EECS 16B   Designing Information Devices and Systems II
Fall 2021   Discussion Worksheet   Discussion 11A

The following notes are useful for this discussion: Note 16

1. **Geometric interpretation of the SVD**

   In this exercise, we explore the geometric interpretation of matrix transformations and how this connects to the SVD. We consider how a real 2×2 matrix acts on the unit circle, transforming it into an ellipse. It turns out that the principal semiaxes of the resulting ellipse are related to the singular values of the matrix, as well as the vectors in the SVD.

   (a) Consider the real 2×2 matrix

   $$A = \begin{bmatrix} 0 & -1 \\ 3 & 0 \end{bmatrix}. \tag{1}$$

   Also consider the unit circle in $\mathbb{R}^2$,

   $$S = \left\{ \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \middle| 0 \leq \theta < 2\pi \right\}. \tag{2}$$

   Plot the transformed circle, AS, on the $\mathbb{R}^2$ plane.

(b) Now let's consider how this transformation looks in the lens of the SVD. The SVD for matrix A is:

$$A = U\Sigma V^\top = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{3}$$

$$A\vec{x} = U\Sigma V^\top \vec{x} = U\left(\Sigma\left(V^\top \vec{x}\right)\right). \tag{4}$$

Let's start by examining the effects of each of these matrices one at a time, right to left, in the same order that they would be applied to a vector $\vec{x}$.

What does the unit circle look like after being transformed by just $V^\top$? Plot $S_1 = V^\top S$ on the $\mathbb{R}^2$ plane. Geometrically speaking, what does $V^\top$ do to any given $\vec{x}$?

(c) **What does the unit circle look like after being transformed by $\Sigma V^\top$? Plot $S_2 = \Sigma V^\top S$ on the $\mathbb{R}^2$ plane. Geometrically speaking, what is the Σ matrix doing to any given $V^\top \vec{x}$?**

(d) **What does the unit circle look like after being transformed by $U\Sigma V^\top$? Plot $S_3 = U\Sigma V^\top S$ on the $\mathbb{R}^2$ plane. Geometrically speaking, what is the U matrix doing to any given $\Sigma V^\top \vec{x}$?**

(e) Consider the columns of the matrices U, V from the SVD of A in part (b), and treat them as vectors in $\mathbb{R}^2$. Let $U = (\vec{u_1} \ \vec{u_2})$, $V = (\vec{v_1} \ \vec{v_2})$. Let σ_1, σ_2 be the singular values of A, where $\sigma_1 \geq \sigma_2$.

In your plot of AS, draw the vectors $\sigma_1 \vec{u_1}$ and $\sigma_2 \vec{u_2}$ from the origin. What do these vectors correspond to geometrically?

(f) **Repeat parts (b-e) for the following matrices, and note down any interesting things you notice.**

i. A 3D matrix, $X = \begin{bmatrix} 1 & 2 & -1 \\ 1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$

ii. A rotation matrix, $A_1 = \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}$.

iii. A diagonal matrix, $A_2 = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix}$.

iv. A symmetric matrix, $A_3 = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$.

v. A matrix with non-trivial nullspace, $A_4 = \begin{bmatrix} 4 & 2 \\ -2 & -1 \end{bmatrix}$.

vi. An arbitrary matrix, $A_5 = \begin{bmatrix} 1.6 & 2.4 \\ -0.4 & -1 \end{bmatrix}$.

2. PCA on Artificial Data

In this example we'll create a synthetic dataset and walk through how SVD reveals the "most important" directions for PCA. As an overview, in this subpart,

- we generate four clusters of data points on a plane
- combine these clusters, and embed the dataset randomly in higher dimensional space
- add quite a lot of noise
- try to use PCA to separate the data into the original clusters.

We'll use `dis11A_pca.ipynb` jupyter notebook for this discussion.

(a) Let's generate the clusters of points around different means μ_i for $i \in \{1, 2, 3, 4\}$. Here, generating a cluster corresponds to sampling from a Gaussian probability distribution, where the probability of sampling x follows

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\|x - \mu_i\|_2^2}{2\sigma^2}} \tag{5}$$

The details of this probability are not important, nor is it even important that you understand what probability means in this context. All that matters is that this is a way of randomly throwing darts at these mean vectors in a manner that makes us miss by a little more than we miss by a lot.

Generate points corresponding to the 4 clusters centered at $\{(-5, -5), (-5, 5), (5, -5), (5, 5)\}$ **with standard deviation** $\sigma = 1$. **Merge into a single dataset** $D = \begin{bmatrix} x_1 & x_2 & \cdots & x_{4N} \end{bmatrix}$. **Scatter plot the dataset.** The function `cluster(mean, stddev, N)` generates a dataset of N such points for 1 cluster.

(b) Any vector $\vec{x} \in \mathbb{R}^m$, can be *embedded* onto a higher dimensional space $\mathbb{R}^n$, with $n > m$ using $n \times m$ matrix M such that

$$\vec{x}_{\text{embed}} = M\vec{x} \tag{6}$$

Therefore for a dataset D, we have $D_{\text{embed}} = MD$. **Embed the 2D-dataset from (a) into a 3D-space using** $M = \begin{bmatrix} 1 & 0 \\ 1 & -1 \\ 1 & 0 \end{bmatrix}_{(3 \times 2)}$

(c) Instead of 3D, we now embed our dataset to a 25D dimensional space, using a random 25×2 matrix M. We can still visualize this embedding in 2D/3D by looking at slices of our dataset. **Generate a** 25×2 **random matrix M (of rank 2), embed the dataset D as** $D_{\text{clean}} = MD$ **and visualize the dataset in 2D, 3D. Are the cluster boundaries still well defined?**

(d) We're now ready to add noise to the dataset. To this effect we sample $4N$ points D_{noise} from a distribution with mean $\mu = 0$, standard-deviation $\sigma = 5$

$$D_{\text{noisy}} = D_{\text{clean}} + D_{\text{noise}} \tag{7}$$

Plot the noisy dataset similar to (c). Are the cluster boundaries still well defined?

(e) Singular Value Decomposition is a useful tool for recovering the original clustering from the newly constructed dataset D_{noisy}. In particular, we perform SVD and project the noisy dataset over the columns of U that correspond to the singular values with the highest magnitude. In other words, we find the projection

$$U\Sigma V^\top = SVD(D_{\text{noisy}}) \tag{8}$$

$$P = U[: \#\text{components}] \tag{9}$$

$$D_{\text{PCA}} = P^\top D_{\text{noisy}} \tag{10}$$

This algorithm is referred to as Principal Component Analysis, and vectors in P are called the principal components. **Perform SVD on the dataset D_{noisy}. Using two principal components, project both the datasets $D_{\text{noisy}}, D_{\text{clean}}$ onto the above components. Are the cluster boundaries well defined?**

(f) **Perform SVD on the dataset D_{noisy}. Using three principal components, project both the datasets $D_{\text{noisy}}, D_{\text{clean}}$ onto the above components. Are the cluster boundaries well defined?**

(g) **The singular values are ordered such that $\sigma_1 \geq \sigma_2...\sigma_n$. Instead of components corresponding to the two largest singular values, use components corresponding to σ_3, σ_4. Repeat the same for using $\sigma_3, \sigma_4, \sigma_5$**

(h) **Plot all the singular values for the matrix D_{noisy}. Do you observe something about the distribution of the singular values?**

Contributors:

- Lynn Chua.

- Shane Barratt.

- Kuan-Yun Lee.

- Anant Sahai.

- Kareem Ahmad.

- Kumar Krishna Agrawal.

Discussion 11A, © UCB EECS 16B, Fall 2021. 7