Outline

- Basic review on linear algebra
- Introduction to dimensionality reduction
- Principal component analysis: formulation and computation
- Applications

Transposition

$$\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \qquad \mathbf{b}^T = \begin{bmatrix} 1 & 1 & 2 \end{bmatrix} \qquad \mathbf{d} = \begin{bmatrix} 3 & 4 & 9 \end{bmatrix} \qquad \mathbf{d}^T = \begin{bmatrix} 3 \\ 4 \\ 9 \end{bmatrix}$$

column $\longrightarrow$ row   row $\longrightarrow$ column

$$\mathbf{A} = \begin{bmatrix} \boxed{1 \quad 2 \quad 3} \\ \boxed{5 \quad 4 \quad 1} \\ \boxed{6 \quad 7 \quad 4} \end{bmatrix} \qquad \mathbf{A}^T = \begin{bmatrix} 1 & 5 & 6 \\ 2 & 4 & 7 \\ 3 & 1 & 4 \end{bmatrix}$$

Matrix Calculations

Addition

- Commutative: **A+B=B+A**
- Associative: **(A+B)+C=A+(B+C)**

$$\mathbf{A}+\mathbf{B}=\begin{bmatrix}2 & 4\\ 2 & 5\end{bmatrix}+\begin{bmatrix}1 & 0\\ 3 & 1\end{bmatrix}=\begin{bmatrix}2+1 & 4+0\\ 2+3 & 5+1\end{bmatrix}=\begin{bmatrix}3 & 4\\ 5 & 6\end{bmatrix}$$

Subtraction

- By adding a negative matrix

$$\mathbf{A}-\mathbf{B}=\begin{bmatrix}2 & 4\\ 5 & 3\end{bmatrix}-\begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix}=\begin{bmatrix}2 & 4\\ 5 & 3\end{bmatrix}+\begin{bmatrix}-1 & -2\\ -3 & -4\end{bmatrix}=\begin{bmatrix}1 & 2\\ 2 & -1\end{bmatrix}$$

Matrix Multiplication

*"When **A** is a mxn matrix & **B** is a kxl matrix, AB is only possible if n=k. The result will be an mxl matrix"*

$$\begin{array}{cccc} & & \mathbf{n} \longrightarrow & \\ \mathbf{m}\downarrow & A_1 & A_2 & A_3 \\ & A_4 & A_5 & A_6 \\ & A_7 & A_8 & A_9 \\ & A_{10} & A_{11} & A_{12} \end{array} \quad \times \quad \begin{array}{ccc} \mathbf{l} \longrightarrow & & \\ B_{13} & B_{14} & \\ B_{15} & B_{16} & \downarrow\mathbf{k} \\ B_{17} & B_{18} & \end{array} \quad = \mathbf{m} \times \mathbf{l} \text{ matrix}$$

Number of columns in A = Number of rows in B

Matrix multiplication

- Multiplication method:

Sum over product of respective rows and columns

$$\begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix} \times \begin{pmatrix} 2 & 1 \\ 3 & 1 \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$$

A **B**

Define output matrix

$$= \begin{bmatrix} (1\times 2)+(0\times 3) & (1\times 1)+(0\times 1) \\ (2\times 2)+(3\times 3) & (2\times 1)+(3\times 1) \end{bmatrix}$$

$$= \begin{pmatrix} \mathbf{2} & \mathbf{1} \\ \mathbf{13} & \mathbf{5} \end{pmatrix}$$

$$\boxed{C_{i,j} = \sum_{k=1}^{n} A_{i,k} B_{k,j}}$$

Matrix multiplication

- Matrix multiplication is NOT commutative
- $AB \neq BA$
- Matrix multiplication IS associative
- $A(BC)=(AB)C$
- Matrix multiplication IS distributive
- $A(B+C)=AB+AC$
- $(A+B)C=AC+BC$
- I is identity matrix, then $AI=IA=A$ if A is square

Vector Product

Two vectors: $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$

Inner product = scalar

Inner product $X^T Y$ is a scalar
(1xn) (nx1)

$$\mathbf{x}^T \mathbf{y} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = x_1 y_1 + x_2 y_2 + x_3 y_3 = \sum_{i=1}^{3} x_i y_i$$

Two vectors x and y are orthogonal if $x^T y = 0$

e.g., $x = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and $y = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

Vector norm

- Euclidean norm or 2-norm $\|\ x\ \|_2$ for $x \in R^D$

$$||x||_2^2 = x^T x = \sum_{k=1}^{D} x_k^2$$

- Unit vector v if $\|\ v\ \|_2 = 1$

e.g., $v = [1\ 0\ 0]$

Eigenvector and Eigenvalue

$$\mathbf{Ax = \lambda x}$$

A: Square Matrix

x: Eigenvector

λ : Eigenvalue

Example

$$Show\ x = \begin{bmatrix} 2 \\ 1 \end{bmatrix} is\,an\,eigenvector\ for\ A = \begin{bmatrix} 2 & -4 \\ 3 & -6 \end{bmatrix}$$

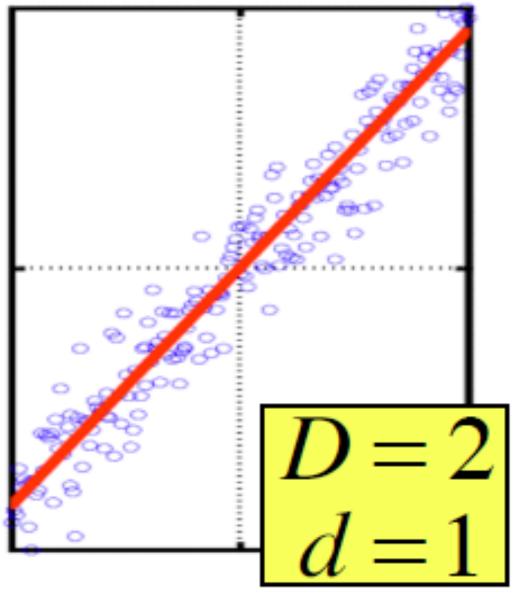
$$Solution : Ax = \begin{bmatrix} 2 & -4 \\ 3 & -6 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

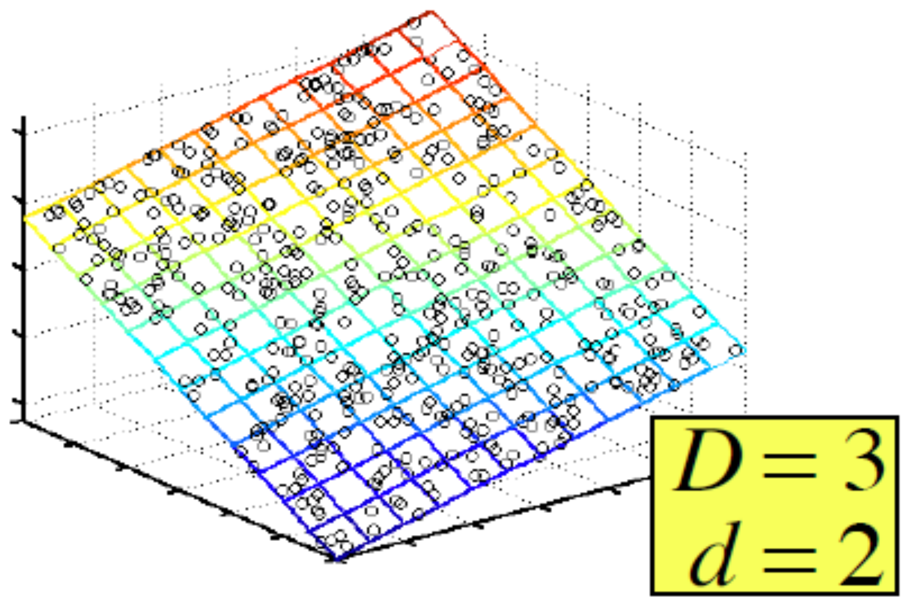
$$But\quad for\ \ \lambda = 0,\ \ \lambda x = 0 \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$Thus, x\ is\ an\ eigenvector\ of\ A, and\ \lambda = 0\ is\ an\ eigenvalue.$$

- ***The zero vector can not be an eigenvector***
- ***The value zero can be eigenvalue***

PCA formulation





- Reduce from 2-dimension to 1-dimension: Find a direction

(a red vector $v_1 \in R^D$) onto which to project the data so as to minimize the projection error.

- Reduce from D-dimension to d-dimension: Find d vectors $v_i \in R^D, i = 1,2,..,d$ onto which to project the data, so as to minimize the projection error.
- v_i is called a principal component (PC)

Andrew Ng

Principal Component Analysis

Input: *N* data points (D-dim vectors)

$$\mathbf{x} \in \mathbb{R}^D: \ \mathcal{D} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$$

Output:

- *d* principal components (PCs)

$$v_j \in R^D, j = 1,2,..,d$$

s.t.,

$$v^T{}_i \cdot v_j = 0, i \neq j \text{ and } v^T{}_i \cdot v_i = 1, \ i = j$$

- For each x_i, it's project coordinates on $\{v_j\}$:

$$w_{i,j} = v_j^T * \mathrm{x}_i, j = 1,2, \ldots, d$$

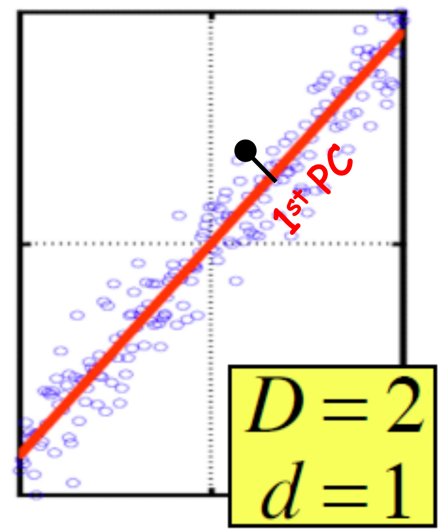
- Now x_i, a D-dim vector can be represented by a d-dim vector (d<D)

$$\Rightarrow \quad [w_{i,1}, w_{i,2}, \ldots, w_{i,d}]$$

Principal Component Analysis (PCA)

Principal Components (PC) are orthogonal directions that capture most of the variance in the data.

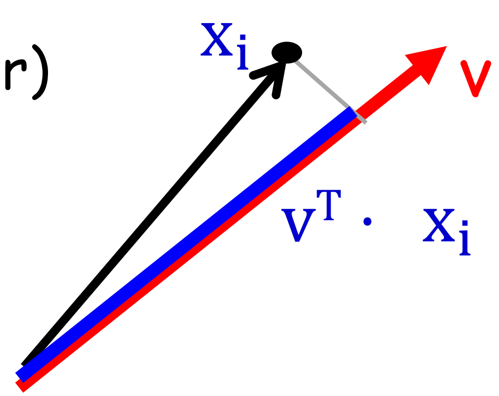
- First PC – direction of greatest variability in data.
- Projection of data points along first PC discriminates data most along any one direction
  (pts are the most spread out when we project the data on that direction compared to any other directions).



Quick reminder:

$||v||=1$, Point x_i (D-dimensional vector)

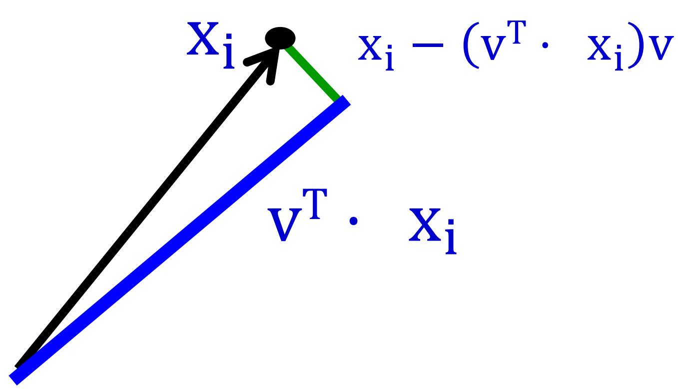
Projection of x_i onto v is $v^T \cdot x_i$

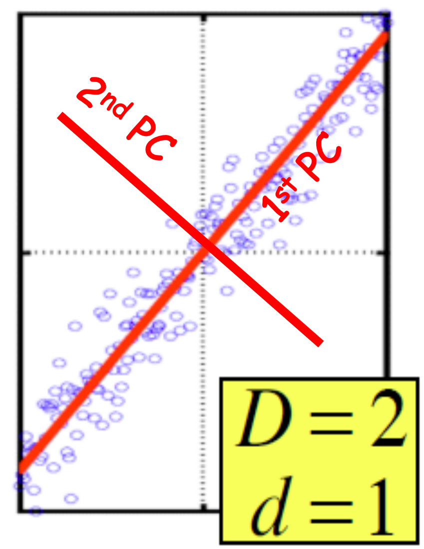


Principal Component Analysis (PCA)

Principal Components (PC) are orthogonal directions that capture most of the variance in the data.

- 1st PC – direction of greatest variability in data.





- 2nd PC – Next orthogonal (uncorrelated) direction of greatest variability

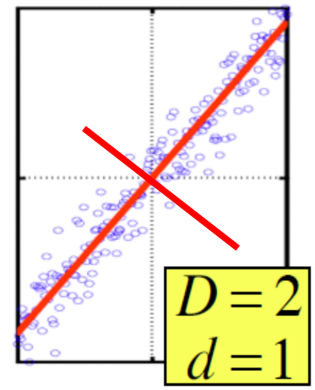
  (remove all variability in first direction, then find next direction of greatest variability)

- And so on …

Principal Component Analysis (PCA)

$(\mathrm{X}\,\mathrm{X}^{\mathrm{T}})\mathrm{v} = \lambda\mathrm{v}$, so v (the first PC) is the eigenvector of sample correlation/covariance matrix $X\,X^T$

Sample variance of projection $\mathrm{v}^T X\,X^T \mathrm{v} = \lambda \mathrm{v}^T\mathrm{v} = \lambda$



Thus, the eigenvalue λ denotes the amount of variability captured along that dimension (aka amount of energy along that dimension).

Eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots$

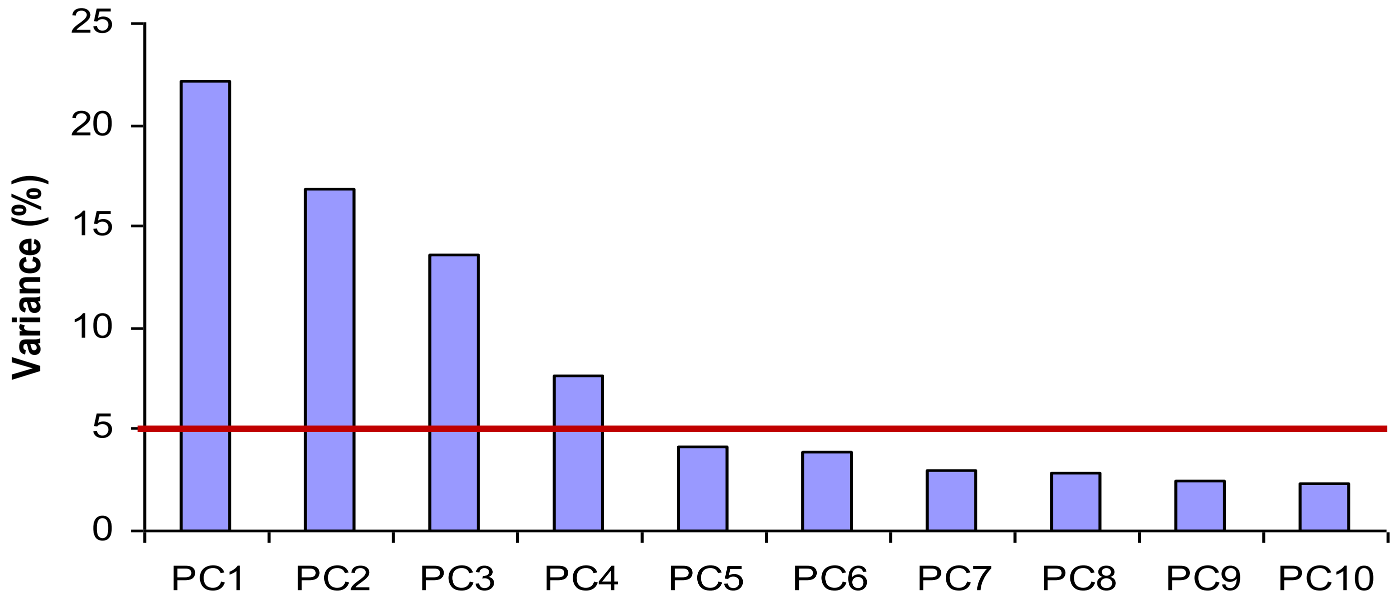
- The 1st PC v_1 is the eigenvector of the sample covariance matrix $X\,X^T$ associated with the largest eigenvalue
- The 2nd PC v_2 is the eigenvector of the sample covariance matrix $X\,X^T$ associated with the second largest eigenvalue
- And so on …

Dimensionality Reduction using PCA

In high-dimensional problems, data sometimes lies near a linear subspace, as noise introduces small variability

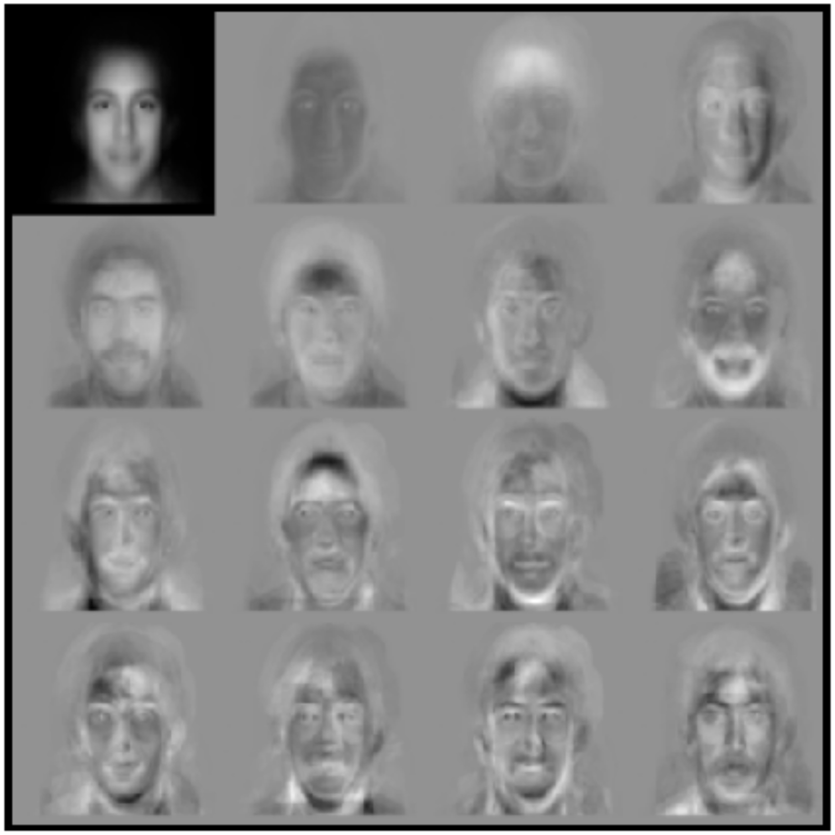
Only keep data projections onto principal components with large eigenvalues

Can *ignore* the components of smaller significance.



Might lose some info, but if eigenvalues are small, do not lose much

Example: faces



Eigenfaces from 7562 images:

top left image is linear combination of rest.

Sirovich & Kirby (1987)
Turk & Pentland (1991)

Can represent a face image using just 15 numbers!