Recitation 7

Alex Dong

CDS, NYU

Fall 2020

Interpretation of Matrices/Matrix Multiplication

Previously in the course...

- Matrices as *linear tranformations*
- Matrices *act* on vectors

Now...

- Matrices as *data matrix*
  - rows are instances of data, columns are features
- Not as interpretable to think of linear transformations
- Matrix multiplications used to condense calculations...
- That being said, it can be useful to think of data matrices as linear transformations.
- (!) Think about which framework makes sense in your proofs!

Questions: Principal Component Analysis

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?

Solutions 1: 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.

Solution

1. *Center your data: Let* $X \in \mathbb{R}^{n \times d}$ *be your data matrix.*
   If $X = \begin{bmatrix} - & x_1 & - \\ & \vdots & \\ - & x_n & - \end{bmatrix}$. *Then,* $X_c = X - \begin{bmatrix} - & \frac{1}{n}\sum_{i=1}^n x_i & - \\ & \vdots & \\ - & \frac{1}{n}\sum_{i=1}^n x_i & - \end{bmatrix}$
2. *Construct the covariance matrix* $S = X_c^T X_c \in \mathbb{R}^{d \times d}$
3. *Take the Spectral Decomposition of* $S : S = V \Lambda V^T$.
4. *Choose the top* k *eigenval-vecs* $\lambda_1, ..., \lambda_k$ *from* Λ *and* $v_1, ..., v_k$ *from* V.
5. *Construct your new data matrix as*
   $$X_{new} = X_c V_k = \begin{bmatrix} - & x_1 & - \\ & \vdots & \\ - & x_n & - \end{bmatrix} \begin{bmatrix} | & & | \\ v_1 & \cdots & v_k \\ | & & | \end{bmatrix}$$

Solutions 2: PCA

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

Solution

2. *How can you implement PCA using SVD? Let X be our centered data matrix. Let $X = U\Sigma V^T$.*
   $X^T X = V\Sigma^T U^T U \Sigma V^T = V\Sigma^T \Sigma V^T$.
   From the framework of the previous question, we can set $\Lambda = \Sigma^T\Sigma$, *and* $V = V$. *Then* $\lambda_i = \sigma_i^2$, *and* v_i *as before.*
3. *How do we determine an 'optimal' value for k?*
   i. *Scree Plot*
   ii. *Fraction of variance explained by top k eigenvalues.*

Considerations of PCA

- Loss of interpretibility.
  - Each principal component is a linear combination of all other features.
  - E.g $v_1 = 0.8x_1 + 0.3x_2 + 0.3x_3 + 0.3x_4 + 0.3x_5$
- To standardize or not to standardize?
  - Depends on your data. (No easy answers here)
- Linearity
  - PCA assumes that the components are linear combinations of all other features.

Question: Prelude to SVD

Let $X \in \mathbb{R}^{n \times d}$. Let $X^T X$ have Spectral Decomposition $X^T X = V \Lambda V^T$; where $\lambda_1, ..., \lambda_d > 0$ are the entries of Λ; $v_1, ..., v_n$ are the columns of V.

Let $\sigma_i = \sqrt{\lambda_i}, \quad \forall i \in \{1, ..., d\}$.
Let $u_i = \frac{1}{\sigma_i} X v_i \quad \forall i \in \{1, ..., d\}$.

1. Show that $u_1, ..., u_d$ form an orthonormal basis.

Solution: Prelude to SVD

Let $X \in \mathbb{R}^{n \times d}$. Let $X^T X$ have Spectral Decomposition $X^T X = V \Lambda V^T$; where $\lambda_1, ..., \lambda_d > 0$ are the entries of Λ; $v_1, ..., v_n$ are the columns of V.

Let $\sigma_i = \sqrt{\lambda_i}, \quad \forall i \in \{1, ..., d\}$.
Let $u_i = \frac{1}{\sigma_i} X v_i \quad \forall i \in \{1, ..., d\}$.

1. Show that $u_1, ..., u_d$ form an orthonormal basis.

Solution

$\langle u_i, u_j \rangle = \langle \frac{1}{\sigma_i} A v_i, \frac{1}{\sigma_j} A v_j \rangle = \frac{1}{\sigma_i \sigma_j} v_i^T A^T A v_j = \frac{1}{\sigma_i \sigma_j} \lambda_j v_i^T v_j$.
When $i \neq j$, $\langle u_i, u_j \rangle = 0 \quad$ *(by orthogonality of* v_i, v_j*)*
When $i = j$, $\langle u_i, u_j \rangle = 1 \quad$ *(by normality of of* v_i*, and* $\sigma_i^2 = \lambda_i$*)*

Review Questions 1

A matrix $M \in \mathbb{R}^{n \times n}$ is diagonalizable if M has a basis of eigenvectors that span $\mathbb{R}^n$. True or False:

- If M is diagonalizable, then M is invertible.
- If M is invertible, then M is diagonalizable.

Solutions: Review Questions 1

A matrix $M \in \mathbb{R}^{n \times n}$ is diagonalizable if M has a basis of eigenvectors that span $\mathbb{R}^n$. True or False:

Solution

- *If M is diagonalizable, then M is invertible.*
  False: Consider $M = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ *Eigenvectors are* $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ *and* $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$
- *If M is invertible, then M is diagonalizable.*
  False. Consider a rotation matrix
  $M = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$

Review Questions 2

True or False:

1. There can exist a set of n non-zero orthogonal vectors in $\mathbb{R}^m$ if $n > m$
2. The matrix corresponding to an orthogonal projection is symmetric
3. The matrix corresponding to an orthogonal projection is orthogonal

Solutions: Review Questions 2

True or False:

Solution

1. *There can exist a set of n non-zero mutually orthogonal vectors in $\mathbb{R}^m$ if $n > m$*
   False, $\mathbb{R}^m$ can at most contain m linearly independent vectors.
2. *The matrix corresponding to an orthogonal projection is symmetric.*
   True. All orthogonal projections can be expressed as VV^T.
3. *The matrix corresponding to an orthogonal projection is orthogonal*
   False. Orthogonal matrices have full rank. Orthogonal projections (usually) don't.

Midterm Review Tips

- Start studying early. Don't try to cram.
- Open book midterm, but *don't assume your notes will help*
    - Consider notes as *reference only*
    - Organize notes beforehand
    - Exam has a notable amount of "time pressure"
- Understand the proofs of the major theorems from lecture.
- Realistically, you will have enough time for one to two "attempts" per question.
- Review: (Suggested order)
    - Midterm Practice Questions
    - Midterm 2019
    - Homework Questions (especially proofs)
    - Recitation Questions
    - Midterm 2018