CS 540 Introduction to Artificial Intelligence
Linear Algebra & PCA

University of Wisconsin-Madison
Fall 2023
Linear Algebra: What is it good for?

• Study of Linear functions: simple, tractable
• In AI/ML: building blocks for all models
  — e.g., linear regression; part of neural networks
Outline

• Basics: vectors, matrices, operations

• Dimensionality reduction

• Principal Components Analysis (PCA)
Basics: Vectors

• Many interpretations
  – List of values (represents information)
  – Point in a space
• Dimension: number of values: $x \in \mathbb{R}^d$
• AI/ML: often use very high dimensions:
  – Ex: images!
Basics: Matrices

• Many interpretations
  – Table of values; list of vectors
  – Represent linear transformations
  – Apply to a vector, get another vector

• Dimensions: #rows × #columns, \( A \in \mathbb{R}^{m \times n} \)
  – Indexing!

\[
A = \begin{bmatrix}
A_{11} & A_{12} & A_{13} \\
A_{21} & A_{22} & A_{23} \\
A_{31} & A_{32} & A_{33} \\
A_{41} & A_{42} & A_{43}
\end{bmatrix}
\]
Basics: Transposition

• Transposes: flip rows and columns
  – Vector: standard is a column. Transpose: row vector
  – Matrix: go from $m \times n$ to $n \times m$

\[
x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad x^T = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}
\]

\[
A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{bmatrix} \quad A^T = \begin{bmatrix} A_{11} & A_{21} \\ A_{12} & A_{22} \\ A_{13} & A_{23} \end{bmatrix}
\]
Matrix & Vector Operations

• Vectors
  – **Addition**: component-wise
    • Commutative: $x + y = y + x$
    • Associative: $(x + y) + z = x + (y + z)$

  \[
  x + y = \begin{bmatrix}
  x_1 + y_1 \\
  x_2 + y_2 \\
  x_3 + y_3
  \end{bmatrix}
  \]

  – **Scalar Multiplication**
    • Uniform stretch / scaling

  \[
  cx = \begin{bmatrix}
  cx_1 \\
  cx_2 \\
  cx_3
  \end{bmatrix}
  \]
Matrix & Vector Operations

• Vector products
  – Inner product (e.g., dot product)
    \[ <x, y> := x^T y = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = x_1 y_1 + x_2 y_2 + x_3 y_3 \]
  – Outer product
    \[ xy^T = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix} = \begin{bmatrix} x_1 y_1 & x_1 y_2 & x_1 y_3 \\ x_2 y_1 & x_2 y_2 & x_2 y_3 \\ x_3 y_1 & x_3 y_2 & x_3 y_3 \end{bmatrix} \]
Matrix & Vector Operations

• $x$ and $y$ are **orthogonal** if $\langle x, y \rangle = 0$

• Vector **norms**: “length”

\[
\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}
\]
Matrix & Vector Operations

• Matrices:
  – **Addition**: Component-wise
  – Commutative, Associative

\[
A + B = \begin{bmatrix}
A_{11} + B_{11} & A_{12} + B_{12} \\
A_{21} + B_{21} & A_{22} + B_{22} \\
A_{31} + B_{31} & A_{32} + B_{32}
\end{bmatrix}
\]

  – **Scalar Multiplication**
  – "Stretching" the linear transformation

\[
cA = \begin{bmatrix}
cA_{11} & cA_{12} \\
cA_{21} & cA_{22} \\
cA_{31} & cA_{32}
\end{bmatrix}
\]
Matrix & Vector Operations

• Matrix-Vector multiplication
  – Linear transformation; plug in vector, get another vector
  – Each entry in $Ax$ is the inner product of a row of $A$ with $x$

\[
x \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n}
\]

\[
Ax = \begin{bmatrix}
\langle A_1, x \rangle \\
\langle A_2, x \rangle \\
\vdots \\
\langle A_m, x \rangle 
\end{bmatrix} = \begin{bmatrix}
A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_n \\
A_{21}x_1 + A_{22}x_2 + \cdots + A_{2n}x_n \\
\vdots \\
A_{m1}x_1 + A_{m2}x_2 + \cdots + A_{mn}x_n
\end{bmatrix}
\]

- Output of layer $k$ is

$$f^{(k)}(x) = \sigma(W_k^T f^{(k-1)}(x)))$$

- Output of layer $k-1$: vector
- Output of layer $k$: vector
- Weight matrix for layer $k$: 
- Note: linear transformation!
Matrix & Vector Operations

• Matrix multiplication
  – \( A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times p} \), then \( AB \in \mathbb{R}^{m \times p} \)
  – “Composition” of linear transformations
  – Not commutative in general!
    \[ AB \neq BA \]
  – Lots of interpretations
Identity Matrix

- Like “1”
- Multiplying by it gets back the same matrix or vector
- Rows & columns are the “standard basis vectors” $e_i$
Break & Quiz

• **Q 1.1:** What is $\begin{bmatrix} 1 & 2 \\ 3 & 1 \\ 1 & 1 \end{bmatrix} \times \begin{bmatrix} 0 \\ 1 \end{bmatrix}$?

• A. $[-1 \ 1 \ 1]^T$
• B. $[2 \ 1 \ 1]^T$
• C. $[1 \ 3 \ 1]^T$
• D. $[1.5 \ 2 \ 1]^T$
Break & Quiz

Q 1.1: What is $\begin{bmatrix} 1 & 2 \\ 3 & 1 \\ 1 & 1 \end{bmatrix} \times \begin{bmatrix} 0 \\ 1 \end{bmatrix}$?

A. $[-1 \ 1 \ 1]^T$
B. $[2 \ 1 \ 1]^T$
C. $[1 \ 3 \ 1]^T$
D. $[1.5 \ 2 \ 1]^T$
Break & Quiz

Q 1.1: What is \[
\begin{bmatrix}
1 & 2 \\
3 & 1 \\
1 & 1
\end{bmatrix}
\times
\begin{bmatrix}
0 \\
1
\end{bmatrix}
\] ?

• A. \([-1 \ 1 \ 1]^T\)

• B. \([2 \ 1 \ 1]^T\)

• C. \([1 \ 3 \ 1]^T\)

• D. \([1.5 \ 2 \ 1]^T\)

Check dimensions: answer must be 3 x 1 matrix (i.e., column vector).

\[
\begin{bmatrix}
1 & 2 \\
3 & 1 \\
1 & 1
\end{bmatrix}
\times
\begin{bmatrix}
0 \\
1
\end{bmatrix}
= 
\begin{bmatrix}
0 \times 1 + 1 \times 2 \\
0 \times 3 + 1 \times 1 \\
0 \times 1 + 1 \times 1
\end{bmatrix}
= 
\begin{bmatrix}
2 \\
1 \\
1
\end{bmatrix}
\]
Break & Quiz

• Q 1.2: Given matrices $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{d \times m}$, $C \in \mathbb{R}^{p \times n}$

What are the dimensions of $BAC^T$?

• A. $n \times p$
• B. $d \times p$
• C. $d \times n$
• D. Undefined
Break & Quiz

• **Q 1.2**: Given matrices $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{d \times m}$, $C \in \mathbb{R}^{p \times n}$.

What are the dimensions of $BAC^T$?

• A. $n \times p$
• B. $d \times p$
• C. $d \times n$
• D. Undefined
Break & Quiz

• **Q 1.2**: Given matrices $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{d \times m}$, $C \in \mathbb{R}^{p \times n}$

What are the dimensions of $BAC^T$?

- A. $n \times p$
- B. $d \times p$
- C. $d \times n$
- D. Undefined

To rule out (D), check that for each pair of adjacent matrices XY, the # of columns of X = # of rows of Y.

Then, B has d rows so solution must have d rows. $C^T$ has p columns so solution has p columns.
Break & Quiz

• Q 1.3: A and B are matrices, neither of which is the identity. Is $AB = BA$?

• A. Never
• B. Always
• C. Sometimes
Break & Quiz

Q 1.3: A and B are matrices, neither of which is the identity. Is $AB = BA$?

• A. Never
• B. Always
• C. Sometimes
Break & Quiz

• **Q 1.3**: A and B are matrices, neither of which is the identity. Is $AB = BA$?

  • A. Never
  • B. Always
  • **C. Sometimes**

Matrix multiplication is not necessarily commutative.
Matrix Inverse

• If there is a $B$ such that $AB = BA = I$
  – Then $A$ is invertible/nonsingular, $B$ is its inverse
  – Some matrices are not invertible!

• Notation: $A^{-1}$

$$
\begin{bmatrix}
1 & 1 \\
2 & 3
\end{bmatrix} \times \begin{bmatrix}
3 & -1 \\
-2 & 1
\end{bmatrix} = I
$$
Eigenvalues & Eigenvectors

• For a square matrix $A$, solutions to $Av = \lambda v$
  – $v$ is a (nonzero) vector: eigenvector
  – $\lambda$ is a scalar: eigenvalue

• Intuition
  – Multiplying by $A$ can stretch/rotate vectors
  – Eigenvectors $v$: only stretched (by $\lambda$)
Dimensionality Reduction

• Vectors store features. Lots of features!
  • Document classification: thousands of words per doc
  • Netflix surveys: 480189 users x 17770 movies
  • MEG Brain Imaging: 120 locations x 500 time points x 20 objects
Dimensionality Reduction

Reduce dimensions

• Why?
  – Lots of features redundant
  – Storage & computation costs

• Goal: take $x \in \mathbb{R}^d \rightarrow x \in \mathbb{R}^r$, for $r \ll d$
  – But, minimize information loss
Dimensionality Reduction

**Examples:** 3D to 2D
Q 2.1: What is the inverse of \( A = \begin{bmatrix} 0 & 2 \\ 3 & 0 \end{bmatrix} \)

A: \( A^{-1} = \begin{bmatrix} -3 & 0 \\ 0 & -2 \end{bmatrix} \)

B: \( A^{-1} = \begin{bmatrix} 0 & \frac{1}{3} \\ \frac{1}{2} & 0 \end{bmatrix} \)

C: Undefined / \( A \) is not invertible
Break & Quiz

Q 2.1: What is the inverse of $A = \begin{bmatrix} 0 & 2 \\ 3 & 0 \end{bmatrix}$

A: $A^{-1} = \begin{bmatrix} -3 & 0 \\ 0 & -2 \end{bmatrix}$

$AA^{-1} = \begin{bmatrix} 0 & 2 \\ 3 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 0 * a + c * 2 & 0 * b + 2 * d \\ 3 * a + c * 0 & 3 * b + 0 * d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

B: $A^{-1} = \begin{bmatrix} 0 & 1/3 \\ 1/2 & 0 \end{bmatrix}$

2c = 1
3a = 0
2d = 0
3b = 1

C: Undefined / $A$ is not invertible

\[
\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 0 & 1/3 \\ 1/2 & 0 \end{bmatrix}
\]
Break & Quiz

Q 2.2: What are the eigenvalues of

\[ A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix} \]

A. -1, 2, 4
B. 0.5, 0.2, 1.0
C. 0, 2, 5
D. 2, 5, 1
Break & Quiz

Q 2.2: What are the eigenvalues of

\[
A = \begin{bmatrix}
2 & 0 & 0 \\
0 & 5 & 0 \\
0 & 0 & 1
\end{bmatrix}
\]

A. -1, 2, 4
B. 0.5, 0.2, 1.0
C. 0, 2, 5
D. 2, 5, 1
Q 2.2: What are the eigenvalues of \[ A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix} \]

A. -1, 2, 4  
B. 0.5, 0.2, 1.0  
C. 0, 2, 5  
D. 2, 5, 1  

Solution #1: You may recall from a linear algebra course that the eigenvalues of a diagonal matrix (in which only diagonal entries are non-zero) are just the entries along the diagonal. Hence D is the correct answer.
Q 2.2: What are the eigenvalues of $A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Solution #2: Use the definition of eigenvectors and values: $Av = \lambda v$

\[
\begin{bmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix} v = \begin{bmatrix} 2v_1 + 0v_2 + 0v_3 \\ 0v_1 + 5v_2 + 0v_3 \\ 0v_1 + 0v_2 + 1v_3 \end{bmatrix} = \begin{bmatrix} \lambda v_1 \\ 5v_2 \\ v_3 \end{bmatrix}
\]

Since $A$ is a 3x3 matrix, $A$ has 3 eigenvalues and so there are 3 combinations of values for $\lambda$ and $v$ that will satisfy the above equation. The simple form of the equations suggests starting by checking each of the standard basis vectors* as $v$ and then solving for $\lambda$. Doing so gives D as the correct answer.

*A each standard basis vector $e_i \in \mathbb{R}^n$ is the vector in which all components are zero except component $i$ is 1.
Q 2.3: Suppose we are given a dataset with $n=10000$ samples with 100-dimensional binary feature vectors. Our storage device has a capacity of 50000 bits. What’s the lower compression ratio we can use?

A. 20X
B. 100X
C. 5X
D. 1X
Q 2.3: Suppose we are given a dataset with $n=10000$ samples with 100-dimensional binary feature vectors. Our storage device has a capacity of 50000 bits. What’s the lower compression ratio we can use?

A. 20X
B. 100X
C. 5X
D. 1X
Q 2.3: Suppose we are given a dataset with \( n = 10000 \) samples with 100-dimensional binary feature vectors. Our storage device has a capacity of 50000 bits. What’s the lower compression ratio we can use?

A. 20X

B. 100X

C. 5X

D. 1X

50,000 bits / 10,000 samples means compressed version must have 5 bits / sample.

Dataset has 100 bits / sample.

Must compress 20x smaller to fit on device.
Principal Components Analysis (PCA)

- A type of dimensionality reduction approach
  - For when data is approximately lower dimensional
Principal Components Analysis (PCA)

• Find axes \( u_1, u_2, \ldots, u_m \in \mathbb{R}^d \) of a subspace
  – Will project to this subspace

• Want to preserve data
  – Minimize projection error

• These vectors are the principal components
Projection: An Example

\[ x_1, x_2, \ldots, x_n \in \mathbb{R}^2 \]
Projection: An Example

$x_1, x_2, \ldots, x_n \in \mathbb{R}^2$

A random line that goes through the origin
Projection: An Example

\[ x_1, x_2, \ldots, x_n \in \mathbb{R}^2 \]

PCA projects data onto this line
Projection: An Example

\[ x_1, x_2, \ldots, x_n \in \mathbb{R}^2 \]

Goal: finding a line that \textbf{minimizes} the sum of squared distances to \( x_i \)'s
Projection: An Example

$x_1, x_2, \ldots, x_n \in \mathbb{R}^2$

The sum of squared distances gets smaller as the line fits better.

The optimal line is called Principal Component 1.
PCA Procedure

• **Inputs:** data $x_1, x_2, \ldots, x_n \in \mathbb{R}^d$
  
  – Center data so that $\frac{1}{n} \sum_{i=1}^{n} x_i = 0$

• **Output:**
  principal components $u_1, \ldots, u_m \in \mathbb{R}^d$
  
  – Orthogonal
  
  – Can show: they are top-$m$ eigenvectors of
  
  $$S = \frac{1}{n-1} \sum_{i=1}^{n} x_i x_i^\top$$ (covariance matrix)

  – Each $x_i$ projected to $x_i^{\text{pca}} = \sum_{j=1}^{m} (u_j^\top x_i) u_j$
Many Variations

• PCA, Kernel PCA, ICA, CCA
  – Extract structure from high dimensional dataset

• Uses:
  – Visualization
  – Efficiency
  – Noise removal
  – Downstream machine learning use
Application: Image Compression

• Start with image; divide into 12x12 patches
  - That is, 144-D vector
  - Original image:
Application: Image Compression

- 6 principal components (as an image)
Application: Image Compression

- Project to 6D
Application: Exploratory Data Analysis

- [Novembre et al. ’08]: Take top two singular vectors of people x SNP matrix (POPRES)

“Genes Mirror Geography in Europe”
Readings

- Vast literature on linear algebra.
- Local class: Math 341
- More on PCA (and other matrix methods in ML): CS 532

**Suggested reading:**
- Lecture notes on PCA by Roughgarden and Valiant  