Linear Algebra Review

Vectors $\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$ feature vector $x \in \mathbb{R}^d$

Algebra

geometric

$[1, 2, 3, 4]$ Array / List

Computation

x - Column vector $(d, 1)$ shape

x^T - Row vector

$D = \{x_1 \dots x_d\}$ $\quad X = \begin{bmatrix} | & & | \\ x_1 & \dots & x_n \\ | & & | \end{bmatrix}$

data matrix

shape (d, n)

$2e_1 - 5e_2$

Basis $\mathbb{R}^d$

$e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ $e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ — not unique
— $\{e_1, e_2\}$ standard basis

represent any vector as algebraic equation

$\mathbb{R}^d$: Dot product $\quad x_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ $x_2 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$

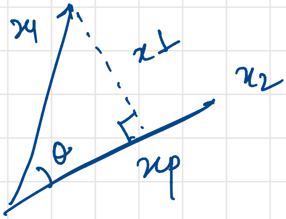
$x_1 \cdot x_2 = 3 \times 1 + 2 \times 1 = 5$

$x_1 \cdot x_2 = x_1^T x_2$ $\quad [1 \ 2] \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 1 \times 3 + 2 \times 1$
$= 5$

Projections $\mathbb{R}^d$

$x_p = c \cdot x_2$

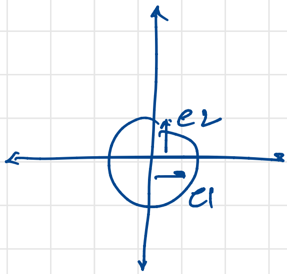
$x_1 = x_p + x_\perp$



$\mathbb{R}^d$: Orthonormal bias

- orthogonal to each other
- unit norm

Basis - 1



$e_1^T e_2 = 0$

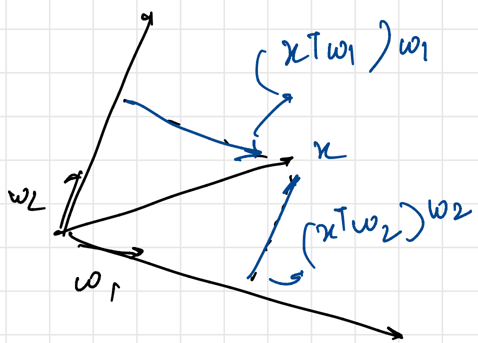
$\|e_1\| = \|e_2\| = 1$

Basis - 2

$w_1^T w_2 = 0$

$w_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad w_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \|w_1\| = \|w_2\| = 1$

Basis 1 $\xleftrightarrow[\text{reflection}]{\text{rotation}}$ Basis 2



Machine Learning Techniques | Applications

Computer Vision

Data driven

Speech → unstructured data

Text → semantics

Generalization + Mathematics

Paradigms of Machine Learning

Supervised

- Classification
- Regression
- Ranking
- Structure Learning

Unsupervised

- Clustering
- Representation Learning

Sequential Learning / Reinforcement

- online learning
- multi armed bandits
- Reinforcement Learning

UNSUPERVISED

↳ REPRESENTATION LEARNING

Data points → vectors in $\mathbb{R}^d$ $\begin{bmatrix} \text{height} \\ \text{weight} \\ \text{age} \end{bmatrix} \in \mathbb{R}^3$

Input $\{x_1, x_2 \dots x_n\}$ $x_i \in \mathbb{R}^d$ ← # d no. of features

output " some understandable representation of dataset "

ex $\left\{ \begin{bmatrix} -7 \\ -14 \end{bmatrix} \begin{bmatrix} 2.5 \\ 5 \end{bmatrix} \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\}$

for compression — 8

Representation $\underbrace{\begin{bmatrix} 1 \\ 2 \end{bmatrix}}$ 6

co-efficients

$\{\underbrace{-7, 2.5, 0.5, 0}\}$

we can reconstruct the dataset exactly

min c length2 (error vector)

↳ $\begin{bmatrix} x_1 - cw_1 \\ x_2 - cw_2 \end{bmatrix}$

$$(x_1 - cw_1)^2 + (x_2 - cw_2)^2$$

$$c^* = \left(\frac{x_1 w_1 + x_2 w_2}{w_1^2 + w_2^2} \right) \text{(c scalar)}$$

$$c^* = (x^T w) \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

1	2	−3	0
2	4	−6	0

(1) minimizing reconstruction error

(2) Different forms of objective function

(3) Covariance Matrix

Reconstruction error for best line is least
↳ finding that line

Dataset $\{x_1 \dots x_n\}$ $x_i \in \mathbb{R}^d$

$$\text{error}(\text{line}, \text{dataset}) = \sum_{i=1}^{n} \text{error}(\text{line}, x_i)$$

represent using w
$\|w\|^2 = 1$

$$= \sum_{i=1}^{n} \text{length}^2\left(x - (x^T w) \cdot w\right)$$

$$= \sum_{i=1}^{n} \|x - (x^T w) \cdot w\|^2$$

$\because \|z\|^2 = z^T z$

$$f(w) = \frac{1}{n} \sum_{i=1}^{n} \|x - (x^T w) \cdot w\|^2$$

we want to minimize this function

$$= \frac{1}{n} \sum_{i=1}^{n} \left(x_i - (x_i^T w) \cdot w\right)^T \left(x_i - (x_i^T w) \cdot w\right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left[x_i^T x_i - \cancel{(x_i^T w)^2} - (x_i^T w)^2 + \cancel{(x_i^T w)^2} \cdot 1 \right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left(\underbrace{x_i^T x_i}_{\text{constant so we can remove}} - (x_i^T w)^2 \right)$$

minimization only depends value of w

$$\min_{\substack{w \\ \|w\|^2=1}} \quad g(w) = \frac{1}{n}\sum_{i=1}^{n} -(x_i^T w)^2$$

$$\max_{\substack{w \\ \|w\|^2=1}} \quad \frac{1}{n}\sum_{i=1}^{n} (x_i^T w)^2 = \frac{1}{n}\sum_{i=1}^{n} \underset{1\times d}{(w^T} \underset{d\times 1}{x_i)} \underset{1\times d}{(x_i^T} \underset{d\times 1}{w)}$$

$$= \frac{1}{n}\sum_{i=1}^{n} w^T (x_i x_i^T) w$$

$$= w^T \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i^T\right)}_{C \to d\times d \text{ matrix}} w$$

equivalently

$$\max_{\substack{w \\ \|w\|^2=1}} \quad w^T C w$$

$$C = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T$$

covariance matrix

→ w is the eigenvector corresponding to the max eigenvalue of C matrix

Going 3-D for data [information in error vector

$x \in \mathbb{R}^d$

↓ find w

$(x^T w . w)$ representation as proxy

↓ Residue / error

$$\underbrace{x - (x^T w) . w}_{\text{might not be all error but "information"}}$$

$\{x_1, \dots x_n\}$ $x_i \in \mathbb{R}^d$ → $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$ $x_i' = x_i' - \mu$

(centering issues)

→ find "best" line $w_1 \in \mathbb{R}^d$

→ Replace $x_i \leftarrow x_i - (x_i^T w) w$

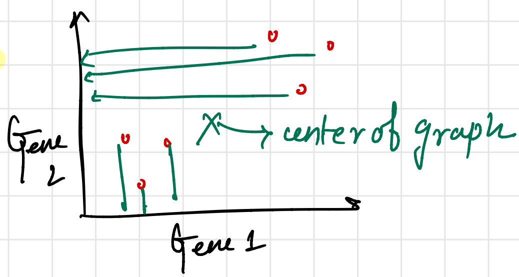
(new dataset created) only having residuals

→ Repeat procedure to obtain w_2

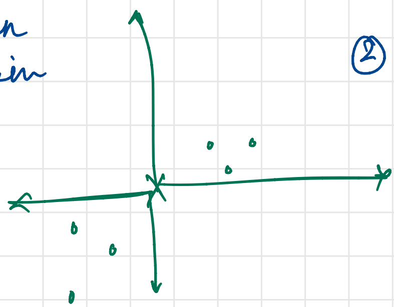
ISSUE: DATA MAY NOT BE CENTERED

Principal component Analysis

(helps plotting data for 4 gene or more)



① centering the mean to origin



② fitting a line through origin

PCA finds the best fitting line by maximizing the sum of the squared distances from the projected points to the origin