frac{1}{m}\sum f_w(x_i,y_i)}_{\text{val loss}} - \frac{1}{m}E[\Sigma]\right| \le \boxed{t/m}$$

$$\frac{1}{m} \cdot \underline{E}\left[\sum f(x_i, y_i)\right]$$

$$= \frac{1}{m} \sum E\left[f(x_i, y_i)\right]$$

$$= \frac{1}{m} \sum_{i=1}^{m} E\left[f(x, y)\right]$$

$$= E\left[f(x, y)\right]$$

= Test loss

w.p. $1 - e^{-t^2/m}$ $\Rightarrow m = \square$

$$|\text{Val} - \text{Test}| \le \boxed{t}/m$$

0.1

$$\Rightarrow \quad t/m = 0.001$$

$$e^{-t^2/m} = 0.01$$

$$\Rightarrow \quad m = \square, \quad t = \square$$

Traing loss $= \frac{1}{n} \sum_{i=1}^{n} \boxed{f_w(x_i, y_i)}$

$n \to \infty$ $\downarrow$ Test loss

How to find w:

$w \leftarrow \arg\min \frac{1}{n} \sum_{i=1}^{n} f_w(x_i, y_i)$

$w = F(\{(x_i, y_i)\}_{i=1}^{n})$

$f_w(x_1, y_1)$ is dependent on $f_w(x_2, y_2)$

$F(\text{train set})$