IE531: Algorithms for Data Analytics
Spring, 2023
Homework 3: SVD and Related Topics
Due Date: 3 March 2023

Instructions

1. You will submit a PDF-version of your answers on Canvas on-or-before midnight of the due date.

Instructions

1. (*40 points*) Suppose a matrix $n \times d$ matrix $\mathbf{A}$ has an SVD decomposition that can be written as

$$\underbrace{\mathbf{A}}_{n\times d} = \left(\underbrace{\mathbf{U}_1}_{n\times r_1} \times \overbrace{\mathbf{\Sigma}_1}^{r_1\times r_1} \times \underbrace{\mathbf{V}_1^T}_{r_1\times d}\right) + \left(\underbrace{\mathbf{U}_2}_{n\times r_2} \times \overbrace{\mathbf{\Sigma}_2}^{r_2\times r_2} \times \underbrace{\mathbf{V}_2^T}_{r_2\times d}\right)$$

where the singular-values in $\mathbf{\Sigma_1}$ (resp. $\mathbf{\Sigma_2}$) are greater than (resp. lesser than) some $\gamma \in \mathcal{R}$. Show that

(a) (*20 points*)
$$\mathbf{U}_1^T \times \mathbf{U}_2 \times \mathbf{\Sigma_2} \times \mathbf{V}_2^T = \underbrace{\mathbf{0}}_{r_1\times d}, \text{ and}$$

(b) (*20 points*)
$$\mathbf{U}_2 \times \mathbf{\Sigma_2} \times \mathbf{V}_2^T \times \mathbf{V}_1 = \underbrace{\mathbf{0}}_{d\times r_1}$$

This follows from the structure of the SVD algorithm covered in class. That is, $\mathbf{U}_1$ (resp. $\mathbf{U}_2$) is a collection of r_1-many (resp. r_2-many) orthogonal, unit-vectors. That is,

$$\mathbf{U}_1^T\mathbf{U}_2 = \underbrace{\mathbf{0}}_{r_1\times r_2} \Leftrightarrow \left(\mathbf{U}_2^T\mathbf{U}_1 = \underbrace{\mathbf{0}}_{r_2\times r_1}\right)$$

That is, if $\mathbf{A} = \mathbf{U\Sigma V}^T$, then $\mathbf{U} = (\widetilde{\mathbf{U}}_1 \mid \widetilde{\mathbf{U}}_2)$ and $\widetilde{\mathbf{V}} = (\widetilde{\mathbf{V}}_1 \mid \widetilde{\mathbf{V}}_2)$, where $\mathbf{U}_1 = (\widetilde{\mathbf{U}}_1 \mid \mathbf{0})$ (resp. $\mathbf{U}_2 = (\mathbf{0} \mid \widetilde{\mathbf{U}}_2)$) and $\mathbf{V}_1 = (\widetilde{\mathbf{V}}_1 \mid \mathbf{0})$ resp. $\mathbf{V}_2 = (\mathbf{0} \mid \widetilde{\mathbf{V}}_2)$) . A similar argument establishes Problem 1b, and you conclude $\mathbf{V}_1$ and $\mathbf{V}_2$ are orthogonal to each other.

2. (*40 points*) Suppose a matrix $n \times d$ matrix $\mathbf{A}$ with rank r, and has an SVD

$$\underbrace{\mathbf{A}}_{n\times d} = \underbrace{\mathbf{U}}_{n\times r} \times \overbrace{\mathbf{\Sigma}}^{r\times r} \times \underbrace{\mathbf{V}^T}_{r\times d}.$$

Let us suppose $\mathbf{A}$ gets "corrupted" by a $n \times d$, *noise-matrix* $\mathbf{E}$, and $\mathbf{A}_1 = \mathbf{A} + \mathbf{E}$. Suppose the corrupted-matrix $\mathbf{A}_1$ has an SVD

$$\underbrace{\mathbf{A}_1}_{n\times d} = \underbrace{\mathbf{U}_1}_{n\times r_1} \times \overbrace{\mathbf{\Sigma}_1}^{r_1\times r_1} \times \underbrace{\mathbf{V}_1^T}_{r_1\times d}.$$

Suppose $\widehat{\mathbf{\Sigma}}_1$ is obtained from $\mathbf{\Sigma}_1$ by keeping only the top r-many entries (i.e. we zero-out all diagonal-values that not in the list of top r-many SVs). Let $\widehat{\mathbf{A}} = \mathbf{U}_1 \times \widehat{\mathbf{\Sigma}}_1 \times \mathbf{V}_1^T$. Show that

$$||\widehat{\mathbf{A}} - \mathbf{A}||_F \leq \sqrt{8r} \times ||\mathbf{E}||_2.$$

See this link for a proof of this claim.

3. (*20 points*) Let $\mathbf{A}$ be an $n \times d$ matrix (of real numbers) that can be partitioned as

$$\underbrace{\mathbf{A}}_{n\times d} = \begin{pmatrix} \underbrace{\mathbf{A}_1}_{n_1\times d_1} & \underbrace{\mathbf{A}_2}_{n_1\times d_2} \\ \underbrace{\mathbf{A}_3}_{n_2\times d_1} & \underbrace{\mathbf{A}_4}_{n_2\times d_2} \end{pmatrix},$$

where $n = n_1 + n_2$ and $d = d_1 + d_2$ (obviously).

(a) (*10 points*) Show that

$$rank(\mathbf{A}) \leq rank(\mathbf{A}_1) + rank(\mathbf{A}_2) + rank(\mathbf{A}_3) + rank(\mathbf{A}_4)$$

Consider the SVD of each $\mathbf{A}_i$, where $\mathbf{A}_i = \mathbf{U}_i\mathbf{\Sigma}_i\mathbf{V}_i^T$, then it follows that

$$\mathbf{A} = \underbrace{\begin{pmatrix} \mathbf{U}_1 & \mathbf{0} & \mathbf{U}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_3 & \mathbf{0} & \mathbf{U}_4 \end{pmatrix}}_{\mathbf{X}} \underbrace{\begin{pmatrix} \mathbf{\Sigma}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Sigma}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Sigma}_4 \end{pmatrix}}_{\mathbf{\Sigma}} \underbrace{\begin{pmatrix} \mathbf{V}_1^T & \mathbf{0} \\ \mathbf{V}_3^T & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2^T \\ \mathbf{0} & \mathbf{V}_4^T \end{pmatrix}}_{\mathbf{Y}^T} \quad (1)$$

We know that $rank(\mathbf{AB}) \leq \min(rank(\mathbf{A}), rank(\mathbf{B}))$[1]. The result follows from the fact that $rank(\mathbf{\Sigma}) = rank(\mathbf{A}_1) + rank(\mathbf{A}_2) + rank(\mathbf{A}_3) + rank(\mathbf{A}_3)$.

(b) (*10 points*) Suppose for each $\{\mathbf{A}_i\}_{i=1}^4$ there is a corresponding set of matrices $\{\mathbf{B}_i\}_{i=1}^4$ such that $\forall i, ||\mathbf{A}_i - \mathbf{B}_i||_F \leq \epsilon$ show that

$$\left\|\mathbf{A} - \begin{pmatrix} \mathbf{B}_1 & \mathbf{B}_2 \\ \mathbf{B}_3 & \mathbf{B}_4 \end{pmatrix}\right\|_F \leq 4\epsilon$$

From equation 1, we note that

[1] Let $\mathbf{A}$ (resp. $\mathbf{B}$) be a $n \times d$ (resp. $d \times m$) matrix. The product $\mathbf{Ax}$ for any $d \times 1$ column vector $\mathbf{x}$ is an $n \times 1$ column vector that can be expressed in terms of the columns of $\mathbf{A}$. Therefore, m-many all columns of $\mathbf{AB}$ can be expressed in terms of the columns of $\mathbf{A}$ (i.e. there are no columns that could not be expressed in terms of columns of $\mathbf{A}$, but now can be expressed in terms of the columns of $\mathbf{AB}$). Therefore $rank(\mathbf{AB}) \leq rank(\mathbf{A})$. If we interpreted $\mathbf{AB}$ as n-many, d-dimensional rows, we get that all rows of $\mathbf{AB}$ can be expressed in terms of the rows of $\mathbf{B}$. This would mean $rank(\mathbf{AB}) \leq rank(\mathbf{B})$, as well.

i. $\mathbf{X}$ is a $(n_1 + n_2) \times (rank(\mathbf{A}_1) + rank(\mathbf{A}_2) + rank(\mathbf{A}_3) + rank(\mathbf{A}_4))$ matrix.

ii. $\mathbf{Y}$ is a $(rank(\mathbf{A}_1) + rank(\mathbf{A}_2) + rank(\mathbf{A}_3) + rank(\mathbf{A}_4)) \times (d_1 + d_2)$ matrix.

iii. $\mathbf{X}$ not orthogonal, in general– more specifically, assuming we have a lot data points (i.e. $n \gg d$), there is no guarantee that $\mathbf{X}^T\mathbf{X}$ is diagonal.

iv. $\mathbf{Y}$ not orthogonal, in general– more specifically, assuming we have a lot data points (i.e. $n \gg d$), there is no guarantee that $\mathbf{Y}\mathbf{Y}^T$ is diagonal.

Let $\mathbf{X} = \mathbf{Q}_x\mathbf{R}_x$ and $\mathbf{Y} = \mathbf{Q}_y\mathbf{R}_y$ be the QR-Decomposition of $\mathbf{X}$ and $\mathbf{Y}$, respectively. It follows that

$$\mathbf{A} = \mathbf{Q}_x \underbrace{\mathbf{R}_x\mathbf{\Sigma}\mathbf{R}_y^T}_{\mathbf{W}=\mathbf{U}_w\mathbf{\Sigma}\mathbf{V}_w^T} \mathbf{Q}_y^t = \underbrace{\mathbf{Q}_x\mathbf{U}_w}_{\mathbf{U}} \mathbf{\Sigma} \underbrace{\mathbf{V}_w^T\mathbf{Q}_y^t}_{\mathbf{V}^T} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

We might have to “move-around” a few columns of $\mathbf{U}$ (and the corresponding rows of $\mathbf{V}^T$) as the singular-values listed in $\mathbf{\Sigma}$ might not necessarily be in increasing order. Regardless, if we truncated/dropped some of these elements out to get within an ϵ-approximation of each of the $\mathbf{A}_i$ matrices, we will not be making a error-in-approximation that exceeds 4ϵ in using these approximations to construct an approximation for $\mathbf{A}$.