PRINCIPAL COMPONENT ANALYSIS:

→ Highest Variance.

General Best-Fit-Subspace Problem

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

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

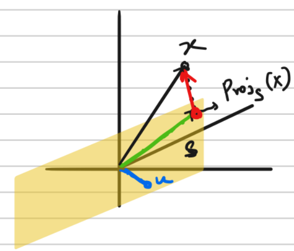
$$\mathrm{ERR}(S;X) = \sum_{i=1}^{n} \| x^i - \mathrm{Proj}_S(x^i) \|_2^2 .$$

OUTPUT: $\underbrace{\text{Find a } k\text{-dim. Subspace}}$ that minimizes $\mathrm{ERR}(S;X)$.

— Return an orthonormal basis for the Subspace.

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 $u \in S$

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

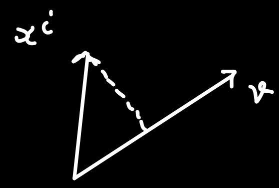
MAXIMIZE

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

Minimize ERR => Maximize Var

$$\text{Proj}_{\vec{v}_s}(x^i) = (x \cos\theta)\,\hat{v}$$

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



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

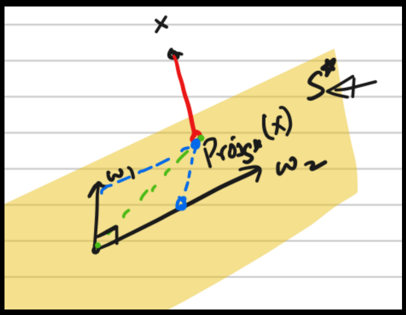
Maximize : $\langle x, v \rangle$!

BEST FIT LINE:

$$\boxed{\underset{v : \|v\| = 1}{\arg\max} \; \|x \cdot v\|^2}$$

If $S = \text{span}\{w_1, w_2, w_3\}$

$$\boxed{\|\text{Proj}_S(x)\|^2 = \|x \cdot w_1\|^2 + \|x \cdot w_2\|^2 + \|x \cdot w_3\|^2}$$



Finding 2 Right Singular vectors, one after other

maximizes $var(S; X)$.

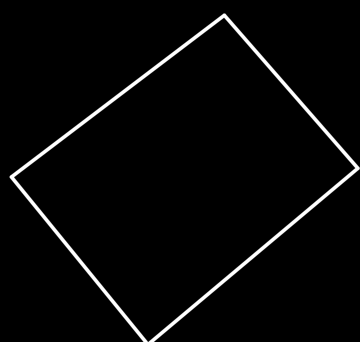
$dim(S = 2)$

PROOF:

We pick v_1: $\underset{v:\|v\|=1}{\arg\max} \|X \cdot v\|^2$

v_2: $\underset{\substack{v:\|v\|=1 \\ v \perp v_1}}{\arg\max} \|X \cdot v\|^2$

Assume the subspace that maximizes $var(S; X)$ is S^*.



S^*

this can be described as span of 2 orthonormal vectors $\rightarrow w_1, w_2$.

Pick w_2 such that $w_2 \perp v_1$.

So $\quad var(S;x) = \| Proj_S(x) \|_2^2$

$S_1 = \{v_1, v_2\} \rightarrow \quad \|x \cdot v_1\|^2 + \|x \cdot v_2\|^2$

$S^* \quad \rightarrow \quad \|x \cdot w_1\|^2 + \|x \cdot w_2\|^2$

$$v_1 \text{ is } \underset{v : \|v\| = 1}{\arg\max} \|x \cdot v\|^2$$

So $\quad \|x \cdot v_1\|^2 \geq \|x \cdot w_1\|^2.$

We know $\quad v_1 \perp w_2 \quad$ and

v_2 is

$$\underset{\substack{v : \|v\| = 1 \\ v \perp v_1}}{\arg\max} \|x \cdot v\|^2$$

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

So $\quad var(S_1 ; x) \geq var(S^* ; x)$

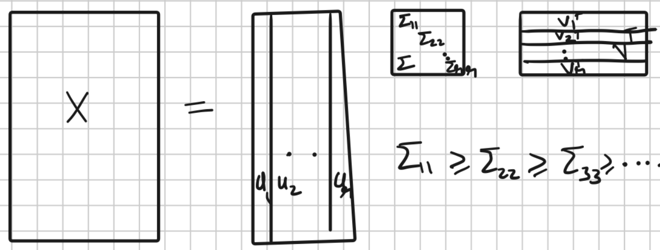
S_1 is S^* //

LOW-RANK APPROXIMATION PROBLEM

INPUT: Given $X \in \mathbb{R}^{n \times d}$, k.

OUTPUT: $\underset{\text{rank}(\tilde{X}) \leq k}{\text{argmin}} \|X - \tilde{X}\|_F$

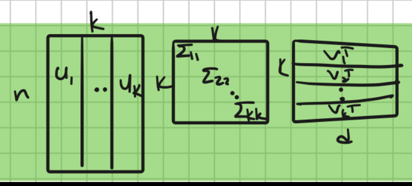
TWO MAGICAL PROPERTIES OF SVD



$\Sigma_{11} \geq \Sigma_{22} \geq \Sigma_{33} \geq \cdots \geq \Sigma_{nn}$

Thm: V_i is the i^{th} Right Singular vector of X.

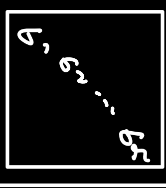
Thm: $X_k \equiv$



is a solution to best rank k approximation!

SVD:

$X = U \Sigma V^T$



$$\boxed{X = \sum_{i=1}^{r} \sigma_i u_i v_i^T}$$

$X v_i = \sigma_i u_i \qquad \|u_i\| = 1$

$$\boxed{\sigma_i = \|X \cdot v_i\|}$$

$$\boxed{u_i = \frac{X v_i}{\|X \cdot v_i\|}}$$

If we find $v_i \rightarrow$ gives σ_i, u_i.

Time to Compute SVD: $\boxed{O(n \cdot d^2)}$

$Y = X^T X \qquad X = U \Sigma V^T$

$Y^l = V \Sigma^{2l} V^T$

↳ σ_1 term dominates!

$$Y^l = \sum_{i=1}^{r} \sigma_i^{2l} u_i v_i^T$$

↳ small σ_i's $\rightarrow 0$ as $l \rightarrow \infty$

Power Iteration

- V_0 is a random unit vector.
- For $t = 1, \ldots, T$:
  - $u = X^T \cdot (X \cdot V_{t-1})$
  - $V_t = \dfrac{u}{\|u\|}$.

$$\boxed{\text{After } \quad T = O\left(\frac{\log(d/\varepsilon)}{\varepsilon}\right) \text{ iterations }, \underbrace{\|X \cdot v_T\|}_{\sigma_r} \geq (1-\varepsilon)\sigma_1}$$

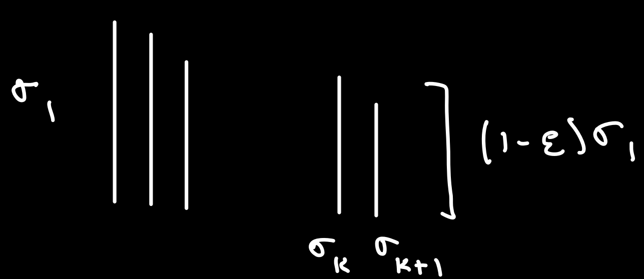
Let k be such that

$$\sigma_k > (1-\varepsilon)\sigma_1$$

$$\sigma_{k+1} \leq (1-\varepsilon)\sigma_1$$

$d \times n \cdot n \times d \cdot d \times 1$

$$\boxed{v \rightarrow d \times 1}$$



$S_k =$ span of $v_1, v_2 \ldots v_k$

After $T = O\left(\frac{\log(d/\delta)}{\varepsilon}\right)$ iterations, v_T is almost entirely within S_k

$$\|v_T - \text{Proj}_{S_k}(v_T)\| \leq \delta$$

ε: gap between σ_1, σ_k $\quad$ $X : n \times d$

After T iterations, $\sigma_{k+1} < (1-\varepsilon)\sigma_1$

$$\|v_T - \text{Proj}_{S_k}(v_T)\| \leq \frac{(1-\varepsilon)^{2T}}{|\langle v_0, v^{(1)}\rangle|}$$

$\hookrightarrow$ First RSV

If $\sigma_2 < (1-\varepsilon)\sigma_1$

To prove:

$$\|v_T - \text{Proj}_{S_1}(v_T)\| \leq \frac{(1-\varepsilon)^{2T}}{|\langle v_0, v^{(1)}\rangle|}$$

$$v_1 = \frac{X^T X v_0}{\|X^T X v_0\|}$$

$$\vdots$$

$$v_t = \frac{(X^T X)^t \cdot v_0}{\|(X^T X)^t \cdot v_0\|}$$

$$X = \sum_{i=1}^{d} \sigma_i u_i v_i^T$$

$$X^T X = V \Sigma U^T U \Sigma V^T = V \Sigma^2 V^T$$

$$= \sum_{i=1}^{d} \sigma_i^2 v_i v_i^T$$

$$(X^T X)^t = \sum_{i=1}^{d} \sigma_i^{2t} v_i v_i^T$$

Steps: i) Find v_t in terms of X, v_0

ii) Write $(X^T X)^t \cdot v_0$ in terms of $\sigma_i, v^{(i)}$

iii) Find $\| v_t - \text{Proj}_{\{v^{(1)}\}}(v_t) \|$

iv) Numerator $\leq$, denominator $\geq$

$$v_t = \frac{(X^T X)^t \cdot v_0}{\| (X^T X)^t \cdot v_0 \|}$$

(i) $(X^T X)^t \cdot v_0$

$$= \left(\sum_{i=1}^{d} \sigma_i^{2t} \underset{1 \times d}{v^{(i)}} \underset{d \times 1}{v^{(i)T}} \right) \cdot v_0$$

$$= \sum_{i=1}^{d} \sigma_i^{2t} < v_0, v^{(i)} > v^{(i)T}$$

$\rightarrow$ orthonormal

(ii) $$\|(X^T X)^t \cdot v_0\|^2 = \sum_{i=1}^{d} \sigma_i^{4t} \cdot < v_0, v^{(i)} >^2$$

(iii) $$\mathrm{Proj}_{\{v^{(1)}\}} ((X^T X)^t v_0) = \sigma_1^{2t} \cdot < v_0, v^{(1)} > v^{(1)}$$

(iv) $$\| v_t - \mathrm{Proj}_{\{v^{(1)}\}} (v_t) \|^2 = \frac{\left\| \sum_{i=2}^{d} \sigma_i^{2t} < v_0, v^{(i)} > v^{(i)T} \right\|^2}{\sum_{i=1}^{d} \sigma_i^{4t} \cdot < v_0, v^{(i)} >^2}$$

$$= \frac{\sum_{i=2}^{d} \sigma_i^{4t} < v_0, v^{(i)} >^2}{\sum_{i=1}^{d} \sigma_i^{4t} < v_0, v^{(i)} >^2}$$

$$\sigma_2 \leq (1-\epsilon)\sigma_1$$

Numerator $$\leq \sum_{i=2}^{d} \sigma_2^{4t} < v_0, v^{(i)} >^2$$

$$= (1-\epsilon)^{4t} \sigma_1^{4t} \left(\sum_{i=2}^{d} \underbrace{< v_0, v^{(i)} >^2}_{\text{both unis}} \right)$$

≤ 1

$$\leq (1-\varepsilon)^{4t} \sigma_1^{4t}$$

$$\text{Denominator} \geq \sigma_1^{4t} \langle v_0, v(i) \rangle^2$$

$$\| v_t - \text{Proj}_{\{v(1)\}}(v_t) \|^2 \leq \frac{\sigma_1^{4t} \cdot (1-\varepsilon)^{4t}}{\sigma_1^{4t} \langle v_0, v(i) \rangle^2}$$

$$\| v_t - \text{Proj}_{\{v(1)\}}(v_t) \| \leq \frac{(1-\varepsilon)^{2t}}{\langle v_0, v(i) \rangle} \quad //$$

General PI:

→ Pick k orthonormal vectors $y_0^{(1)}, y_0^{(2)}, \ldots, y_0^{(k)}$.

→ For $t = 1, \ldots, T$:

- Let $z^{(1)} = X^T \cdot X \cdot y_{t-1}^{(1)}$, $z^{(2)} = X^T \cdot X \cdot y_{t-1}^{(2)}$, $\ldots$, $z^{(k)} = X^T \cdot X \cdot y_{t-1}^{(k)}$.
- Let $y_t^{(1)}, y_t^{(2)}, \ldots, y_t^{(k)}$ be an orthonormal basis for $\text{Span}(\{z^{(1)}, z^{(2)}, \ldots, z^{(k)}\})$.

$$\boxed{\begin{array}{lll} \text{PI} \quad \text{Convergence Rate} & & 1/\varepsilon \\ \quad \text{with momentum} & & 1/\sqrt{\varepsilon} \end{array}}$$

Power Iteration + "Momentum":

→ v_0 : random unit vector

→ For $t = 1, \dots, T$:

$$u = X^T \cdot (X \cdot v_{t-1}) - \beta \cdot V_{t-2} \, .$$

$$V_t = \frac{u}{\|u\|} \, .$$

MATRIX COMPLETION:

$$L(\tilde{x}) = \sum_{i,j \in O} (x_{ij} - \tilde{x}_{ij})^2$$

↘ known entries

$$\boxed{\begin{array}{l} \min \; L(\tilde{x}) \\ \quad \text{given} \;\; \text{Rank}(\tilde{x}) \leq k \end{array}}$$

Singular Value Projection:

Apply GD to the optimization problem.

→ At each step, use SVD to make the new x matrix as rank k.

x_0 : Random

For $t = 1, \dots T$:

$$y_t = x_{t-1} - \eta \nabla L(x_{t-1})$$

$$x_t = (\text{Take top } k\text{-SVD of } y_t)$$

↳ $\text{Proj}_C(y_t)$

↳ set of rank $\leq k$ matrices.

Summary:

1. Best-Fit-Subspace Problem:

$$ERR(S; X) = \sum_i \| x^i - Proj_S(x^i) \|_2^2$$

$$\max \rightarrow var(S; X) = \sum_i \| Proj_S(x^i) \|_2^2$$

- k^{th} RSV

$$\left.\begin{array}{l} v_k = \arg\max \|X \cdot v\|^2 \\ v \perp v_1, v \perp v_2 \ldots v \perp v_{k-1}, \\ \|v\| = 1 \end{array}\right\} \text{Aligns to } X$$

- If $S = \{w_1, w_2, w_3\}$

$$var(S; X) = \|X \cdot w_1\|^2 + \|X \cdot w_2\|^2 + \|X \cdot w_3\|^2$$

2. Low-Rank Approximation:

$$\underset{rank(\tilde{X} \leq k)}{\arg\min} \|X - \tilde{X}\|_F$$

3. SVD:

$O(n \cdot d^2)$

$$X = U \Sigma V^T$$

$$= \sum_{i=1}^{u} \sigma_i u_i v_i^T$$

$$\sigma_i = \|X \cdot v_i\|$$

$$u_i = \frac{X \cdot v_i}{\|X \cdot v_i\|}$$

4. POWER ITERATION:

Find 1^{ST} RSV

$$Y = X^T X$$

$$Y^{\ell} = \sum_{i=1}^{d} \sigma_i^{2\ell} u_i v_i^T$$

→ v_0 : random

→ For $t = 1 \dots T$

* $u = X^T \cdot (X \cdot v_{t-1})$

* $v_t = \dfrac{u}{\|u\|}$

5. BOUNDS

* $T \propto 1/\varepsilon$: After $T = O\left(\dfrac{\log(d/\delta)}{\varepsilon}\right)$ iterations,

$$\|v_T - \mathrm{Proj}_{S_k}(v_T)\| \leq \delta$$

$X : n \times d$, $\sigma_{k+1} < (1-\varepsilon)\sigma_1$, $\sigma_k \geq (1-\varepsilon)\sigma_1$

$$S_k = \mathrm{span}(v^{(1)}, v^{(2)}, \dots, v^{(k)})$$

* After T iterations

$$\| v_T - \text{Proj}_{S_k}(v_T) \| \leq \frac{(1-\varepsilon)^{2T}}{|\langle v_0, v^{(1)} \rangle|}$$

If Distance between $\sigma_2 = \frac{\sigma_1}{2}$, $1-\varepsilon = 1/2$

$\sigma_2 = \sigma_1/3$, $1-\varepsilon = 1/3$

↳ v_T will converge to S_k at $(1/3)^{2T}$ speed!

→ PROOF FOR $\sigma_2 < (1-\varepsilon)\sigma_1$

Steps: i) Find v_t in terms of X, v_0

ii) Write $(X^T X)^t \cdot v_0$ in terms of $\sigma_i, v^{(i)}$

iii) Find $\| v_t - \text{Proj}_{\{v^{(1)}\}}(v_t) \|$

iv) Numerator $\leq$, denominator $\geq$

6.i. General PI:

→ v_0^i → k orthonormal random $[i = 1 \dots k]$

→ $t = 1, \dots T$:

* $z^i = X^T \cdot X \cdot v_{t-1}^i \quad \forall\, i$

* $v_t^1 \dots v_t^k \leftarrow$ orthonormal basis of $\text{span}(z^1, \dots z^k)$

ii. Momentum

$$u = X^T \cdot (X \cdot v_{t-1}) - \beta v_{t-2}$$

Convergence: $1/\sqrt{\varepsilon}$

7. MATRIX COMPLETION:

$$L(\tilde{x}) = \sum_{i,j \in O} (x_{ij} - \tilde{x}_{ij})^2$$

$\to$ known entries

$$\boxed{\begin{array}{l} \min \ L(\tilde{x}) \\ \quad \text{given} \ \ \text{Rank}(\tilde{x}) \leq k \end{array}}$$

x_0 : Random

For $t = 1, \ldots T$:

$$y_t = x_{t-1} - \eta \nabla L(x_{t-1})$$

$$x_t = (\text{Take top } k\text{-SVD of } y_t)$$

HOMEWORK

1. let SVDs be $v_1, v_2 \dots$

$$v = \sum_i \alpha_i v_i$$

$$X = \sum_i \sigma_i u_i v_i^T$$

$$Xv = \sum_i \sigma_i u_i v_i^T \cdot \sum_i \alpha_i v_i$$

$$= \sum_i \sigma_i u_i \alpha_i$$

$$||Xv||^2 = ||\sum_i (\sigma_i \alpha_i)^2||$$