Recitation 6

Carles Domingo

Matrices as maps vs. data

Previously in the course,

- Matrices are *linear tranformations* that act on vectors.

In PCA,

- Matrices as *data matrix*, where rows are instances of data and columns are features.

Remark the two different interpretations of matrices!

Recap questions on PCA

Let $x_1, ..., x_n \in \mathbb{R}^d$. You want to represent these data points in $k < d$ dimensions.

1. Explain how to do this using PCA.
2. How can you implement PCA using SVD?
3. How do we determine an ‘optimal’ value for k?

Recap questions on PCA

Let $x_1, ..., x_n \in \mathbb{R}^d$. You want to represent these data points in $k < d$ dimensions.

1. Explain how to do this using PCA.
2. How can you implement PCA using SVD?
3. How do we determine an 'optimal' value for k?

Recap questions on PCA

Let $x_1, ..., x_n \in \mathbb{R}^d$. You want to represent these data points in $k < d$ dimensions.

1. Explain how to do this using PCA.
2. How can you implement PCA using SVD?
3. How do we determine an ‘optimal’ value for k?

Recap questions on PCA

Let $x_1, ..., x_n \in \mathbb{R}^d$. You want to represent these data points in $k < d$ dimensions.

1. Explain how to do this using PCA.
2. How can you implement PCA using SVD?
3. How do we determine an 'optimal' value for k?

Review: SVD

Theorem

Let $A \in \mathbb{R}^{n\times m}$. Then there exists two orthogonal matrices $U \in R^{n\times n}$ and $V \in \mathbb{R}^{m\times m}$ and a matrix $\Sigma \in \mathbb{R}^{n\times m}$ such that $\Sigma_{1,1} \geq \Sigma_{2,2} \geq \cdots \geq 0$ and $\Sigma_{i,j} = 0$ for $i \neq j,\ A = U\Sigma V^\top$. The columns $u_1, ..., u_n$ of U (respectively the columns $v_1, \ldots, v_m$ of V) are called the left (resp. right) singular vectors of A. The non-negative numbers $\Sigma_{i,i}$ are the singular values of A. Moreover rank$(A) = \#\{i|\Sigma_{i,i} \neq 0\}$.

Reminder of Courant-Fisher

Theorem (Courant-Fisher principle)

Let A be a $n \times n$ symmetric matrix and let $\lambda_1 \geq \cdots \geq \lambda_n$ be its n eigenvalues and $v_1, \ldots, v_n$ be an associated orthonormal family of eigenvectors. Then

$$v_1 = \arg\max_{\|v\|=1} v^\top Av, \text{ and } \forall k = [2:n],\ v_k = \arg\max_{\|v\|=1, v\perp v_1,\ldots,v_{k-1}} v^\top Av.$$

$$v_n = \arg\min_{\|v\|=1} v^\top Av,\ \forall k = [1:(n-1)],\ v_k = \arg\min_{\|v\|=1, v\perp v_{k+1},\ldots,v_n} v^\top Av.$$

Courant-Fisher & SVD

Theorem (Corollary of Courant-Fisher principle)

Let A be a $n \times m$ matrix and let $A = U\Sigma V^\top$ be a singular eigenvalue decomposition of A. Then

$$v_1 = \arg\max_{\|v\|=1} \|Av\|,\ \sigma_1 = \max_{\|v\|=1} \|Av\|,$$

$$\forall k = [2:n],\ v_k = \arg\max_{\|v\|=1, v\perp v_1,\dots,v_{k-1}} \|Av\|, \sigma_k = \max_{\|v\|=1, v\perp v_1,\dots,v_{k-1}} \|Av\|$$

$$v_n = \arg\min_{\|v\|=1} \|Av\|,\ \sigma_n = \min_{\|v\|=1} \|Av\|,$$

$$\forall k = [1:(n-1)],\ v_k = \arg\min_{\|v\|=1, v\perp v_{k+1},\dots,v_n} \|Av\|,$$

$$\sigma_k = \min_{\|v\|=1, v\perp v_{k+1},\dots,v_n} \|Av\|$$

Question: Courant-Fisher & SVD

Suppose that we are given data $\{(x_i, y_i)\}_{i=1}^n$, and we hypothesize that the data approximately satisfy an affine relation of the form $ax_i + by_i = c$, where $(a, b, c) \neq 0$. Define the matrix $A \in \mathbb{R}^{n\times 3}$ as

$$A = \begin{bmatrix} x_1 & y_1 & -1 \\ \vdots & \vdots & \vdots \\ x_n & y_n & -1 \end{bmatrix}.$$

Assume that you have access to its SVD: $A = U\Sigma V^\top$.

1. Prove that for a general matrix B with SVD $B = \tilde{U}\tilde{\Sigma}\tilde{V}^\top$, $Ker(B) = \text{span}(\{\tilde{v}_i | \tilde{\sigma}_i = 0\})$.
2. Use this to show that the data $\{(x_i, y_i)\}_{i=1}^n$ satisfies an affine relation exactly iff some singular value of A is zero.
3. How do we find the best vector (a, b, c) when all the singular values are larger than zero? How do we know if the approximation is good?

Question: Courant-Fisher & SVD

1. Prove that for a general matrix B with SVD $B = \tilde{U}\tilde{\Sigma}\tilde{V}^\top$, $Ker(B) = \text{span}(\{\tilde{v}_i | \tilde{\sigma}_i = 0\})$.

Question: Courant-Fisher & SVD

1. Prove that for a general matrix B with SVD $B = \tilde{U}\tilde{\Sigma}\tilde{V}^\top$, $Ker(B) = \text{span}(\{\tilde{v}_i | \tilde{\sigma}_i = 0\})$.

Question: Courant-Fisher & SVD

1. Prove that for a general matrix B with SVD $B = \tilde{U}\tilde{\Sigma}\tilde{V}^\top$,
$Ker(B) = \text{span}(\{\tilde{v}_i | \tilde{\sigma}_i = 0\})$.

Question: Courant-Fisher & SVD

$$A = \begin{bmatrix} x_1 & y_1 & -1 \\ \vdots & \vdots & \vdots \\ x_n & y_n & -1 \end{bmatrix}.$$

2. Use this to show that the data $\{(x_i, y_i)\}_{i=1}^n$ satisfies an affine relation exactly iff some singular value of A is zero.

Question: Courant-Fisher & SVD

$$A = \begin{bmatrix} x_1 & y_1 & -1 \\ \vdots & \vdots & \vdots \\ x_n & y_n & -1 \end{bmatrix}.$$

2. Use this to show that the data $\{(x_i, y_i)\}_{i=1}^n$ satisfies an affine relation exactly iff some singular value of A is zero.

Question: Courant-Fisher & SVD

$$A = \begin{bmatrix} x_1 & y_1 & -1 \\ \vdots & \vdots & \vdots \\ x_n & y_n & -1 \end{bmatrix}.$$

3. How do we find the best vector (a, b, c) when all the singular values are larger than zero? How do we know if the approximation is good?

Question: Courant-Fisher & SVD

$$A = \begin{bmatrix} x_1 & y_1 & -1 \\ \vdots & \vdots & \vdots \\ x_n & y_n & -1 \end{bmatrix}.$$

3. How do we find the best vector (a, b, c) when all the singular values are larger than zero? How do we know if the approximation is good?

Question: Courant-Fisher & SVD

$$A = \begin{bmatrix} x_1 & y_1 & -1 \\ \vdots & \vdots & \vdots \\ x_n & y_n & -1 \end{bmatrix}.$$

3. How do we find the best vector (a, b, c) when all the singular values are larger than zero? How do we know if the approximation is good?

Question: Courant-Fisher & SVD

Extra Question: How is this exercise related to the "Best Fitting Subspace" Subsection of the notes of Lecture 7?

Question: Courant-Fisher & SVD

Extra Question: How is this exercise related to the "Best Fitting Subspace" Subsection of the notes of Lecture 7?

Question: SVD and ellipsoids

Explain the following statement: For any $A \in \mathbb{R}^{m \times n}$, the set $\{Ax : \|x| = 1\}$ is an ellipsoid. In other words, the image of the sphere under a linear transformation is always an ellipsoid.

Question: SVD and ellipsoids

Explain the following statement: For any $A \in \mathbb{R}^{m\times n}$, the set $\{Ax : \|x| = 1\}$ is an ellipsoid. In other words, the image of the sphere under a linear transformation is always an ellipsoid.

Next week

Next week the recitation will be about review exercises for the midterm.