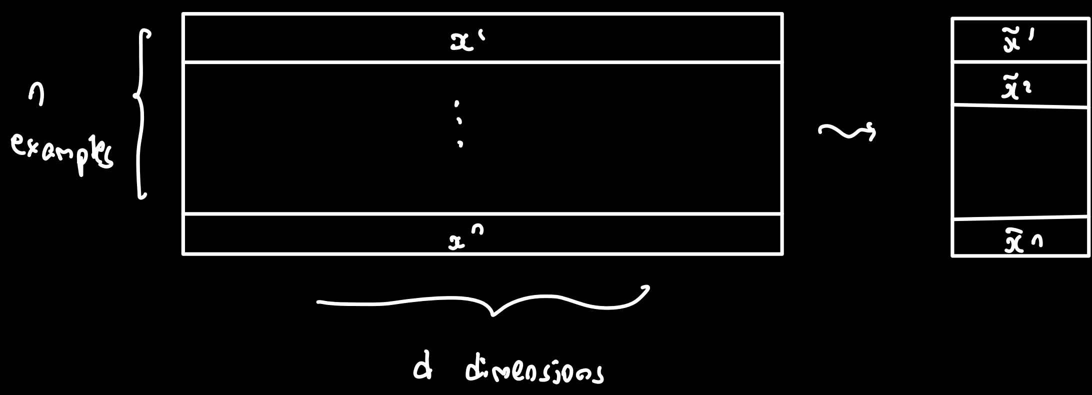
PCA:



$n \ll d$

IDEA:

We want to "represent" x as $\tilde{x}$

$x \rightarrow$ in d dimensions

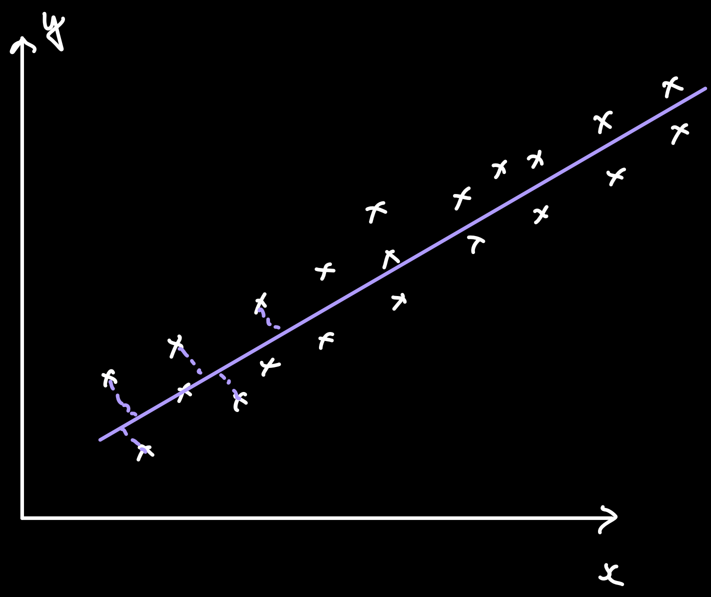
$\tilde{x}$ in low dimensions

PCA assumes

$x \curvearrowleft \tilde{x}$

is a "linear map"

Assume $d = 2$



Project to the line to get $\tilde{x}$.

Best-Fit Line Problem:

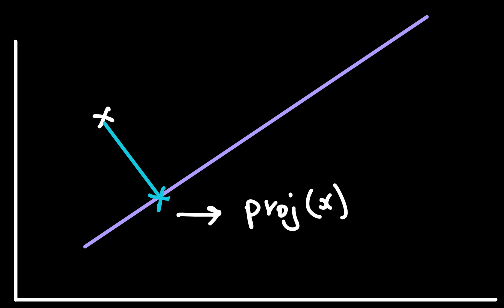
Input: $x^1, x^2 \ldots x^n \in \mathbb{R}^d$

Output: Find the line that is "closest" to the dataset.

→ Defining Closest:

Minimize aggregate distance of points to the line.

→ Distance:



Measuring distance of x to a line as length of the perpendicular to the line

$$\equiv \| x - \mathrm{proj}(x) \|$$ //

Average Distance:

$$\mathrm{Avg\,Distance}(L) = \sum_{i=1}^{n} \| x^i - \mathrm{Proj}_L(x^i) \|_2^2$$

$\mathrm{Proj}_L(x) \equiv$ closest point of x on the line L.

General Best-Fit-Subspace Problem:

Input: $X = \{x^1, x^2, \ldots x^n\}$; dimension of subspace k.

For any k-dimensional subspace $S \subseteq \mathbb{R}^d$,

$$\mathrm{ERR}(S; X) = \sum_{i=1}^{n} \| x^i - \mathrm{Proj}_S(x^i) \|_2^2$$

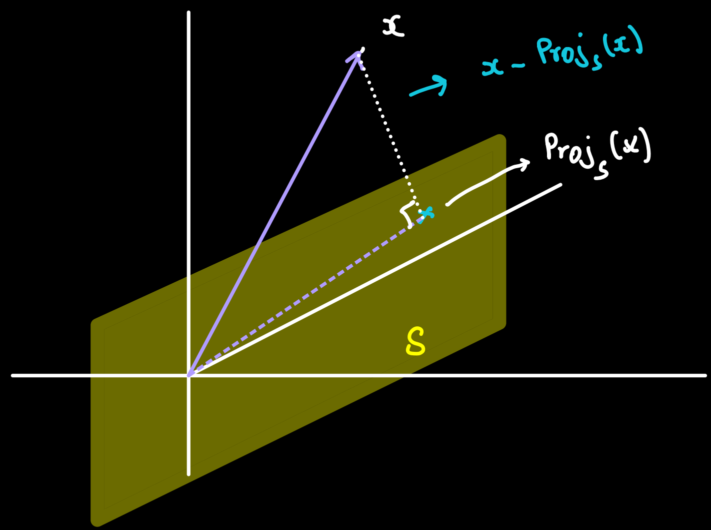
Output: Find a k-dimensional subspace that maximizes $\mathrm{ERR}(S; X)$ //

- Return an orthonormal basis for the subspace.

NEXT GOAL:

How to Solve $k=1$

LINEAR ALGEBRA FACTS:



1. $$\mathrm{Proj}_S(x) = \underset{y \in S}{\arg\min} \|x - y\|_2$$

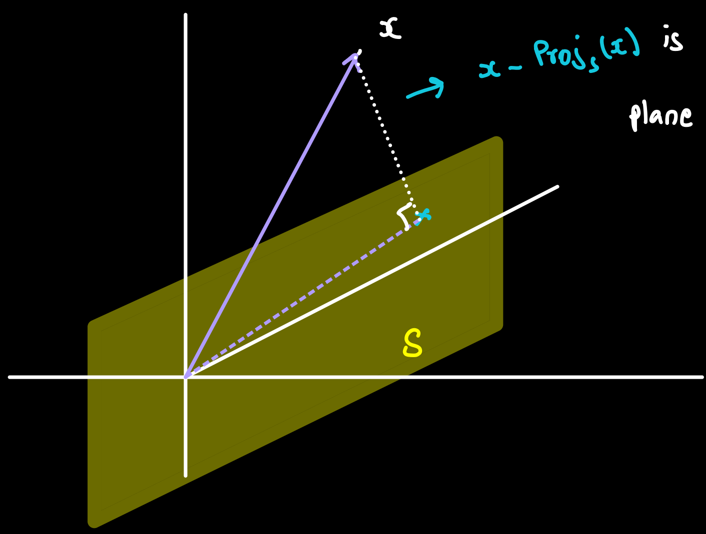
2. $$\langle x - \mathrm{Proj}_S(x), \mathrm{Proj}_S(x) \rangle = 0$$

(angle is $90°$)

3. $$\|x\|^2 = \|x - \mathrm{Proj}_S(x)\|^2 + \|\mathrm{Proj}_S(x)\|^2$$

4. For all points $u \in S$, $\langle u, x - \text{Proj}_S(x) \rangle = 0$.

($90°$)



$x - \text{Proj}_S(x)$ is perpendicular to the plane S.

INPUT : $\mathcal{X} = \{x^1, \dots, x^n\} \subseteq \mathbb{R}^d$ $\quad k = 1$

GOAL : Find 1-dimensional subspace S to minimize

$$\text{ERR}(S; \mathcal{X}) = \sum_{i=1}^{n} \| x^i - \text{Proj}_S(x^i) \|_2^2$$

by ③

$$= \sum_{i=1}^{n} \left(\| x^i \|^2 - \| \text{Proj}_S(x^i) \|_2^2 \right)$$

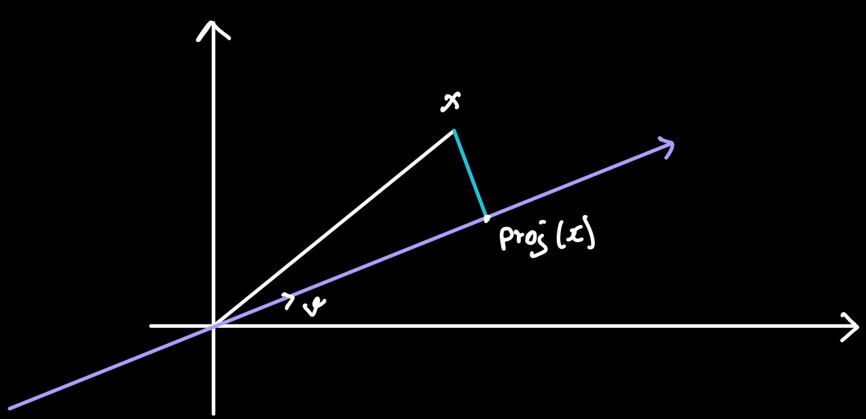
$$= \sum_{i=1}^{n} \| x^i \|^2 - \sum_{i=1}^{n} \| \text{Proj}_S(x^i) \|_2^2$$

So looking for an S to minimize $ERR(S;x)$ is the same as looking for an S to maximize

$$var(S;x) = \sum_{i=1}^{n} \| Proj_S(x^i) \|_2^2$$

Minimizing $ERR(S)$ over k-dimensional subspaces	$\equiv$	Maximizing $Var(S)$ over k-dimensional subspaces.

Focusing on $k=1$ case:



$S \equiv span\{v\}$

↳ is a unit vector

$Proj_S(x) = \langle v, x \rangle \cdot v$

If v is not a unit vector

$$\text{Proj}_S(x) = \frac{\langle v . x \rangle}{\|v\|^2} \cdot v$$

So for $k=1$ if $S = \text{span}\{v\}$

$$\text{Var}(S) = \sum_{i=1}^{n} \|\text{Proj}_S(x^i)\|^2$$

$$= \sum_{i=1}^{n} \left\| \frac{\langle v, x^i \rangle}{\|v\|^2} \cdot v \right\|^2$$

$$= \sum_{i=1}^{n} \frac{\langle v, x^i \rangle^2}{\|v\|^2}$$

$$= \frac{1}{\|v\|^2} \sum_{i=1}^{n} \langle x^i, v \rangle^2 = \frac{\|X \cdot v\|^2}{\|v\|^2}$$

$$X\left\{\begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{bmatrix}\right. \begin{bmatrix} \\ v \\ \\ \end{bmatrix} = \begin{bmatrix} \langle x^1, v \rangle \\ \langle x^2, v \rangle \\ \vdots \\ \langle x^n, v \rangle \end{bmatrix}$$

$S = \text{span}\{v\}$

$$\text{var}(S; X) = \frac{\|X \cdot v\|^2}{\|v\|^2}$$

The best-fit-line problem is equivalent to

$$\underset{v \neq 0}{\arg\max} \frac{\|X \cdot v\|^2}{\|v\|^2} \longrightarrow$$

* First Right Singular Vector of X
* First Principal Component of X.

EXAMPLE 1:

$$X = \begin{bmatrix} 10 & 0 & \cdots & 0 \\ 9 & 0 & \cdots & 0 \\ 1000 & 0 & \cdots & 0 \\ & & & \\ 1 & 0 & \cdots & 0 \end{bmatrix}$$

First PC of $X = (1, 0 \ldots, 0)$

Example 2:

$$X = \begin{bmatrix} u \\ \hline 0.1u \\ \hline 100u \\ \vdots \\ \hline 11u \end{bmatrix}$$

First PC of $X \equiv u$

Summary:

Solving for Best-Fit-Line $\equiv$ Solving for max var $\equiv$ Solving for $\underset{v : \|v\| = 1}{\arg\max} \|X \cdot v\|^2 \rightarrow$ First PC

HOW ABOUT $k=2$:

Recall: $$\underset{S:\, \dim(S)=2}{\arg\max} \sum_{i=1}^{n} \|\text{Proj}_S(x^i)\|^2$$

IDEA: To use a "greedy" approach

Find first PC and then find first PC of the "left-over".

IDEA:

1. Find First PC - v_1
2. Replace $\tilde{x}^i = x^i - \text{Proj}_{\{v_1\}}(x^i)$
3. Find first PC of the "new" dataset.

IDEA 2:

$$\underset{v:\, \|v\|=1}{\arg\max} \|X \cdot v\|^2$$

and $v \perp v_1$

(v is perpendicular to first PC)

* Second right singular vector of X
* Second principal component of X.

If $v_1, \ldots v_{i-1}$ are the first, second, third, ... $(i-1)^{th}$ PCs of x, then

$$\left.\begin{array}{c}\arg\max\limits_{\substack{v : \|v\| = 1 \\ v \perp v_1,\ v \perp v_2, \\ \ldots\ v \perp v_{i-1}}} \|X \cdot v\|^2\end{array}\right\} \begin{array}{l} i^{th} \text{ PC of } x \\ \text{or} \\ i^{th} \text{ Right singular vector of } x. \end{array}$$

THEOREM:

The span of first k singular vectors minimizes

$$ERR(S; x)$$

$$(\equiv \text{minimizes } var(S; x))$$

Remark: For $ERR_1(S; x) = \sum_{i=1}^{n} \|x^i - Proj_S(x^i)\|$, finding best-fit-line and then best-fit-line of left-over does not work.

THEOREM 1:

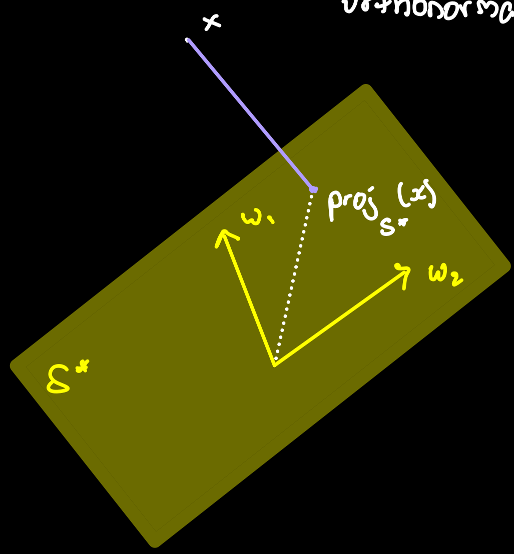
Span of first two right singular vectors maximizes

$$\max_{\dim(S) = 2} var(S; x)$$

PROOF:

$$S^* = \underset{\dim(S)=2}{\arg\max}\ \mathrm{Var}(S; x)$$

$$S^* = \mathrm{span}\{\underbrace{w_1, w_2}_{\text{orthonormal basis for } S^*}\}$$



$$v_1 := \underset{\|v\|=1}{\arg\max}\ \|x \cdot v\|$$

$$v_2 := \underset{\substack{\|v\|=1 \\ v \perp v_1}}{\arg\max}\ \|x \cdot v\|$$

v_1, v_2 are perpendicular to each other.

CLAIM:

1. $\|Proj_{S^*}(x)\|^2 = \langle x, \omega_1 \rangle^2 + \langle x, \omega_2 \rangle^2$

2. If $S = span\{v_1, v_2\}$

$$\|Proj_S(x)\|^2 = \langle x, v_1 \rangle^2 + \langle x, v_2 \rangle^2$$

$\Rightarrow$ $var(S^*; x) = \|X \cdot \omega_1\|^2 + \|X \cdot \omega_2\|^2$

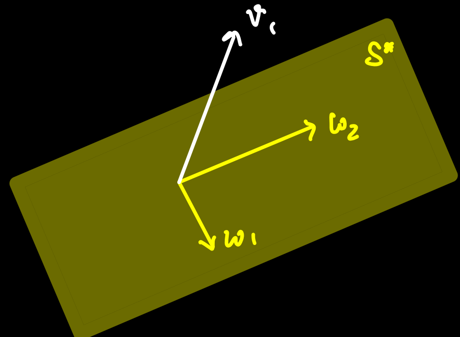
$var(S; x) = \|X \cdot v_1\|^2 + \|X \cdot v_2\|^2$

At the very least, by definition

$$\|X \cdot \omega_1\|^2 \leq \|X \cdot v_1\|^2$$

CLAIM:

I can always pick an orthonormal basis $\{\omega_1, \omega_2\}$ for S^* where $\omega_2 \perp v_1$.



$\Rightarrow$ For this basis :

$$\mathrm{var}(S^*, x) = \|x \cdot w_1\|^2 + \|x \cdot w_2\|^2$$

$$\mathrm{var}(S ; x) = \|x \cdot v_1\|^2 + \|x \cdot v_2\|^2$$

We know out of all vectors perpendicular to v_1, v_2 gives the highest variance. So,

$$\|x \cdot v_2\|^2 \geq \|x \cdot w_2\|^2$$

$$\Rightarrow \mathrm{var}(S^*, x) \leq \mathrm{var}(S; x)$$

$\Rightarrow S$ maximizes the variance.