DS Assignment 2

Rohit Patil — 17110126

March 18, 2020

1 Question 1

Colab Notebook Link : link

	EM method	K-Means ++
error for dataset 0 is	1.4960849841587072	0.5455393578370766
error for dataset 1 is	1.4150054792113034	0.8786377312912221
error for dataset 2 is	1.323143370781267	0.6411169195844564
error for dataset 3 is	1.2345612601486615	0.7526543380792334
error for dataset 4 is	1.0968649995115272	1.0841567916366697
error for dataset 5 is	2.469538812882015	1.2168355419086527
error for dataset 6 is	3.5471508682996724	1.9424907887065643

2 Question 2 [3.11]

let A be the matrix with right singular vectors $v_1, v_2, ...v_r$ and $\sigma_1, \sigma_2, ..\sigma_r$ are the corresponding singular values

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T \tag{1}$$

$$\therefore A^T = \sum_{i=1}^{r} \sigma_i v_i u_i^T \tag{2}$$

$$A^T A = \sum_{i=1}^{r} \sigma_i^2 v_i u_i^T u_i v_i^T \tag{3}$$

$$= \sum_{i=1}^{r} \sigma_i^2 v_i v_i^T \text{.....................} (u_i^T u_i = 1) \tag{4}$$

mulitiplying by v_j to both sides

$$A^T A v_j = (\sum_{i=1}^{r} \sigma_i^2 v_i v_i^T) v_j \tag{5}$$

now, $v_i^T v_j = 0$ if $i \neq j$ and $v_i^T v_j = 1$ if $i = j$
$\therefore$

$$A^T A v_j = \sum_{i=1}^{r} \sigma_i^2 v_j$$

Hence $v_1, v_2, v_3, ...v_r$ are eigenvectors of $A^T A$ and $\sigma_1^2, \sigma_2^2, ...\sigma_r^2$ are corresponding eigenvalues

The eigenvectors of $A^T A$ are unique up to multiplicative constant, i,e there are distinct eigenvalues. Singular values of A are also distinct as they are the square root of these eigenvalues(non-negative real). Hence all σs are distinct, hence we can infer that right singular vector v_1 corresponding to first singular value is unique upto sign, hence u_1 is unique upto sign. Now we incur other singular vectors in space orthogonal to v_1 which is defined as v_1 is upto sign. Similary we can do above procedure for every v_i, Hence the singular vectors are unique upto sign.

3 Question 3 [3.12]

(a) $\|A_k\|_F^2$

The rows of A_k are the projections of the rows of A onto the subspace V_k spanned by the first k singular vectors of A.

As proved in the class, frobenius norm is sum of squares of 2 norms of rows. i.e., sum of squares of first k singular values

Hence $\|A_k\|_F^2 = \sum_1^k \sigma_i^2$

(b) $\|A_k\|_2^2 = \sigma_1^2$

Spectral norm is the highest singular value, i.e. σ_1^2

$A = \sum_1^r \sigma_i u_i v_i^T$

let v be top singular vector of A, expressing v as linear combination of $v_1, v_2, ..v_r$ i.e $v = \sum_1^r c_j v_j$,
Then, $|(A)v| = \sqrt{\sum_{i=1}^r c_i^2 \sigma_i^2}$....ref(BHK)

since the u_i are orthonormal. the last quantity is maximized by $c_1 = 1$ and the rest of the c_i are zero ,hence $\|A\|_2^2 = \sigma_1^2$

(c) $\|A - A_k\|_F^2$

$$\|A - A_k\|_F^2 = \left\| \sum_{k+1}^{r} \sigma_i u_i v_i^T \right\|_F^2 \tag{6}$$

$$\tag{7}$$

now, we know

$$A - A_k = \sum_{k+1}^{r} \sigma_i u_i v_i^T \tag{8}$$

$$A - A_k = U(\Sigma - \Sigma_k)V^T \tag{9}$$

now, if U is orthogonal matrix:

$$\|A\| = \|AU\| = \|UA\|$$

hence,

$$\|A - A_k\|_F^2 = \left\|U^T(A - A_k)V\right\|_F^2 = \left\|U^T U(\Sigma - \Sigma_k)V^T V\right\|_F^2$$

now, $U^T U = V^T V = 1$ as they are orthogonal,
$\therefore \|A - A_k\|_F^2 = \|Diag(0, 0, 0, 0 ... \sigma_{k+1,..\sigma_r})\|_F^2 = \sum_{k+1}^{r} \sigma_i^2$

(d) $\|A - A_k\|_2^2$
$\|A - A_k\|_F^2 = \sigma_{k+1}^2$
proof:
$A - A_k = \sum_{k+1}^{r} \sigma_i u_i v_i^T$
let v be top singular vector of $A - A_k$, expressing v as linear combination of $v_1, v_2, ..v_r$ i.e $v = \sum_1^r c_j v_j$,
Then, $|(A - A_k)v| = \sqrt{\sum_{i=k+1}^{r} c_i^2 \sigma_i^2}$....ref(BHK page no: 50)
since the u_i are orthonormal. the last quantity is maximized by $c_1 = 1$ and the rest of the c_i are zero ,hence $\|A - A_k\|_2^2 = \sigma_{k+1}^2$

4 Question 4 [3.26]

(a) maximizing sum of squared similarities of vector v with each of document d_i, i.e
$max(\sum_1^m (d_i^T.v)^2) = \max_v \|Av\|_F^2$

$\therefore v$ -> first right singular vector of A

(b) The centre of Gravity or centroid minimizes the sum of euclidian distances. let v be the C.O.G So, $v = \frac{1}{d}\sum_1^m d_i$
i.e. minimize w.r.t. v

$$= \sum_1^m (d_i - v)^T.(d_i - v) \tag{10}$$

$$= \sum_1^m (2 - 2d_i^T.v) \tag{11}$$

which means maximizing, $\sum_1^m (d_i^T.v)$ which is different from synthetic document because we were maximizing $\sum_1^m (d_i^T.v)^2$ for the synthetic document.

(c) We have to find k synthetic documents such that
$\|Av\|_F^2$ is maximized for each $v_i \forall i \in [1,k]$ and all v_i are orthogonal hence, if $k \leq rank(A)$, v_i 's are first k right singular vectors of A.

(d) Hint: the right-singular vectors of A are eigenvectors of $A^T.A$.

To Prove: For any block diagonal matrix M, the eigenvectors are concatenations of individual eigenvectors of blocks.

let v be eigenvector of M

let the first block be $k \times k$, hence the eigenvector is $k \times 1$

concatenate a vector of [0,0,0,...,0] to v, hence v looks like $[1,1,....1,0,0,...,0]$
Here we can see, v is an eigenvector of M let v = $\sum v_j$ where j corresponds to j'th block of M,

hence $i'th$ block of v is eigenvector to $i'th$ block of M, and v is eigenvector of M

Now, obviously $A^T A$ is block diagonal as A is block diagonal and A^T is block diagonal, and each row in $A^T A$ is normalized.

By the previous claim, any block diagonal matrix $A^T A$ has the property that blocks of its eigenvector are eigenvectors of individual blocks of $A^T A$

Now, we know eigenvectors of $P^T P$ are same as right singular vectors of P Hence, eigenvectors(of $A^T A$)'s block are right singular vectors of corresponding block of A

(e) From 2nd part, we obtained the synthetic documents,i.e by taking the svd, we find k right singular vectors. for document d, take dot product with each of the singular vectors, and include in j'th cluster where $d^T v_j$ is maximum(similarity). If two or many vectors have same similarity with document d, include in the cluster with larger singukar value.

5 Question 5 [3.28]

Colab Notebook Link : link

Collaborators: Harshil Jain, Anubhav Jain