Lecture 4. High dimensionality & Random Projection

Sep. 11, 2017

Recall

data $X = [x_1, \dots, x_n]^{p \times n}$

centered data $X_c = XH$, $H = I - \frac{1}{n} 1 1^T$, $1 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^p$

<u>k-SVD</u> $X_c \cong \hat{U}_k \hat{S}_k \hat{V}_k^T$ as best rank-k approximation of X_c

$\left.\begin{array}{l} \hat{U}_k \in \mathbb{R}^{p \times k} \\ \hat{V}_k \in \mathbb{R}^{n \times k} \end{array}\right\}$ orthogonal column mat.

$\hat{S}_k = \mathrm{diag}(\hat{\sigma}_1, \dots \hat{\sigma}_k)$, $\hat{\sigma}_1 \geq \hat{\sigma}_2 \geq \dots \geq \hat{\sigma}_k \geq 0$

"<u>k-PCA</u>" is given by $(\hat{U}_k, \hat{S}_k)$ with projection

$$\hat{\beta} = [\hat{\beta}_1, \dots, \hat{\beta}_n] = \hat{U}_k^T X_c \simeq \underline{\hat{S}_k \hat{V}_k^T}$$

each column gives new coordinates

$\Longleftrightarrow$ Eigenvalue Decomposition of Covariance Mat.

$$\hat{\Sigma}_n = \frac{1}{n} X_c X_c^T \cong \hat{U}_k \hat{\Lambda}_k \hat{U}_k^T, \quad \Lambda_k = S_k^2$$

"<u>k-MDS</u>" is given by $(\hat{S}_k, \hat{V}_k)$ with data representation

$$\underline{\hat{S}_k \hat{V}_k^T} \in \mathbb{R}^{k \times n}$$

$\Longleftrightarrow$ Eigenvalue Decomp. of Kernel Mat.

$$\hat{K} = \frac{1}{n} X_c^T X_c \cong \hat{V}_k \hat{\Lambda}_k \hat{V}_k^T$$

"<u>kernel-PCA/MDS</u>" $K \succeq 0$ is positive semidefinite

$$B = HKH^T \cong \hat{V}_k \hat{\Lambda}_k \hat{V}_k^T$$

Problem: What about big data & high dimensionality?
$n \gg 1$ $\qquad$ $p \gg 1$

big data $n \gg 1$. $\hat{\Sigma}_n = \frac{1}{n} \sum_i (x_i - \hat{\mu})(x_i - \hat{\mu})^T$

down sample $n' \to$ good approximation of $\hat{\Sigma}_n$ & $\hat{U}_k, \hat{S}_k$

$\hat{V}_k$ restricted on subsample

high dimensionality $p \gg 1$, $\hat{\Sigma}_n \in \mathbb{R}^{p \times p}$ too big to compute

$K = X_c^T X_c$?

easy to approximate?

Random Projection!

e.g. $R = \frac{1}{\sqrt{d}} A^{d \times p}$ where $A_{ij} \sim \mathcal{N}(0, 1)$.

$$X_c^{p \times n} \mapsto (R X_c)^{d \times n}, \quad d \ll p.$$

$K_R = X_c^T R^T R X_c$ a good approximation of K!

- $A_{ij} = \begin{cases} 1 & p = \frac{1}{2} \\ -1 & (1-p) = \frac{1}{2} \end{cases}$

- $A_{ij} = \begin{cases} 1 & p = \frac{1}{6} \\ 0 & 1-2p = 2/3 \\ -1 & p = 1/6 \end{cases}$ sparse with many zeros!

Examp (Human Genome Diversity Project / HGDP)

http://www.cephb.fr/en/hgdp_panel.php

$n = 1064$ persons $\quad p = 644{,}258$ SNPs

$X^{p \times n}$: $X_{ij} = 0$: "AA"; 1: "AC"; 2: "CC"; 9: "Missing"

Removing 21 persons with missing values.

$$X^{644{,}258 \times 1043}$$

$R^{d \times p}$ randomly select d rows/SNPs of X.

$$\tilde{X}^{d \times n} = R X_c = R X H \cong \underset{d \times k}{\hat{U}_k} \underset{k \times k}{\hat{S}_k} \underset{k \times n}{\hat{V}_k^T}$$

$d = p = 600k \qquad d = 5k \qquad d = 100k.$

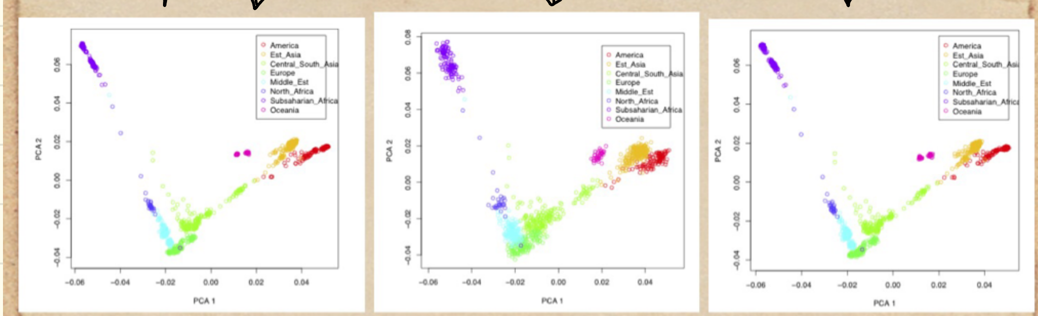

FIGURE 1. (Left) Projection of 1043 individuals on the top 2 MDS principal components. (Middle) MDS computed from 5,000 random projections. (Right) MDS computed from 100,000 random projections. Pictures are due to Qing Wang.

In all cases, $(\hat{S}_{k,d}, \hat{V}_{k,d})$ are good results!

Here: PCA coordinates $\hat{S}_{k,d} \hat{V}_{k,d}^T \in \mathbb{R}^{k \times n}$, $\quad k = 2$.

Why does it work?

Johnson-Lindenstrauss Lemma

Idea: $x_i \in \mathbb{R}^p$, $d_{ij} = \|x_i - x_j\|$, $i=1,\dots,n$

Look for a transform $f: X_i \mapsto Y_i \in \mathbb{R}^d$, $d = O(c(\varepsilon)\log n)$

s.t.
$$1-\varepsilon \le \frac{\|Y_i - Y_j\|}{\|X_i - X_j\|} \le 1+\varepsilon.$$
with probability $\ge 1 - n^{\alpha}$, $\alpha > 0$ $\xrightarrow[n\to\infty]{} 0$

Uniform ε-Isometry!

relative metric-distortion is uniformly bounded by ε!

f is a random projection!

1980s Johnson-Lindenstrauss Lipschitz Extension

2003 Sanjoy Dasgupta, Anupam Gupta

Dimitris Achlioptas.

Computer Science, data compression

nearest neighbor search

Thm Given $\varepsilon \in (0,1)$, $n, \alpha > 0$.

Let
$$k = c(\alpha, \varepsilon)\log n = (4+2\alpha)\left(\frac{\varepsilon^2}{2} - \frac{\varepsilon^3}{3}\right)^{-1}\log n$$

Then for any n points $x_i \in \mathbb{R}^D$ $(i=1,\dots,n)$, there exists a map $f: \mathbb{R}^D \to \mathbb{R}^k$ s.t. $\forall x_i, x_j$

$$1-\varepsilon \le \frac{\|f(x_i) - f(x_j)\|^2}{\|x_i - x_j\|^2} \le 1+\varepsilon \qquad (*)$$

with prob $\ge 1 - n^{\alpha}$.

- $(*)$ holds with probability at least $1-n^{\alpha}$.
- f can be found in randomized polynomial time

(random projections)

e.g. $f(x) = Rx$, $R = [r_1, \cdots, r_K]^T \in \mathbb{R}^{K \times D}$, $x \in \mathbb{R}^D$

$r_i \in S^{D-1}$ sphere of $D-1$ dim.

e.g. $r_i = \dfrac{(a_1^i \cdots a_D^i)}{\| a^i \|}$, $a_j^i \sim N(0,1)$