VECTOR BASICS

1) Dot product
defined as

$$x^T y$$

$$\underset{(1 \times n)}{\begin{bmatrix} a & b & c & \dots \end{bmatrix}} \underset{(n \times 1)}{\begin{bmatrix} p \\ q \\ r \\ \vdots \end{bmatrix}} = (a \times p) + (b \times q) + (c \times r) + \dots$$

(1)

2) Length
defined as

$$\sqrt{x^T x}$$

3) Distance between two vectors x, y

$$|x - y| = \sqrt{(x-y)^T (x-y)}$$

4) inner product generalises the notion of dot product

$$\langle x, y \rangle$$

5) Length

$$\|x\| = \sqrt{\langle x, x \rangle}$$

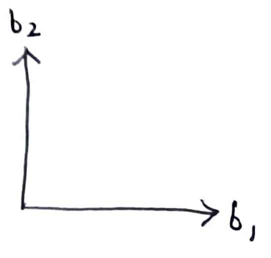
6) Two vectors b_1 & b_2 form an orthonormal basis
if

$$\langle b_1, b_2 \rangle = 0$$

and length is

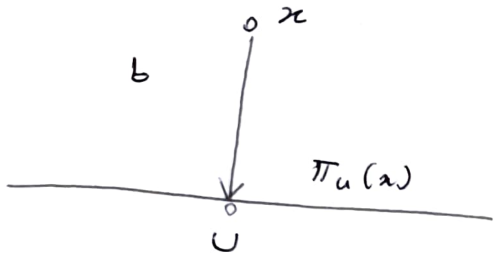
$$\| b_1 \| = 1$$

$$\| b_2 \| = 1$$



SOUMYA BANERJEE

ADVANCED PCA NOTES

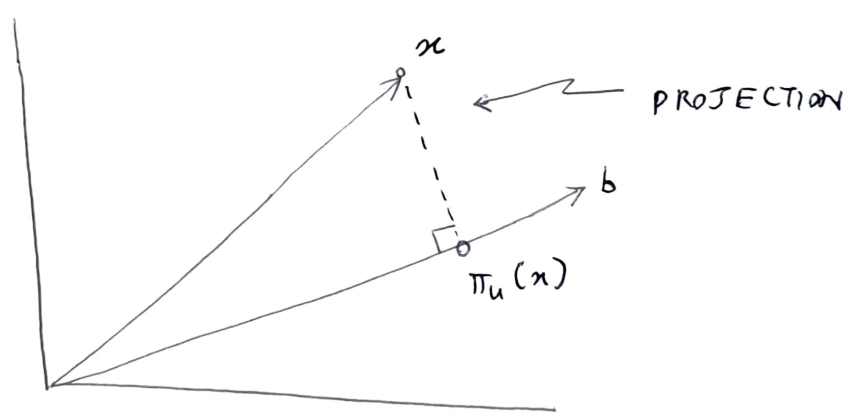


$$\min \|x - \pi_u(x)\|$$

b, $x - \pi_u(x)$ orthogonal.

inner product

$$\langle b, x - \pi_u(x) \rangle = 0$$



$\pi_u(x)$ is a vector.

CONCEPT

projection $\pi_u(x)$ is "closest" ~~to~~ to x

$\|x - \pi_u(x)\|$ is minimized

$$\Rightarrow \boxed{\langle b, x - \pi_u(x) \rangle = 0}$$

$$\boxed{\pi_u(x) = \lambda b}$$

$\langle b$

$$\langle b, x - \pi_u(x) \rangle = 0$$

$$\langle b, x - \lambda b \rangle = 0$$

$$\langle x, b \rangle - \lambda \langle b, b \rangle = 0.$$

$$\lambda = \frac{\langle x, b \rangle}{\langle b, b \rangle}$$

if dot product

$$\lambda = \frac{b^T x}{b^T b}$$

$$\boxed{\lambda = \frac{b^T x}{\|b\|^2}}$$

orthonormal basis $\|b\| = 1$

$$\boxed{\lambda = b^T x}$$

$$\pi_u(x) = \lambda b$$

$$= b^T x b$$

$$\boxed{\pi_u(x) = b^T x b}$$

λ is the eigenvalue
b is the eigenvector

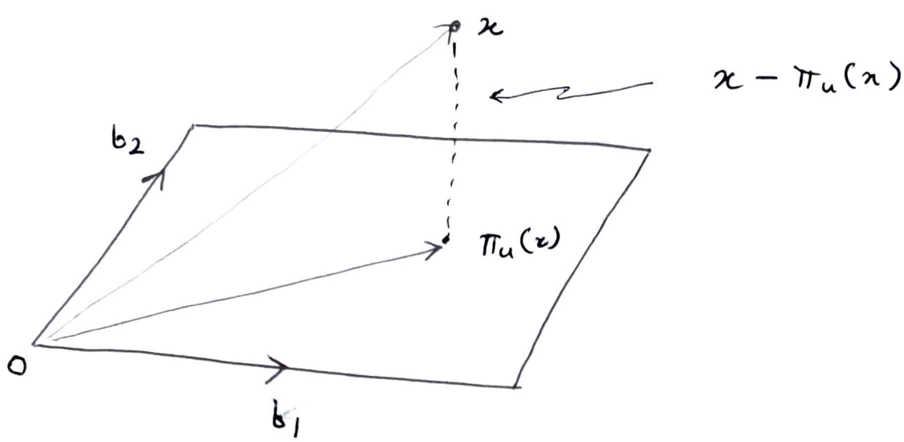
CONCEPT

eigenvalue - how much do you stretch

eigenvector - direction of transformation

PCA DERIVATIONS

This works in higher dimensions as well



"Iterative" algorithm.

Find the first principal component (eigen vector)
then second (orthogonal to that)

⋮

Generalize to n-dimensions

$\pi_u(x)$ is the projection and hence must be a linear combination of basis vectors $b_1, b_2, \ldots,$

$$\pi_u(x) = \sum \lambda_i b_i$$

$$\pi_u(x) = B\lambda \quad \ldots\ldots$$

Now we want the minimum distance.
Hence the basis vector b_1

and $\pi_u(x) - x$

must be orthogonal.

$\Rightarrow$ inner product must be 0.

$$\langle b_1, \pi_u(x) - x \rangle = 0$$

and so on for all basis vectors

$$\langle b_2, \pi_u(x) - x \rangle = 0$$

$$\langle b_3, \pi_u(x) - x \rangle = 0$$

$$\vdots$$

In matrix notation

$$\langle B, \pi_u(x) - x \rangle = 0.$$

But $\pi_u(x) = B\lambda$

Hence

$$\langle B, B\lambda - x \rangle = 0.$$

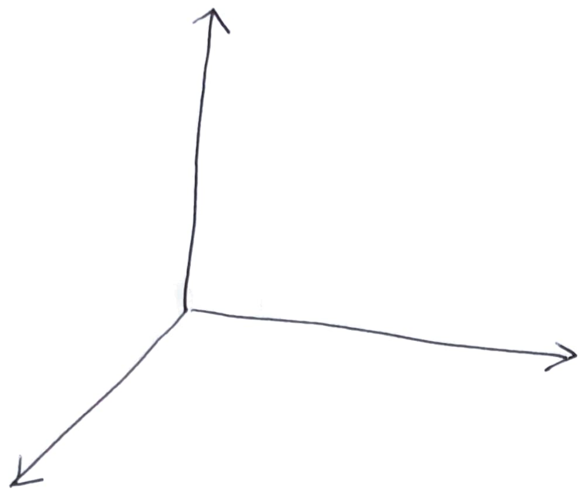
Assuming dot product,

$$B^T(B\lambda - x) = 0.$$

$$B^T B\lambda - B^T x = 0$$

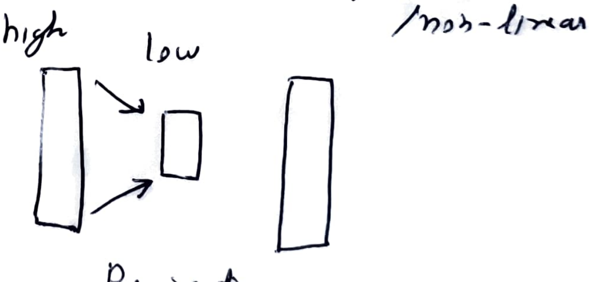
$$B^T B\lambda = B^T x$$

$$\boxed{\lambda = (B^T B)^{-1} B^T x}$$ eigen value



iterative algorithm

orthogonal to each other



Coder / Decoder

Reconstruction error / loss

→ RECOMMENDED VIEWING

Video by 3 blue 1 brown on eigenvectors

→ see page 10, 11 of notes here for INTUITION

PCA DERIVATION 1

The covariance between two random variables is the
expected value of the product of deviation from their means

Let x_n be data points with mean 0.

The data covariance matrix is

$$S = \frac{1}{N} \sum_{n=1}^{N} x_n x_n^T$$

The data points can be compressed / projected onto a lower dimensional space z_n

$$z_n = B^T x_n$$

$\downarrow$

projection matrix

$$B = [b_1, b_2, \dots]$$

each column has vectors

orthogonal

$$b_i^T b_i = 1$$

$$b_i^T b_j = 0 \quad (i \neq j)$$

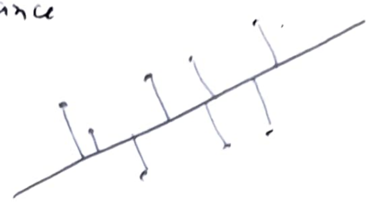
CONCEPT

PCA assumes a linear relationship between data point and lower dimensional representation

$$z_n = B^T x_n$$

We aim to maximize the variance

$$V = \frac{1}{N} \sum_{n=1}^{N} z_{1n}^{2}$$



$$z_{1n} = b_1^T x_n$$ projected co-ordinate

Maximize

$$V = \frac{1}{N} \sum (b^T x_n)^2$$

PCA uses the dot product (symmetric)

$$b_1^T x_n = x_n^T b_1$$

$$V = \frac{1}{N} \sum_{n=1}^{N} b_1^T x_n \, x^{nT} b_1$$

b_1 does not depend on

$$= \frac{1}{N} \; b_1^T \underbrace{\left(\sum_{n=1}^{N} x_n \, x^{nT} \right)} \; b_1$$

$$= b_1^T \left(\frac{1}{N} \sum_{n=1}^{N} x_n \, x^T \right) b_1$$

$$V = b_1^T S \, b_1$$

$$S = \frac{1}{N} \sum_{n=1}^{N} x_n \, x^T$$

find b_1 such that $b_1^T S \, b_1$ is maximized.

CONCEPT

$$\boxed{S b_1 = \lambda_1 b_1}$$

$b_1^T b_1 = 1$ length 1

We choose basis vector corresponding to largest eigenvalue of data covariance matrix (S). λ represents variance explained by first co-ordinate. $\sqrt{\lambda} \rightarrow$ loading

$$V_1 = \frac{1}{N} \sum_{n=1}^{N} z_{1n}^2$$

maximize
variance

$$= b_1^T S b_1$$

$$S b_1 = \lambda_1 b_1$$

$$V_1 = \cancel{b_1^T \lambda b}$$
$$\lambda_1 b_1^T b_1$$

$$= \lambda_1 \qquad (b_1^T b_1 = 1)$$

$$\boxed{V_1 = \lambda_1}$$

CONCEPT

standard deviation of data accounted for by the ~~first~~ first principal component	←	eigenvalue and it captures the amount of variance in the data

eigenvector
direction of transformation

WATCH

- video by 3blue 1 brown youtube
  - eigenvector
  - linear transformations
- rotation demonstration cricket ball

YET ANOTHER DERIVATION

- see mathematics_data_science.pdf

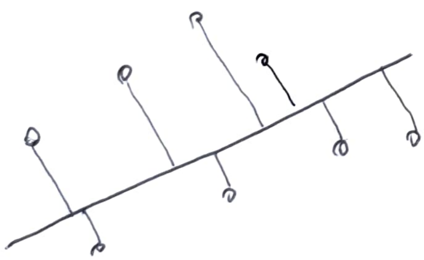
NOTES

page 16.

see the repository

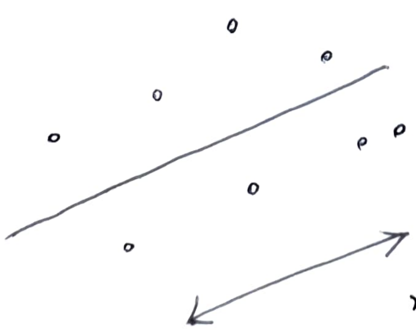
https://github.com/neelsoumya/public_teaching_unsupervised_learning

- application to LLMs. (see videos)



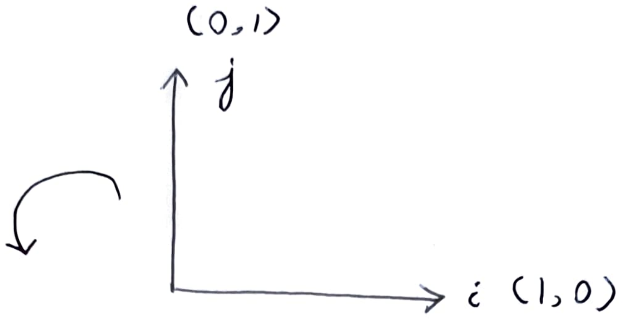
minimize these residuals

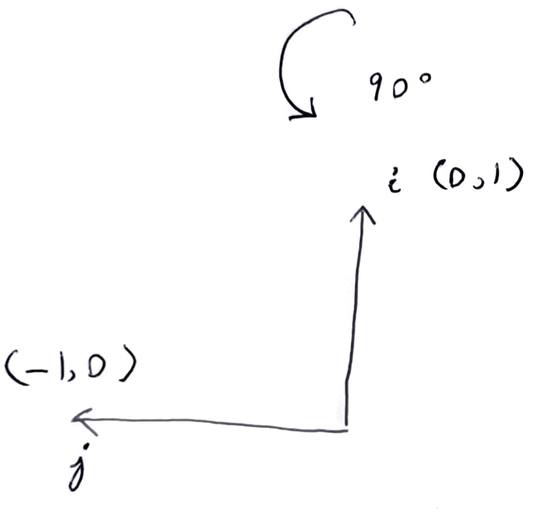
OR



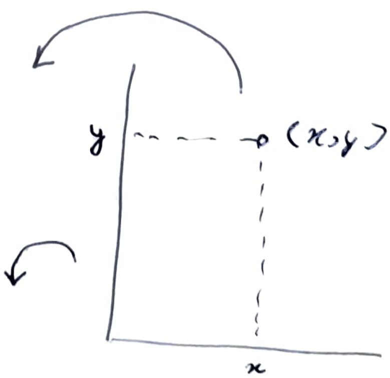
maximize the variance

MATRIX MULTIPLICATION AS LINEAR TRANSFORMATION



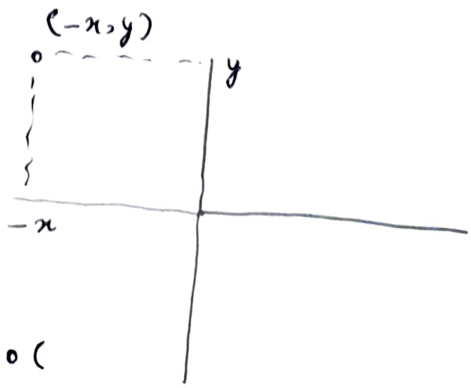


$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

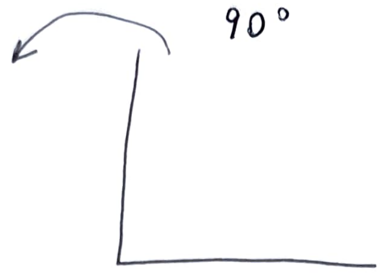


$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$$= \begin{bmatrix} -y \\ x \end{bmatrix}$$



What are the eigen vectors
and eigen values?



- any line or vector on the plane
  which does not change?
- outside plane?
- cricket ball rotation metaphor.