Dimensionality Reduction

L16: Random Projections

Jeff M. Phillips

March 26, 2018

Peer Feedback : Group Grouping

15 minutes for each group

- 5-10 minutes, group explain to others about project/progress
- 5-10 minutes feedback

1, 7, 22	2, 5, 32	3, 17, 33
4, 11	6, 10, 21	8, 19, 37
9, 39, 40	12, 35, 41	16, 20, 31
18, 24, 38	25, 29, 36	15, 26, 30
13, 23, 27	28, 34, 42	

Data Set $A = \{a_1, a_2, \dots a_n\}$ Note $A \in \mathbb{R}^{d \times n}$

$a_i \in \mathbb{R}^d$

Map $A \to B \subset \mathbb{R}^{d'}$ $d' \ll d$

Previous PCA found best rank-k apx A

by using SVD $[U, S, V^T] = A$

right sing. vectors $V = [V_1, V_2, V_3, \dots V_k \dots V_d]$

$b_i = (\langle v_1, a_i \rangle, \langle v_2, a_i \rangle, \dots \langle v_k, a_i \rangle) \in \mathbb{R}^k$

minimized $\sum_{i=1}^{n} \| a_i - \pi_{V_k}(a_i) \|^2$ $b_i = \pi_{V_k}(a_i)$

JL Objective

(Johnson-Linden Strauss 1985)

for all pairs $a, a' \in A \subset \mathbb{R}^d$

find (linear) mapping $u : \mathbb{R}^d \to \mathbb{R}^k$

so

$$(1-\varepsilon)\,\|a - a'\| \le \|u(a) - u(a')\| \le (1+\varepsilon)\,\|a - a'\|$$

for error parameter $\varepsilon \in (0, 1/2)$

$\varepsilon = \frac{1}{10}$ or $\frac{1}{100}$

$\varepsilon = \frac{1}{10}$ $\quad \|a - a'\| = 20$

then $\|u(a) - u(a')\| \in [18, 22]$

dimension k needed

is $O\left(\frac{1}{\varepsilon^2} \log \frac{n}{\delta}\right)$ w.p. $1-\delta$

How do create mapping $u : \mathbb{R}^d \to \mathbb{R}^p$

Random Projection

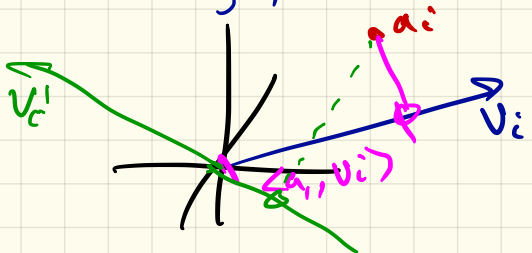
$\underline{u_i : \mathbb{R}^d \to \mathbb{R}^k}$ $\qquad u(a) = (u_1(a), u_2(a), \ldots)$

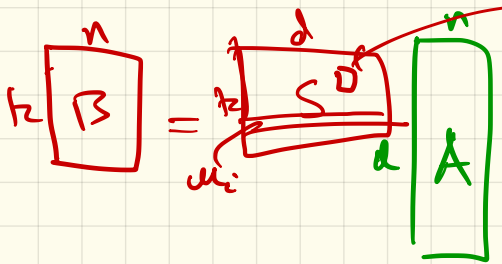
Create $v_i \in d\text{-Gaussian}(0, I)$

↑ — Box-Mueller Transform $[d\ [0,1]]$

Normalize $v_i = \frac{\sqrt{d}}{\sqrt{k}} \frac{v_i}{\|v_i\|}$

Now $u_i(a) = \langle a_i, v_i \rangle = \sum_{j=1}^{d} a_{ij} \cdot v_{ij}$

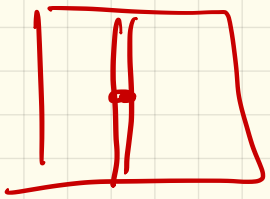




$S_{ij} \sim \text{Gaussian}(0, 1)$

$S_{ij} = \text{Uniform}\{0, -1, +1\}$

S



{all 0s except one 1 at random

User Guide

when $d \geq 10,000$

$k \geq 500$