CSCI3230 (ESTR3108)
Fundamentals of Artificial Intelligence

Tutorial 6

Yuehao Wang

Email: yhwang21@cse.cuhk.edu.hk
Office: Room 1024, 10/F, SHB

Dept. of Computer Science & Engineering
The Chinese University of Hong Kong

Outline

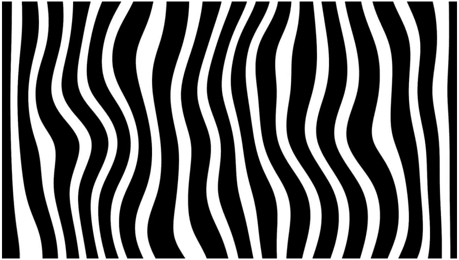
Part 1. Change of basis

Part 2. Spectral theory

Part 3. Robust PCA

Principal components

Principal components



Principal components

Principal components

Part 1. Change of basis

Independence of vectors

- Suppose we have vectors $u_1 \dots u_n \in \mathbb{R}^n$
- If $u_1 \dots u_n$ are independent then:

$$c_1 u_1 + \dots + c_n u_n = 0, \quad c_i \in \mathbb{R}, i = 1 \dots n$$

  iff $c_1 = \dots = c_n = 0$
- Alternatively, we can write this statement in matrix-vector multiplication: Let $A = [u_1 \dots u_n]$, $c \in \mathbb{R}^n$:

$$Ac = 0$$

  iff $c = 0$, which also implies A is full-rank.

Quiz

Given $u_1 \dots u_n \in \mathbb{R}^n$, for any pair (u_i, u_j), $i \neq j$, u_i and u_j are independent. Can we say $u_1 \dots u_n$ are independent to each other?

Basis

- Basis is a set of independent vectors, which can span the vector space by linear combination.
- Denote basis as $B = \{u_1, \dots, u_n\}$, where $u_1 \dots u_n$ are independent.
- Let $U = [u_1 \dots u_n]$. The subspace spanned by B is exactly the range space (column space) of U.
- If S is spanned by B, then any vector $v \in S$, we can find scalars $c_1, \dots, c_n$ such that:

$$v = c_1 u_1 + \cdots + c_n u_n$$

- The number of vectors in the basis = dimension: $dim(S) = n$.

Basis

Typical basis:

- Canonical basis: $\{e_1, \ldots, e_n\}$, a very useful and common basis of $\mathbb{R}^n$. For any i, e_i is a vector where only the i-th element equal to 1 and others are zeros.
  1. For example, in R^2 space, $e_1 = [1, 0]$ and $e_2 = [0, 1]$ are the two canoical basis vectors.
  2. Coordinates like $(x_1, \ldots, x_n)$ without other context is actually the shorthand of $x_1 e_1 + \cdots + x_n e_n$.
- Orthonormal basis: a basis $[u_1, \ldots, u_n]$ is orthonormal iff $u_j^T u_i = 0$ when $i \neq j$ and $u_i^T u_i = 1$.
  1. For example, canoical basis is orthonormal.
  2. Basis $\{[\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}], [\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}]\}$ is orthonormal.

Basis

- Note that basis of a subspace is not unique.
  1. Besides of the canonical basis e_1, e_2, e_3, $u_1 = (1,1,1)^T, u_2 = (0,1,0)^T, u_3 = (1,2,2)^T$ can also be a basis for 3-D Euclidean space.
  2. Basis for a plane in $\mathbb{R}^3$ is not unique as well. Spanning of $(0,1,0)^T, (1,0,0)^T$ and $(1,1,0)^T, (1,0,0)^T$ are the same plane.
- Different basis may have different use.
  1. In computer graphics, we often need to transform a point from the world coordinate system to the camera coordinate system.
  2. Discrete signals are generally recorded in time domain. But we usually do Fourier transform to get the frequency domain information. Fourier transform: canonical basis $=>$ fourier basis.
- This is the reason why we need change of basis.

Change of basis

Change of basis transformation

Given a basis $B = \{u_1, \ldots, u_n\}$, let $A = [u_1 \ldots u_n]$, the change of basis transformation from canonical basis to B is A^{-1}.

- Convsersely, A is the change of basis transformation from B to canonical basis.
- For example: Suppose we have basis $[1, 0, 0]$ and $[0, 0, 1]$. A coordinate $p = [2, 3]$ represented by this basis is actually the coordinate $[2, 0, 3]$ under the canoical basis. By applying the theorem,
$$Ap = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix}.$$

Part 2. Spectral theory

Eigenvalues and eigenvectors

$\forall A \in \mathbb{R}^{n\times n}$, if we have the equation below for some λ and v,

$$Av = \lambda v$$

- λ is called eigenvalue and the corresponding non-zero v is called eigenvector asscoiated with the λ.
- Eigenvalues are solutions to $det(A - \lambda I) = 0$, which is a n-order polynomial.
- When there are n independent eigenvectors, A can be diagonalizable.

$$A = U\Lambda U^{-1}$$

where Λ is a diagonal matrix with eigenvalues $\lambda_1 \dots \lambda_n$ on its diagonal, U's i-th column is the eigenvector associated with λ_i

Eigenvalues and eigenvectors

Spectral theory for symmetric matrix

- For symmetric A (i.e., $A^T = A$), A must have n real eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ and the associated orthonormal eigenvectors $v_1, v_2, \ldots, v_n$:

$$Av_i = \lambda_i v_i$$

  - A can be diagonalized as

$$A = U \Lambda U^T$$

    where $U = [v_1 \ldots v_n]$ and $\Lambda = diag(\lambda_1, \ldots, \lambda_n)$
  - Spectral resolution: $A = \sum_{i=1}^{n} \lambda_i v_i v_i^T$
  - Courant-Fischer theorem:

$$\lambda_i = \max_{dimV=i} \min_{x \in V, \|x\|_2=1} x^T A x$$

Spectral theory for symmetric matrix

Quiz

Let this quiz help with recapping spectral theory. Which statement below about eigenvalues and eigenvectors is correct?

- A Eigenvectors of $A \in \mathbb{R}^{n \times n}$ need not to be independent.
- B Any matrix $A \in \mathbb{R}^{n \times n}$ must have n eigenvalues.
- C Symmetric matrix $A \in \mathbb{R}^{n \times n}$ must have n non-repeated eigenvalues.
- D Any matrix $A \in \mathbb{R}^{n \times n}$ is NOT invertible iff 0 is an eigenvalue of A.

Quiz

Let this quiz help with recapping spectral theory. Which statement below about eigenvalues and eigenvectors is correct?

A Eigenvectors of $A \in \mathbb{R}^{n \times n}$ need not to be independent.

B Any matrix $A \in \mathbb{R}^{n \times n}$ must have n eigenvalues.

C Symmetric matrix $A \in \mathbb{R}^{n \times n}$ must have n non-repeated eigenvalues.

D Any matrix $A \in \mathbb{R}^{n \times n}$ is NOT invertible iff 0 is an eigenvalue of A.

Correct Answer: D

Part 3. Robust PCA

Dense noise

Suppose data matrix $X^* \in \mathbb{R}^{D\times N}$, $rank(X^*) = d, d < \min\{D, N\}$. $\tilde{X}$ is a corruption of X^*. Problem:

$$\tilde{X} = X^* + E$$

where E is a dense noise matrix (suppose Gaussian noise).
Solve this problem through optimization:

$$\min_{rank(X)=d} \|\tilde{X} - X\|_F$$

Sparse noise

Problem:

$$\tilde{X} = X^* + E$$

where E is a sparse noise matrix, i.e., $\|E\|_0$ is small.
Solve this problem through optimization:

$$\min_{X \in \mathbb{R}^{D \times N}} rank(X) + \|\tilde{X} - X\|_0$$

The objective function is not convex. Solution: convex envelope.
The convex envelope of $\|\cdot\|_0$ is $\|\cdot\|_1$. The convex envelope of $rank(\cdot)$ is the nuclear norm $\|\cdot\|_*$.

Missing entries (Matrix completion)

Problem:

$$\tilde{X} = X^* \otimes E_\Omega, \text{ given } \Omega \subset [D] \times [N]$$

$$E_\Omega(i,j) = 1, \text{ if}(i,j) \in \Omega$$

$$E_\Omega(i,j) = 0, \text{ otherwise}$$

Application: Netflix recommendation systems.
This problem can be solved via optimization if E_Ω is not dense and $rank(X^*)$ is small:

$$\min_X \; rank(X)$$