PCA

- DIMENSIONALITY REDUCTION
- COMPARE TO "RANDOM PROJECTION" "JL-LEMMA"

  RANDOMLY PICKED VECTORS $r_1, \dots, r_k$

  PROJECTED x $\underline{\langle x, r_1 \rangle, \dots, \langle x, r_k \rangle}$

  - r_i's WERE <u>NOT</u> MEANINGFUL WITH RESPECT TO S
  - PRESERVE EUCLIDEAN DISTANCE BTWN POINTS
  - IN PRACTICE $K > 100$ FOR RANDOM PROJECTION TO WORK.
  - FOR PCA WE CAN CHOOSE $K = 2$
  - PCA LOOKS AT S TO COME UP WITH A NEW REPRESENTATION

HIGH LEVEL GOAL OF PCA IS TO FIND VECTORS

$$v_1, \dots, v_k \quad \text{s.t.} \quad \forall x \in S \quad x \approx \sum_{j=1}^{k} a_j v_j$$

NOTE ABOUT PRE-PROCESSING OF S

- SUBTRACT THE MEAN OR CENTER OF MASS FROM EACH DATA POINT.
- NORMALIZE THE STANDARD DEVIATION OF EACH FEATURE. BY

$$\forall i \quad \text{COMPUTE} \quad \sqrt{\frac{1}{m}\sum_{j=1}^{m} (x_i^j)^2} = \sigma_i$$

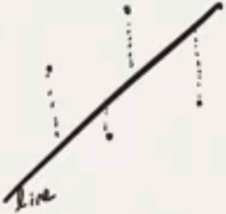
DIVIDE ALL THE i^{th} FEATURES BY σ_i

HOW TO BEGIN? FIND V_1

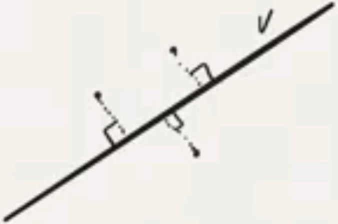
LOOK FOR A VECTOR THAT MINIMIZES SQ-DISTANCE

$$\min_{V,\ \|V\|_2=1} \frac{1}{m}\sum_{j=1}^{m}\left(\text{DISTANCE BTWN } x^j \text{ AND } V\right)^2$$

PICTURE

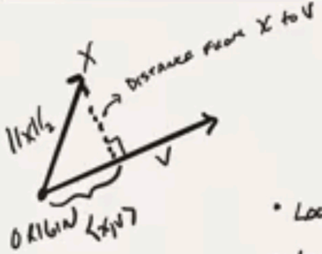


REGRESSION



PCA

$$\boxed{\min_{v,\ \|v\|_2 = 1} \frac{1}{m}\sum_{j=1}^{m} \left(\text{DISTANCE BTWN } x^j \text{ and } v\right)^2}$$



$$\underline{\langle x, v\rangle^2} + (\text{DIST FROM } x \text{ to } v)^2 = \|x\|^2$$

FIXED

- LOOKING FOR V SMALL
- LOOK FOR A V TO MAKE $\langle x,v\rangle^2$ LARGE.

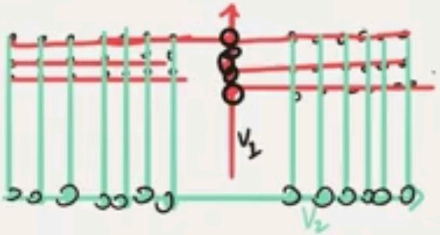
EQUIVALENTLY, WE WANT TO FIND A V THAT MAXIMIZES $\langle x,v\rangle^2$

$$\boxed{\text{FIND } v,\ \|v\|_2 = 1 \qquad \max \underline{\frac{1}{m}\sum_{j=1}^{m} \langle x^j, v\rangle^2}}$$

DIRECTION OF MAXIMAL VARIANCE.

VAR $\underline{E[x^2]} - (E[x])^2$

0



MAIN IDEA VISUALLY

THIS WAS FOR 1 VECTOR   WHAT ABOUT FINDING K-VECTORS
K-COMPONENTS

MAX SUBSPACES S OF DIMENSION K $\frac{1}{m}\sum_{j=1}^{m}$ (LENGTH OF x^j PROJECTED ONTO $S)^2$

A REALLY NICE/PREFERABLE BASIS WOULD BE AN ORTHONORMAL BASIS $\underline{v_1, \dots, v_k}$

$$(\text{DISTANCE FROM } x \text{ to } S)^2 = \|x\|^2 - [\langle x, v_1\rangle^2 + \dots + \langle x, v_k\rangle^2]$$

$$\boxed{\max_{\substack{v_1,\dots,v_k \\ \text{ORTHOGONAL}}} \frac{1}{m}\sum_{j=1}^{m}\sum_{i=1}^{k}\langle x^j, v_i\rangle^2} \longleftarrow \boxed{\text{PCA OBJECTIVE}}$$

ASSUME WE HAVE $V_1, \ldots, V_k$

$$\underline{X} = \underline{\langle x, v_1 \rangle \cdot V_1} + \underline{\langle x, v_2 \rangle \cdot V_2} + \ldots + \underline{\langle x, v_k \rangle \cdot V_k}$$

X CAN BE WRITTEN AS A VECTOR IN $\mathbb{R}^k$ CORRESPONDING TO THESE PROJECTIONS.

APPLICATION 1: UNDERSTANDING GENOMES

TOOK 1400 PEOPLE FROM EUROPE

EACH PERSON WAS REPRESENTED ACCORDING 200,000 GENETIC MARKERS IN THEIR GENOME.

CORRESPONDS TO MATRIX OF DIM $1400 \times 200{,}000$

- RAN PCA ON THIS DATA TO FIND VECTORS V_1 and V_2
- EACH PERSON CORRESPONDS TO 2 NUMBERS.

THEY PLOTTED THESE 2 NUMBERS; COLOR CODE EACH POINT ACCORDING TO COUNTRY OF ORIGIN.

ANOTHER APPLICATION:

IMAGE DATA COMPRESSION

STRATEGY FOR COMPRESSING DATA:

EACH DATA POINT IS AN IMAGE (VECTOR OF PIXELS)

EACH IMAGE HAS 65,000 PIXELS (65,000 FEATURES)

IMAGES OF FACES. RUN PCA ON DATA SET

$K \approx 100-150$

IMAGE $\approx$ LINEAR COMBINATION OF 150 VECTORS OF LENGTH 65,000

IS AN IMAGE!

BIG QUESTION:

HOW DO WE FIND THESE $V_1, \dots, V_k$'s ?

$$\max_{\substack{V_1, \dots, V_k \\ \text{orthonormal}}} \frac{1}{m} \sum_{j=1}^{m} \sum_{i=1}^{k} \langle x^j, V_i \rangle^2 \quad \text{IS MAXIMIZED}$$

Let X be an m by n matrix $\underbrace{\frac{1}{m} X^T X}$ (n by n MATRIX)

SAMPLE COVARIANCE MATRIX

$v \leftarrow$ a column vector

$v^T \leftarrow$ a row vector

$v^T v \leftarrow$ INNER PRODUCT (SCALAR)

$v v^T \leftarrow$ OUTER PRODUCT (MATRIX)

$(i,j)^{th}$ entry of $X^T X$ CORRESPONDS TO "HOW SIMILAR IS FEATURE i to FEATURE j?"

NOTE THAT X^TX IS A SYMMETRIC MATRIX

$X^TX \begin{pmatrix} ij \end{pmatrix}$ = INNER PRODUCT OF ROW i OF X^T WITH COLUMN j OF X

" COLUMN i OF X WITH COLUMN j OF X

= COLUMN j OF X WITH " i of X

<u>FACT:</u> ALL EIGENVALUES OF SYMMETRIC MATRICES ≥ 0.

FOR A MATRIX A, v IS AN EIGENVECTOR IF $A \cdot v = \lambda \cdot v$ $\lambda \in \mathbb{R}$

EIGENVALUE.

<u>DEFINITION:</u> AN ORTHOGONAL MATRIX IS ONE WHERE ALL COLUMNS ARE ORTHONORMAL. $\Leftrightarrow$

$A^TA = I$ $(AA^T = I)$

$\boxed{\text{SPECTRAL THM:}}$ EVERY SYMMETRIC MATRIX A HAS AN EIGENDECOMPOSITION:

$$A = \underbrace{Q}_{\text{ORTHOGONAL MATRIX}} \cdot \underbrace{D}_{\text{DIAGONAL MATRIX}} \cdot Q^T$$

ENTRIES OF D ARE THE EIGENVALUES OF A.

Let's try to compute V_1.

X is matrix corresponding to S.

X is m by n

X is m by n $\qquad X \cdot v = \begin{bmatrix} \langle x_1, v \rangle \\ \vdots \\ \langle x_m, v \rangle \end{bmatrix}$ ← ROWS OF X

$$v^T X^T X v = \underline{(Xv)}^T \cdot \underline{(Xv)} = \sum_{i=1}^{m} \langle x_i, v \rangle^2$$

FIND A v THAT MAXIMIZES $\qquad$ FIND v THAT MAXIMIZES $v^T \underbrace{(X^T X)}_{A} \cdot v$

"MAXIMIZING A QUADRATIC FORM"

MAXIMIZE $v^T A \cdot v$ "MAX QUADRATIC FORM"

$v, \|v\|_2 = 1$

LET'S LOOK AT A SIMPLE CASE: A IS DIAGONAL.

$$A = \begin{pmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \\ & & & \lambda_n \end{pmatrix}$$

$\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n \geq 0.$

PICK $v = (1, \cdots, 0)$

$A = X^T X$

$$v^T A v = v_1, \ldots, v_n \cdot \begin{pmatrix} \lambda v_1 \\ \vdots \\ \lambda_n v_n \end{pmatrix} = \sum_{i=1}^{n} v_i^2 \cdot \lambda_i$$

DIAGONAL

WE DON'T KNOW IF A IS DIAGONAL IN GENERAL $A = Q \cdot D Q^T$

A IS "ALMOST" DIAGONAL. $e_1 = (1, 0 \ldots 0, 0)$

CHOOSE $v = (Q \cdot e_1)$ MAXIMIZES TOP EIGENVECTOR OF A

LAST TIME: WE ALSO DISCUSSED THE "EASY CASE"

WHEN A WAS A DIAGONAL MATRIX

$\begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$

MAPS UNIT CIRCLE INTO AN ELLIPSE.

RECALL FROM LINEAR ALGEBRA: ROTATION MATRICES

ORTHOGONAL MATRICES

FOR EXAMPLE $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ ROTATE θ degrees COUNTERCLOCKWISE

$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$ ROTATE YOUR AXES θ degrees CLOCKWISE

$\begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$ THE SOLUTION TO $\max_{v,\ \|v\|_2=1} v^T \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} v$

$v = (1, 0)$

$\underline{v^T A v} = \underline{\sum_{i=1}^{n} v_i^2 \lambda_i}$ $\quad \sum v_i^2 = 1$ CORRESPONDS TO CHOOSING 3

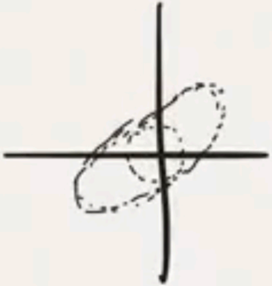
A IS DIAGONAL

$\lambda_1, \dots, \lambda_n \geq 0$

$v = (1, 0)$

FOR EXAMPLE

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = \underbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}}_{\text{ROTATE COUNTER CLOCKWISE}} \cdot \underbrace{\begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}}_{\text{STRETCHING}} \cdot \underbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}}_{\text{ROTATION } 45^\circ \text{ CLOCKWISE}}$$



SPECTRAL THM: ANY SYMMETRIC MATRIX CAN BE WRITTEN AS QDQ^T WHERE Q IS ORTHOGONAL AND D IS DIAGONAL WITH REAL $\underbrace{\text{VALUES ON THE DIAGONAL}}_{\text{EIGENVALUES}}$.

FURTHERMORE: IF $A = \underline{X^T X}$ THEN ALL EIGENVALUES ≥ 0.

CLAIM 1: FOR ANY v, $v^T A v \geq 0$

BECAUSE $A = X^T X$ $\quad v^T A v = (\underline{Xv})^T \cdot \underline{Xv} \geq 0$.

CLAIM 2 A CANNOT HAVE NEGATIVE EIGENVALUES. (PROOF IS BY CONTRADICTION)

Let's assume by contradiction that $\lambda_i < 0$ (i^{th} eigenvalue is negative)

$A = QDQ^T$ Let's consider vector $Q \cdot e_i$ $\quad (0\,0\,0\,0\,0\,\underset{i^{th}}{1}\,0\,0\,0\,0\,0) = e_i$

$v = Q \cdot e_i$ $\quad v^T A v$

$$e_i^T \underbrace{\cancel{Q^T Q}}_{} D \underbrace{\cancel{Q^T Q}}_{} e_i = \underline{e_i^T} \underline{D e_i} < 0$$

TO RECAP:

PCA:

1) SUBTRACT THE MEAN FROM YOUR DATA
2) NORMALIZE THE COLUMNS OF YOUR DATA
3) COMPUTE EIGENVALUE/EIGENVECTOR DECOMPOSITION OF YOUR MATRIX

$$QDQ^T$$

4) THE FIRST K ROWS OF Q^T ARE THE K EIGENVECTORS YOU'RE LOOKING FOR

TOP K PRINCIPAL COMPONENTS.

PROVE i^{th} ROW OF Q^T IS AN EIGENVECTOR OF A

i^{th} ROW OF Q^T = $Q \cdot e_i$

$$\underline{A} \cdot \underbrace{Q \cdot e_i} = QDQ^TQ \cdot e_i = QDe_i$$
$$= \underbrace{\lambda_i} \cdot \underbrace{Q \cdot e_i}$$

HOW DO WE COMPUTE THIS DECOMPOSITION?

ANOTHER PROBLEM "SINGULAR VALUE DECOMPOSITION"
SVD

KNOWN: POLYNOMIAL-TIME ALGORITHMS FOR COMPUTING SVD.