PCA + SVD

Principal component Analysis.

[4 views]

1. → max var ← most popular
2. → min dist
3. → max R^2 ← most intuitive
4. → linear algebra view

PCA.

$x_1 \quad x_2 \ldots x_k$

$X\,[n \times k]$

n - obs,
k - predictors.

goal?
new artificial variables.

$p_1 \quad p_2 \ldots p_d$

$P\,[n \times d]$

n - obs.
d - number of new vars.
$d \leq k$

Informal: new variables should preserve as much info from original var-s as possible.

Formal approach.

[Idea 1] Remove units of meas-t of original variables.

$$\begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{bmatrix} \longrightarrow \begin{bmatrix} x^s_{11} \\ x^s_{21} \\ x^s_{31} \\ \vdots \\ x^s_{n1} \end{bmatrix}$$

$$se(x_j) = \sqrt{\frac{\sum_i (x_{ij} - \bar{x}_j)^2}{n-1}}$$

i - obs. no
j - var. no

$$x^s_{ij} = \frac{x_{ij} - \bar{x}_j}{se(x_j)}$$

from now on we assume that original variables are already standar[d]

omit the ~~superscript~~ standar "s"

$$\sum_{i=1}^{n} x_{ij} = 0 \qquad se(x_j) = 1$$

$$x_j = \begin{bmatrix} \vdots \\ \vdots \\ \vdots \end{bmatrix}$$

$$\underset{[n \times k]}{X} \xrightarrow{\text{create}} \underset{[n \times d]}{P}$$

$d \leq k$.

max-var approach to PCA.

every new variable is a linear comb-n of original [stand] variables.

$$\begin{bmatrix} \vdots \\ p_1 \\ \vdots \end{bmatrix} = v_{11} \cdot \begin{bmatrix} | \\ x_1 \\ | \end{bmatrix} + v_{21} \cdot \begin{bmatrix} | \\ x_2 \\ | \end{bmatrix} + \dots + v_{k1} \cdot \begin{bmatrix} | \\ x_k \\ | \end{bmatrix}$$

$$p_1 = X \cdot v_1$$

$X\ [n \times k]$
$v_1\ [k \times 1]$

intuit

exam1	exam2
5	2
6	2
1	2
2	2
3	2
some info	no info

formally

$$\max_{v_1} \text{ sample variance } (p_1)$$

subject to $\|v_1\| = 1$

$$v_{11}^2 + v_{21}^2 + v_{31}^2 + \dots + v_{k1}^2 = 1$$

As $x_1, x_2 \dots x_k$ are (already) centered

and $p_1 = ?x_1 + ?x_2 + \dots ?\cdot x_k$

p_1 is also centered ($\bar{p}_1 = 0$)

Step 1.

$$\boxed{\max_{\|v_1\|=1} \|p_1\|^2 \qquad p_1 = X \cdot v_1}$$

first principal component !

max $\|p_1\|$ $\Leftrightarrow$ max $se(p_1)$ $\Leftrightarrow$ max stdev(p_1)

Step 2

create p_2

$$\max_{v_2} \|p_2\|^2 \qquad p_2 = X \cdot v_2$$

s.t. $\|v_2\| = 1$ and $v_2 \perp v_1$

Step 3

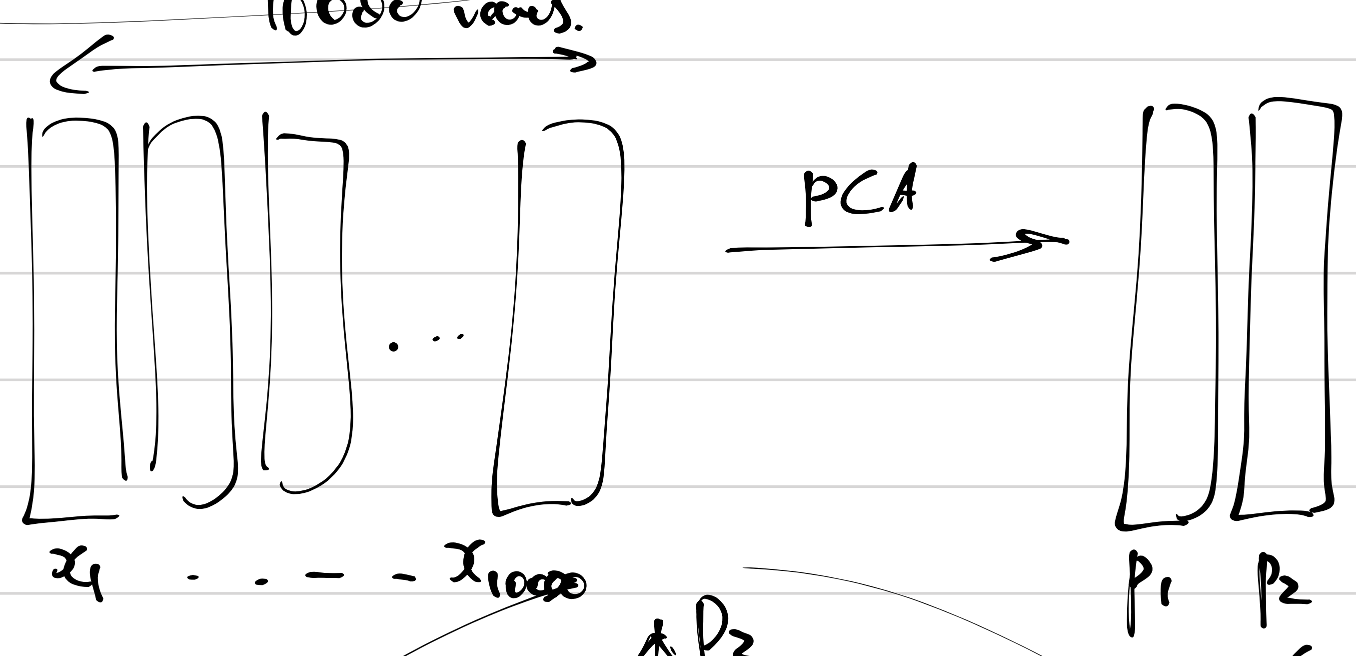
create p_3

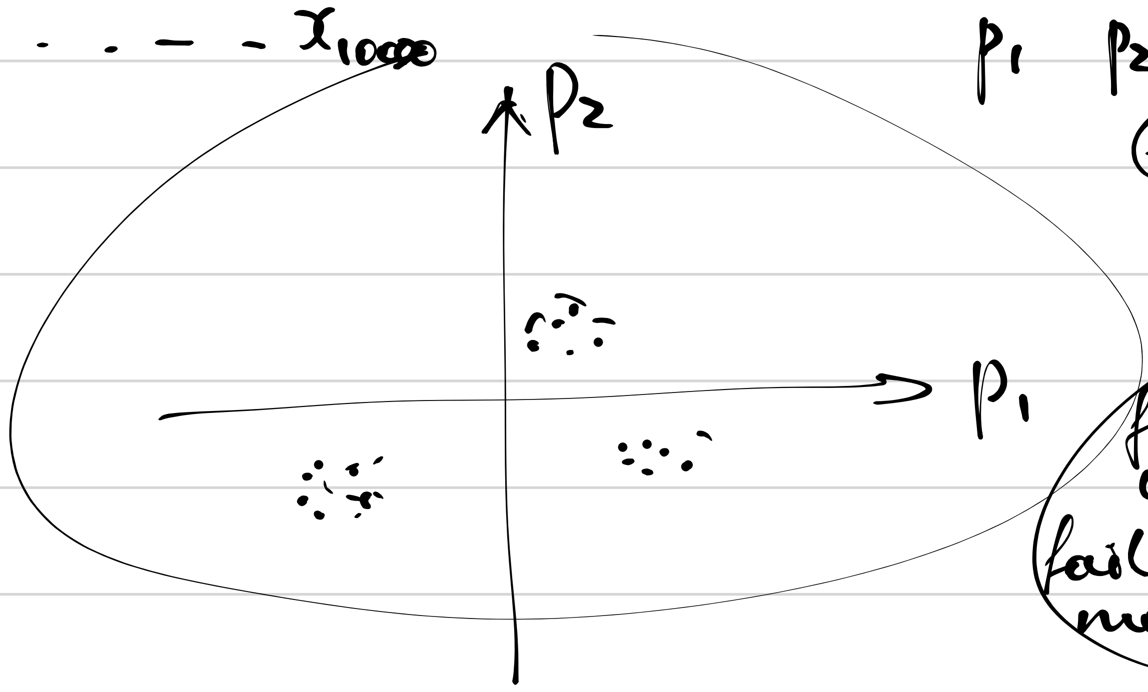
$$\max_{v_3} \|p_3\|^2 \qquad p_3 = X \cdot v_3$$

s.t. $\|v_3\| = 1$ and $v_3 \perp v_1, v_2$

. . . . .

What is the main use?





(1) visualize complex data

(2) use forecasting algo-s that fail with too many variables

Ex

$n=3$

$k=2$

$$\begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{3} & -2/\sqrt{3} \\ -2/\sqrt{3} & 1/\sqrt{3} \end{pmatrix}$$

scaled orig. variables

unscaled

$$x_1 = \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} \qquad \bar{x}_1 = 0$$

$$se(x_1) = \sqrt{\frac{1^2+1^2+(-2)^2}{3-1}} = \sqrt{\frac{6}{2}} = \sqrt{3}$$

find p_1?

$$p_1 = v_{11} \cdot \underbrace{\begin{pmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ -2/\sqrt{3} \end{pmatrix}}_{x_1} + v_{21} \cdot \underbrace{\begin{pmatrix} 1/\sqrt{3} \\ -2/\sqrt{3} \\ 1/\sqrt{3} \end{pmatrix}}_{x_2}$$

$\max\limits_{v_1} se^2(p_1)$

$\max\limits_{\|v_1\|=1} \|p_1\|^2$

$$\|p_1\|^2 = p_1^T \cdot p_1 = (v_{11} \cdot x_1^T + v_{21} \cdot x_2^T) \cdot (v_{11} \cdot x_1 + v_{21} \cdot x_2) =$$

$$= v_{11}^2 \cdot x_1^T \cdot x_1 + v_{21}^2 \cdot x_2^T \cdot x_2 + 2 v_{11} v_{21} \cdot \underbrace{x_1^T \cdot x_2}_{= x_2^T \cdot x_1} =$$

$$x_1^T x_1 = \left(\frac{1}{\sqrt{3}} \;\; \frac{1}{\sqrt{3}} \;\; -\frac{2}{\sqrt{3}}\right) \cdot \begin{pmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ -2/\sqrt{2} \end{pmatrix} = \frac{1}{3} + \frac{1}{3} + \frac{4}{3} = 2$$

$$x_1^T \cdot x_2 = \left(\frac{1}{\sqrt{3}} \;\; \frac{1}{\sqrt{3}} \;\; -\frac{2}{\sqrt{3}}\right) \cdot \begin{pmatrix} 1/\sqrt{3} \\ -2/\sqrt{3} \\ 1/\sqrt{3} \end{pmatrix} = \frac{1}{3} - \frac{2}{3} - \frac{2}{3} = -1$$

$$\|p_1\|^2 = 2 \cdot v_{11}^2 + 2 v_{21}^2 - 2 \cdot v_{11} \cdot v_{21} \rightarrow \max_{\|v_1\|=1}$$

? trick ? divide by v_{11}^2

$$\|v_1\| = 1$$
$$\|$$
$$v_{11}^2 + v_{21}^2 = 1$$

$$\|p_1\|^2 = 2\cdot v_{11}^2 + 2v_{21}^2 - 2\cdot v_{11}\cdot v_{21} =$$

$$v_{11}^2 + v_{21}^2 = 1$$

$$= \underbrace{(v_{11}^2 + v_{21}^2)}_{1} \cdot \left(2 - 2\frac{v_{11}\cdot v_{21}}{v_{11}^2 + v_{21}^2}\right) =$$

$$= 2 - 2\frac{v_{21}/v_{11}}{1 + \frac{v_{21}^2}{v_{11}^2}} = \qquad \cdot / v_{11}^2$$

$$\frac{v_{21}}{v_{11}} = t$$

$$h(t) = 2 - 2\frac{t}{1+t^2} \to \max_t$$

$$\min_t \frac{t}{1+t^2} \qquad \left(\frac{t}{1+t^2}\right)' = \frac{1(1+t^2) - 2t\cdot t}{(1+t^2)^2} =$$

$$= \frac{1+t^2-2t^2}{\cdots} = 0.$$

$$t = 1 \qquad t = -1$$

opt.

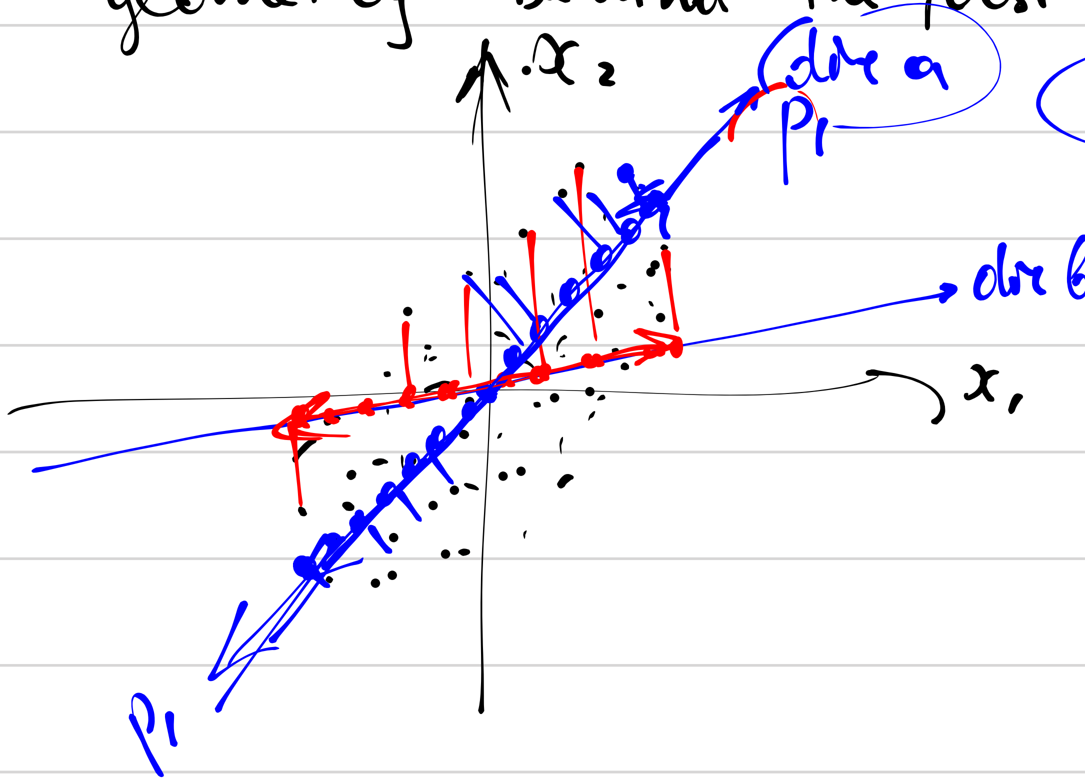
$$t^* = -1 \qquad \frac{v_{21}}{v_{11}} = -1 \qquad v_{11}^2 + v_{21}^2 = 1$$

two optima:

$$v_{11} = \frac{1}{\sqrt{2}}, \quad v_{21} = -\frac{1}{\sqrt{2}} \qquad \Bigg| \qquad v_{11} = -\frac{1}{\sqrt{2}}, \quad v_{21} = \frac{1}{\sqrt{2}}$$

$$p_1 = \frac{1}{\sqrt{2}}\cdot x_1 - \frac{1}{\sqrt{2}}\cdot x_2 \quad \text{or} \quad p_1 = -\frac{1}{\sqrt{2}}x_1 + \frac{1}{\sqrt{2}}\cdot x_2$$

geometry behind the first approach.



choose direction p_1 to maximize sample variance