Dimesion Reduction 2: PCA

Feb 17 2022

J.L :

+ve : data independent

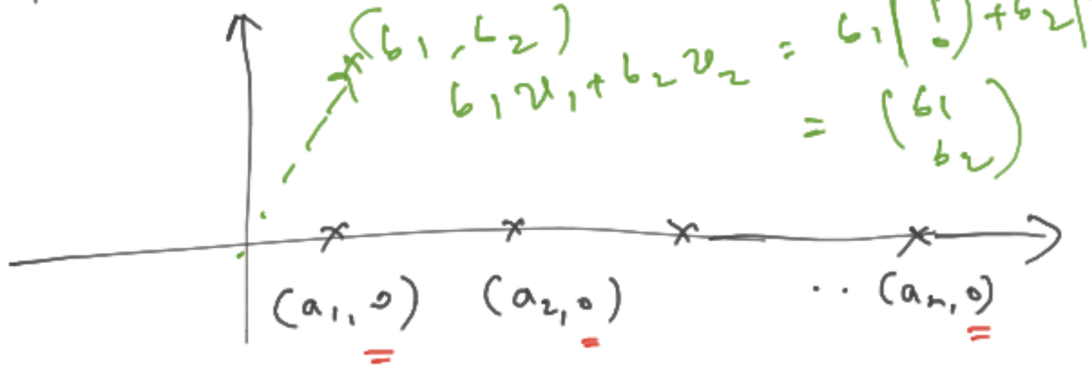
$K = O\left(\frac{\log n}{\varepsilon^2}\right)$

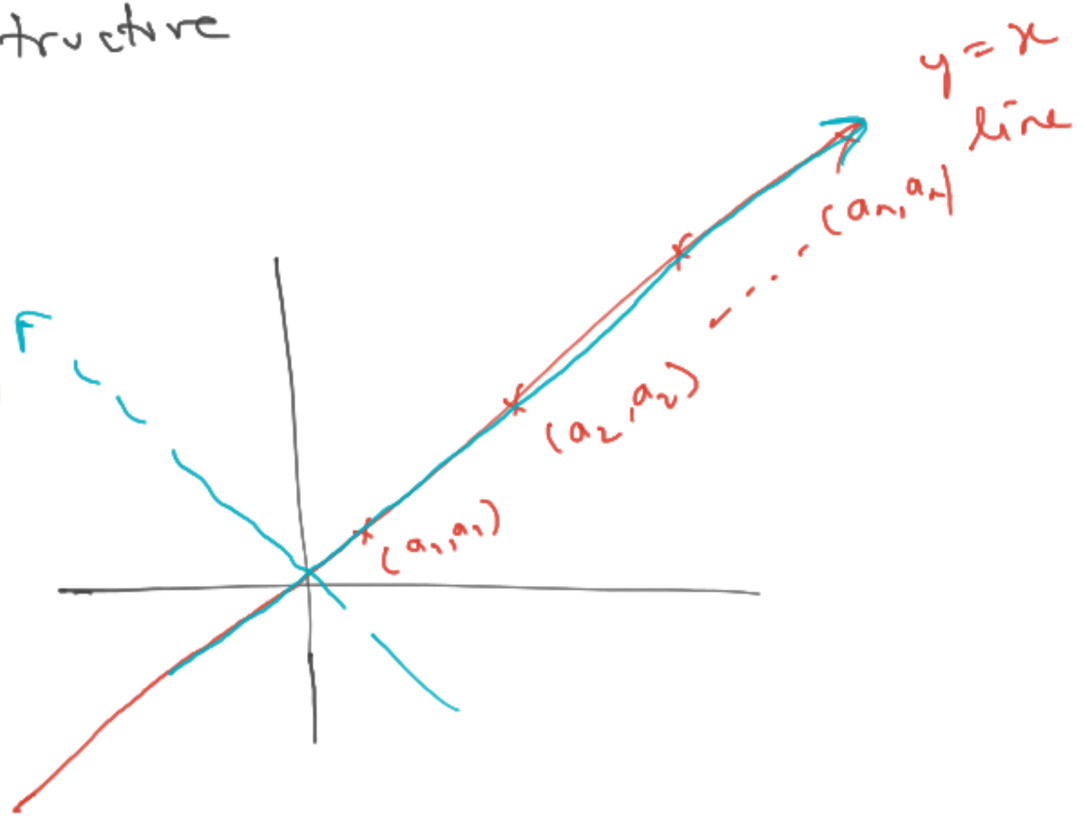
indep. of D

-ve : ——— " ———

Cannot exploit structure in data.

Toy example :

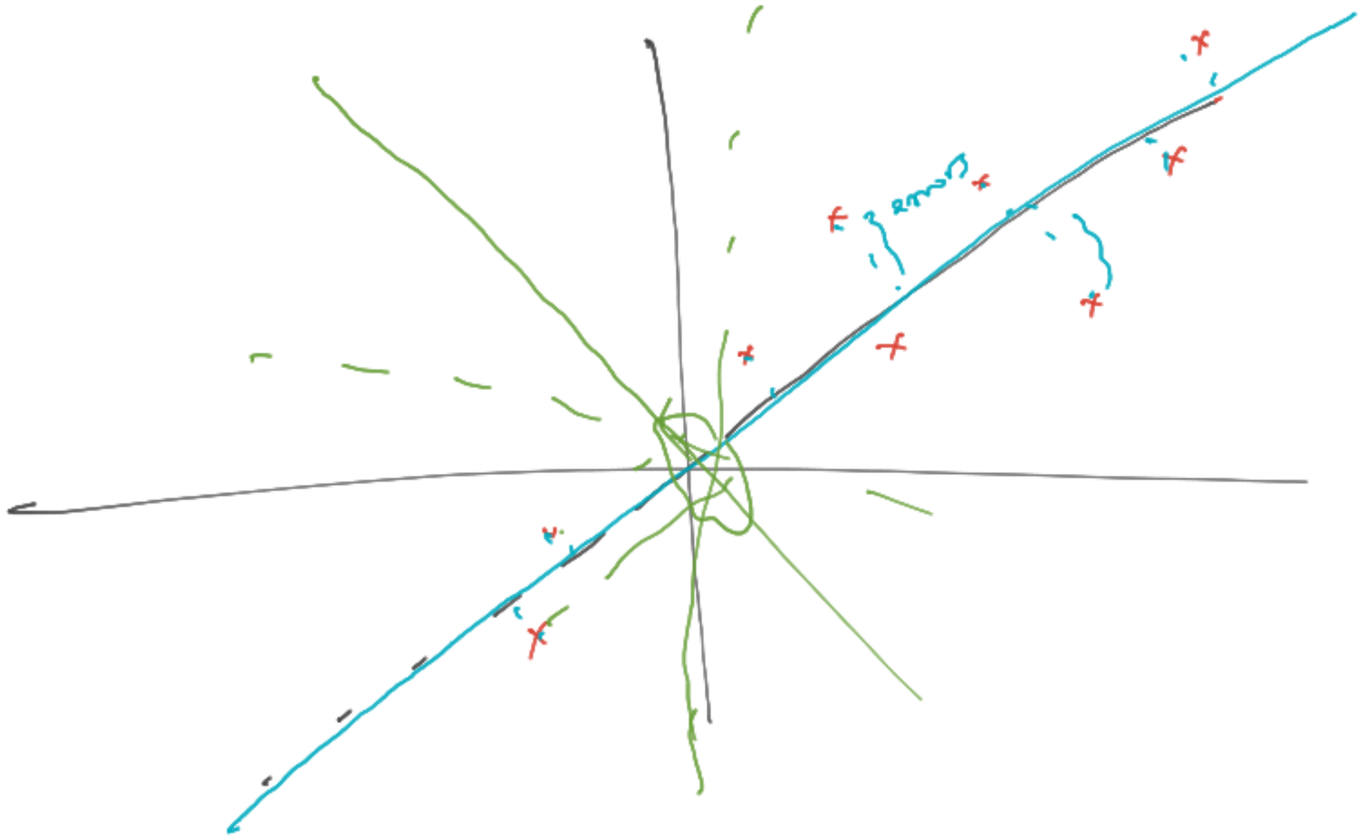




(b_1, b_2)

$v_1 = (1,0) \rightarrow x\text{-axis}$

$v_2 = (0,1) \rightarrow y\text{-axis}$



PCA / SVD:

Maximize variance/info in the "chosen" dimensions.

Applications:

1. Analysis of genome data: Gene microarray data

   original dimension $D = 10,000$ genes

   n vectors in $\mathbb{R}^D$ (individuals)

2. Denoising stock market data

Principal Component Analysis (PCA)

Goal of PCA: To find K orthonormal vectors s.t the points in the dataset have a good approximation in the subspace generated by these vectors

Recall 'basis vectors'

Good approximation: is in L2 sense

↓

(squared error)

between original points and their approximation in the low-dimensional subspace

$\{ v_1, v_2 \dots v_k \}$ orthonormal "basis" vectors

Looking at subspace spanned by these vectors $v_i \in \mathbb{R}^D$

$$\hat{\omega} = \sum_{i=1}^{k} \alpha_i v_i \in \mathbb{R}^D \qquad \omega \in \mathbb{R}^D$$

$\hat{\omega}$: Approximation of vector ω in the chosen subspace

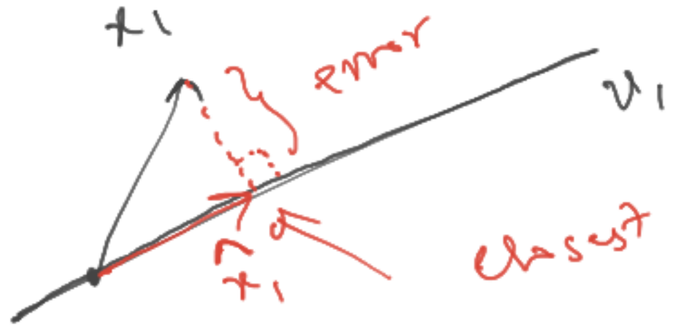
α_i: scalars

$(\alpha_1, \alpha_2, \dots \alpha_k) \in \mathbb{R}^K$

ω: arbitrary vector

$$\hat{\omega} = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_k v_k \in \mathbb{R}^D$$

Error $\| \omega - \hat{\omega} \|$



closest approx of x_1 along v_1 = projection of x_1 along v_1

$$\hat{x}_1 = (\text{scalar})\, v_1 = \langle x_1, v_1 \rangle v_1 = (x_1^T v_1) v_1$$

$\langle \cdot, \cdot \rangle$: inner product

By By thagons,

$$\|x_i\|^2 = |\text{projection}|^2 + |\text{error}|^2$$

$\|x_i\|$: fixed

$|\text{projection}|$: maximize

$|\text{error}|$: minimize

∴ Minimizing l2 error of approx.

= maximizing projected distances

= maximizing the variance / "information" in the chosen dimensions

Preprocessing :

PCA is very sensitive to scaling

⇒ Need preprocessing.

1. data zero mean (by subtracting the sample mean)
2. Scale each co-ordinate so that they are comparable
   (divide by sample deviation)

Solving PCA : 1- dim.

Given a unit vector u

a point x

| length of the projection of x onto u is $x^T u$.

n data points $x_1 \dots x_n$

Find a vector u s.t maximize the projected distances.

$$\text{maximize} \sum_{i=1}^{n} (x_i^T u)^2 = \sum_{i=1}^{n} u^T x_i \, x_i^T u$$

x_i | x_i^T

$$= u^T \left(\sum_{i=1}^{n} x_i x_i^T \right) u$$

$$= \text{maximize } u^T M u$$

M → Covariance Matrix

$$\underset{\substack{u \in \mathbb{R}^D \\ \|u_2\| = 1}}{\arg\max} \; u^T M u = \text{principal eigen vector of } M$$

K - dimensions:

Let $M = \sum_{i=1}^{n} x_i x_i^T = X^T X$ → Covariance matrix

Where $X = \begin{bmatrix} - & x_1^T & - \\ - & x_2^T & - \\ & \vdots & \\ - & x_n^T & - \end{bmatrix}$

Pos. Sem. Def (PSD) ⇒ Non-neg. eigen values

Eigen Value Decomposition (EVD) of $M = V \Lambda V^T$

Eigen values

Eigen vectors

$\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \end{bmatrix}$

Let $\lambda_1 \geq \lambda_2 \geq \lambda_3 \cdots$.

Let $v_1, v_2 \ldots, v_k$ be the first k columns of V.

Then $\{v_1, \ldots v_k\}$ are k orthonormal vectors that maximize the projected distances (i.e variance)

Transformation matrix $S = \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_k^T \end{bmatrix}$

Computing eigen values/vectors

Power Iteration:

(recall $M = \sum_{i=1}^{r} \lambda_i x_i x_i^T$)

High level steps:

1. Compute the principal eigen vector
2. Subtract out its contribution $\rightarrow M - \lambda_1 u_1 v_1^T$
3. Iterate

1. Pick a random vector y_0.
2. $y_{i+1} = \dfrac{M y_i}{\| M y_i \|_2}$ → rescaling to norm 1
3. Iterate until convergence to get (approximate) u_1
4. $\lambda_1 = v_1^T M v$ → (lin. algebra concept)

Intuition behind power iteration:

Let $y_0 = \sum_{i=1}^{k} \alpha_i v_i = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k$

$M = V \Lambda V^T$

V: orthonormal columns

$$y_1 = M y_0 = (V \Lambda V^T)(\alpha_1 v_1 + \cdots + \alpha_k v_k)$$

'diagonal'

$$= [\alpha_1 \lambda_1 v_1 + \alpha_2 \lambda_2 v_2 + \cdots + \alpha_k \lambda_k v_k]$$

$$y_t = [\alpha_1 \lambda_1^t v_1 + \alpha_2 \lambda_2^t v_2 + \cdots + \alpha_k \lambda_k^t v_k]$$

Projection of y_t along v_1

$$|\langle y_t, v_1 \rangle| = \frac{\alpha_1 \lambda_1^t}{\sqrt{\sum_{i=1}^{k} (\alpha_i \lambda_i^t)^2}}$$

Boils down to λ_1 / λ_2

When PCA does not work?

1. find "linear approximation" : $y = x^2$

   ⇒ if there are nonlinear relationships

2. If normalization is not done correctly

Excercise: Recall Relationship btwn PCA and SVD