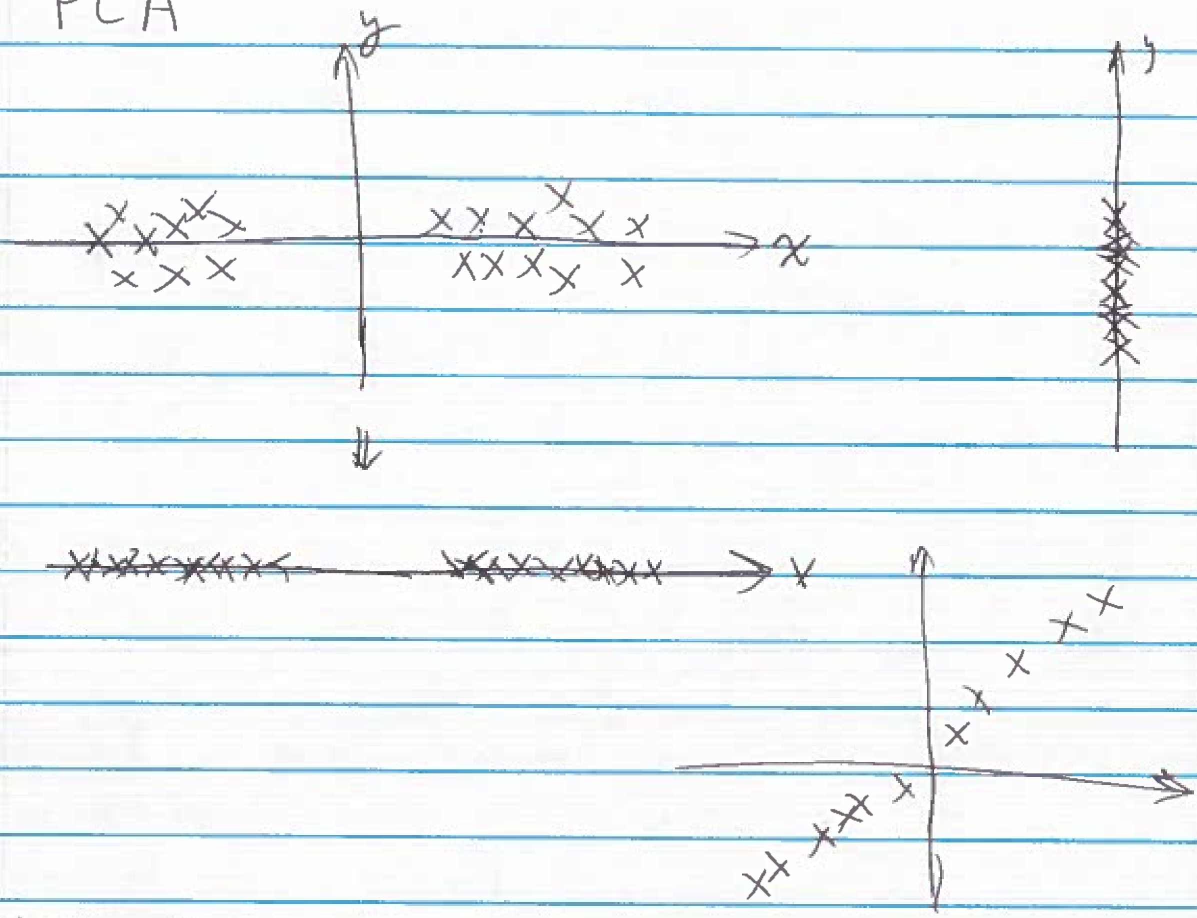
PCA



Objective

Input $D = \{x_1, x_2, \ldots, x_n\} \subset \mathbb{R}^d$

$$\text{variability} = \frac{1}{n}\sum_{i=1}^{n} \left\| x_i - \frac{1}{n}\sum_{j=1}^{n} x_j \right\|^2$$

Project D onto some ~~vector~~ subspace while "keeping" as much variability as we can.

Uses of PCA

Visualization! Use PCA to map $\mathbb{R}^d \to \mathbb{R}^2$.

Pre-processing. $\Rightarrow$ run faster!

PCA as minimizing error.

input $x_1, x_2, \dots, x_n \in \mathbb{R}^d$

find: linear transform $A: \mathbb{R}^d \to \mathbb{R}^m$
and a transform $B: \mathbb{R}^m \to \mathbb{R}^d$

s.t.

minimize $\frac{1}{n} \sum_{i=1}^{n} \| B A \hat{x}_i - \hat{x}_i \|^2$

equivalently, find $W \in \mathbb{R}^{d \times m}$ s.t. $W^T W = I$

minimize $\frac{1}{n} \sum_{i=1}^{n} \| W W^T \hat{x}_i - \hat{x}_i \|^2$

PCA solves this too!

Step 1: Remove the Mean;

Let $\hat{x}_i = x_i - \frac{1}{n}\sum_{j=1}^{n} x_j$

variability $= \frac{1}{n}\sum_{i=1}^{n} \|\hat{x}_i\|^2$

Linear projection (map to $\mathbb{R}^1$)

$x \to u^T x$ for some u s.t. $\|u\|_2 = 1$

In $\mathbb{R}^2$: projecting onto x-axis..

$$x \to x_1 \equiv x \to \begin{bmatrix}1\\0\end{bmatrix}^T x$$

$$x \to x_2 \equiv x \to \begin{bmatrix}0\\1\end{bmatrix}^T x$$

---

max variability $= \frac{1}{n}\sum_{i=1}^{n} \left(u^T \hat{x}_i\right)^2$

subject to $\|u\|_2 = 1$

note for $z \in \mathbb{R}$
$\|z\|^2 = |z|^2 = z^2$

1d PCA problem!

$$\underset{\text{over } u \in \mathbb{R}^d}{\text{maximize}} \quad \overset{J}{\overset{\|}{\frac{1}{n}\sum_{i=1}^{n} (u^T \hat{x}_i)^2}} \quad \text{subject to } \|u\|_2 = 1$$

$$J = \frac{1}{n}\sum_{i=1}^{n} (u^T \hat{x}_i)^2 = \frac{1}{n}\sum_{i=1}^{n} (u^T \hat{x}_i)(\hat{x}_i^T u)$$

$$= \frac{1}{n}\sum_{i=1}^{n} u^T (\hat{x}_i \hat{x}_i^T) u$$

$$= u^T \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} \hat{x}_i \hat{x}_i^T\right)}_{} u = u^T \Sigma u$$

$$\underset{\text{over } u \in \mathbb{R}^d}{\text{maximize}} \quad u^T \Sigma u \quad \text{subject to } \|u\|_2 = 1$$

where $\Sigma \in \mathbb{R}^{d \times d}$, symmetric, PSD.

Why symmetric? $$\left(\frac{1}{n}\sum_{i=1}^{n} \hat{x}_i \hat{x}_i^T\right)^T = \frac{1}{n}\sum_{i=1}^{n} (\hat{x}_i \hat{x}_i^T)^T$$

$$= \frac{1}{n}\sum_{i=1}^{n} (\hat{x}_i^T)^T (\hat{x}_i)^T = \frac{1}{n}\sum_{i=1}^{n} \hat{x}_i \hat{x}_i^T = \Sigma \quad \checkmark$$

Why PSD? A is PSD iff $\underset{\text{for all}}{\forall} \omega$, $\omega^T A \omega \geq 0$.

$$\omega^T \Sigma \omega = \omega^T \left(\frac{1}{n}\sum_{i=1}^{n} \hat{x}_i \hat{x}_i^T\right) \omega = \frac{1}{n}\sum_{i=1}^{n} (\omega^T \hat{x}_i)^2 \geq 0$$

$$\mathcal{L}(u)$$

Maximize $\quad u^T \Sigma u \quad$ s.t. $\|u\|_2 = 1$

$$\Updownarrow$$

Maximize $\quad \dfrac{u^T \Sigma u}{\|u\|^2} = \dfrac{u^T \Sigma u}{u^T u} \quad$ s.t. $\|u\|_2 \neq 0$. $\quad u \neq 0$

$\nabla u^T \Sigma u = \Sigma u + \Sigma^T u$
$= 2\Sigma u$

because $\dfrac{u}{\|u\|}$ is a unit vector $\left(\dfrac{u}{\|u\|}\right)^T \Sigma \left(\dfrac{u}{\|u\|}\right)$

$$\nabla \mathcal{L}(u) = \frac{u^T u(2\Sigma u) - (u^T \Sigma u)(2u)}{(u^T u)^2} = 0$$

$\nabla u^T u$
$= \nabla \|u\|^2$
$= 2u$

$$\frac{2u^T u}{(u^T u)^2} \Sigma u = \left(\frac{2u^T \Sigma u}{(u^T u)^2}\right) u$$

$$\Sigma u = \frac{u^T \Sigma u}{u^T u} \cdot u$$

let $\lambda = \dfrac{u^T \Sigma u}{u^T u}$; $\quad \Sigma u = \lambda u$

$\therefore$ u is an eigenvector of Σ.

Σ has eigenvectors $u_1, u_2, \ldots, u_d$
with corr. eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d \geq 0$.
i.e. $\Sigma u_i = \lambda_i u_i$

$$\mathcal{L}(u_i) = \frac{u_i^T \Sigma u_i}{u_i^T u_i} = \frac{u_i^T(\lambda u_i)}{u_i^T u_i} = \lambda_i$$

which eigenvector maximizes $\mathcal{L}$? $\quad u_1$

For mapping to $m > 1$, we pick the subspace spanned by the top m eigenvectors of Σ.

$$\text{variability} \approx \sum_{i=1}^{m} \lambda_i \qquad (\lambda_i \text{ eigenvectors of } \Sigma)$$

Important: In practice, use SVD.
singular value decomposition

$$\Sigma = \frac{1}{n}\sum_{i=1}^{n} \hat{x}_i \hat{x}_i^T = \frac{1}{n} \hat{X} \hat{X}^T$$

if $\hat{X}$ be the matrix w/ cols $\hat{x}_i$

eigenvectors of $\Sigma \iff$ singular vectors $\hat{X}$

$\Downarrow$

"principal components" of D

$$\text{variability of } D \approx \sum_{i=1}^{d} \lambda_i$$