* Contents from Hopcroft & Kannan's (not yet published) Book.

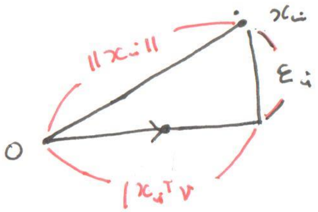
SVD - Low rank approximation

$$A = n\left\{\overbrace{\begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix}}^{d}\right.$$

x_i : a point in $\mathbb{R}^d$

What is the best k-dimensional subspace for these points $x_1 \cdots x_n$?

<I> Best 1-dimensional subspace span$\{v^*\}$



$$\varepsilon_i^2 = \|x_i\|^2 - |x_i^T v|^2$$

$$v^* = \arg\min_{v:\,\|v\|=1} \Sigma\, \varepsilon_i^2$$

$$= \arg\max_{v:\,\|v\|=1} \Sigma\, |x_i^T v|^2$$

$$= \arg\max_{v:\,\|v\|=1} \|Av\|^2$$

$$= \arg\max_{v:\,\|v\|=1} \frac{v^T A^T A v}{v^T v}$$

$= v_1$ eigenvector of A^TA corresponding to the largest eigenvalue λ_1

$$A^T A v_1 = \lambda_1 v_1$$

$$\|A v_1\| = \sqrt{\lambda_1} = \sigma_1$$

Facts)

$$v_2 = \arg\max_{\substack{v:\,\|v\|=1 \\ v \perp v_1}} \|Av\|^2 \quad \cdots ①$$

$$v_3 = \arg\max_{\substack{v:\,\|v\|=1 \\ v\perp v_1,\ v \perp v_2}} \|Av\|^2$$

and so on

Proof for ①)

$$\|Av\|^2 = v^T (A^T A) v$$

$$= v^T (Q \Lambda Q^T) v$$

$v \perp v_1$, $\|v\|=1$ mean that $Q^T v = \begin{bmatrix} 0 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}$ $d_2^2 + \cdots + d_n^2 = 1$

$$= \lambda_2 d_2^2 + \cdots + \lambda_n d_n^2$$

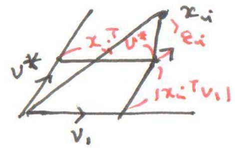
It is maximized when $d_2 = 0$, $d_3 = \cdots = d_n = 0$

or $v = v_2$

(II-1) Best 2-dimensional subspace span$\{v_1, v^*\}$

$\boxed{\text{i.e. best space containing } v_1}$ constraint.

$v^* \perp v$



$$\varepsilon_i^2 = x_i^T x_i - (x_i^T v_1)^2 - (x_i^T v^*)$$

$$v^* = \underset{\substack{\|v\|=1 \\ v \perp v_1}}{\arg\max} \ \Sigma (x_i^T v_1)^2 + (x_i^T v)^2$$

$$= \underset{\|v\|=1 \ \ v \perp v_1}{\arg\max} \ \Sigma (x_i^T v)^2$$

$$= \underset{\|v\|=1 \ \ v \perp v_1}{\arg\max} \ \|Av\|$$

$$= v_2$$

(II-2) Unconstrained best 2-D subspace $W = \text{span}\{w_1, w_2\}$

$\|w_1\| = \|w_2\| = 1, \quad w_1 \perp w_2$

Theorem. It is indeed $\text{span}\{v_1, v_2\}$

Proof) Suppose that $v_1 \notin W$

$\exists\, w_2 \in W$ such that $w_2 = \alpha_2 v_2 + \cdots + \alpha_d v_d$

$\because \underset{\text{dim } 2}{W} \cap \underset{\text{dim } d-1}{\text{span}\{v_2, \cdots, v_d\}} \neq 0$

Then $w_1 \triangleq$ Gram-Schmidt orthogonalization of W using w_2

Note that $\|Aw_1\|^2 \leq \|Av_1\|^2$

$$\|Aw_2\|^2 = w_2 A^T A w_2$$

$$= [0\ \alpha_2 \cdots \alpha_d] Q^T Q \Lambda Q^T Q \begin{bmatrix} 0 \\ \alpha_2 \\ \vdots \\ \alpha_d \end{bmatrix}$$

$$= \lambda_2 \alpha_2^2 + \cdots + \lambda_d \alpha_d^2$$

$$\leq \lambda_2 = \|Av_2\|^2$$

$$\therefore \underbrace{\|Aw_1\|^2 + \|Aw_2\|^2}_{\substack{\Sigma \text{ of projected lengths} \\ \text{onto span}\{w_1, w_2\}}} \leq \underbrace{\|Av_1\|^2 + \|Av_2\|^2}_{\substack{'' \\ \text{span}\{v_1, v_2\}}}$$

$\therefore$ It is contradiction.

W should contain v_1

and by (II-1) $W = \text{span}\{v_1, v_2\}$.

(III) Unconstrained best 3-D subspace $W = \text{span}\{w_1, w_2, w_3\}$

Theorem. $W = \{v_1, v_2, v_3\}$

$w_1 \perp w_2,\ w_2 \perp w_3,\ w_3 \perp w_1,$
$\|w_1\| = \|w_2\| = \|w_3\|$

Proof) Suppose that $v_1 \notin W$ and $v_2 \notin W$.

Then, $\exists\, w_3 \in W$ such that $w_3 = \alpha_3 v_3 + \cdots + \alpha_d v_d$

$w_1, w_2 =$ G.S.O of W starting from w_3

$$\|Aw_1\|^2 + \|Aw_2\|^2 \leq \|Av_1\|^2 + \|Av_2\|^2$$

$$\|Aw_3\|^2 \leq \|Av_3\|^2$$

$$\therefore \|Aw_1\|^2 + \|Aw_2\|^2 + \|Aw_3\|^2 \leq \|Av_1\|^2 + \|Av_2\|^2 + \|Av_3\|^2$$

Contradiction Q.E.D

(IV) By induction,

the best k-D subspace $\cdot$ $\text{span}\{v_1, \cdots, v_k\}$

Approximation errors.

① Original signal energy

$$\|A\|_F^2 = \sum_{i=1}^{n} x_i^T x_i$$

$$= \sum_{j=1}^{d} \sum_{i=1}^{n} a_{ij}^2$$

$\boxed{\|A\|_F = \text{Frobenius norm of } A}$

② Sum of squared projection errors onto span$\{v_i\}$

$$= \sum_{i=1}^{n} \left(x_i^T x_i - (x_i^T v_i)^2 \right)$$

$$= \|A\|_F^2 - \|A v_i\|^2$$

$$= \|A\|_F^2 - \sigma_i^2$$

③ Smallest sum of squared projection errors onto k-D subspace

$$= \|A\|_F^2 - \sum_{i=1}^{k} \sigma_i^2$$

④ $$\boxed{\|A\|_F^2 = \sum_{i=1}^{n} \sigma_i^2 = \sum_{i=1}^{r} \sigma_i^2}$$

Also note that

$$\|A\|_F = \operatorname{tr}(A^T A) = \sum \lambda_i = \sum \sigma_i^2$$

Another matrix norm: $\|A\|_2$

$$\|A\|_2 = \max_{\|v\|=1} \|Av\|$$

$$= \sqrt{\max_{\|v\|=1} v^T A^T A v}$$

$$= \sqrt{\lambda_1}$$

$$= \sigma_1$$

Best Rank-k Approximation of A

$$A = U \Sigma V^T$$
$$= \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

Partial sum

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

Theorem A_k is the best rank-k approximation of A both in terms of the Frobenius norm or the 2-norm

In other words

$$A_k = \arg\min_{\text{rank}(B)=k} \|A - B\|_F \quad \cdots \text{①}$$
$$= \arg\min_{\text{rank}(B)=k} \|A - B\|_2 \quad \cdots \text{②}$$

Proof of ①. Note that $\|A-B\|_F^2 = \sum_{i=1}^{n} \underbrace{\|x_i - \tilde{x}_i\|^2}_{\text{row approximation error.}}$

Suppose that B is optimal and $R(B^T) \triangleq \text{span}\{w_1, \cdots, w_k\}$

Then each row of B should be the projection of the row of A onto $\text{span}\{w_1 \cdots w_k\}$. otherwise by making so we can reduce $\|A - B\|_F$.

We know that $R(B^T) = \text{span}\{v_1 \cdots v_k\}$

and

$$B = \begin{bmatrix} \tilde{x}_1^T \\ \vdots \\ \tilde{x}_n^T \end{bmatrix}$$

$\tilde{x}_j^T$ should be the projection of x_j onto V_k

* Projection of x onto $\text{span}\{v_1, \cdots, v_k\}$

where $v_i^T v_j = \delta_{ij}$

$$x^P = (x^T v_1) v_1 + \cdots + (x^T v_k) v_k$$
$$= (v_1 v_1^T + \cdots + v_k v_k^T) x$$
$$= V_k V_k^T x$$

where $V_k = [v_1 \cdots v_k]$

Therefore

$$B = \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix} V_k V_k^T$$
$$= U \Sigma V^T V_k V_k^T$$
$$= \left[U_k \,\middle|\, U' \right] \left[\begin{array}{c|c} \Sigma_k & \\ \hline & \Sigma' \end{array} \right] \left[\begin{array}{c} V_k^T \\ \hline (V')^T \end{array} \right] V_k V_k^T$$
$$= U_k \Sigma_k V_k$$
$$= \sum_{i=1}^{k} \sigma_i u_i v_i^T$$

Q.E.D.

Proof of ②

Lemma $\|A - A_k\|_2 = \sigma_{k+1}$

Suppose that B is optimal and

$$\|A - B\|_2 < \sigma_{k+1}.$$

Let $z \in \underbrace{N(B)}_{\dim\ n-k} \cap \underbrace{\mathrm{span}\{v_1, \cdots, v_{k+1}\}}_{k+1} \neq 0$

and $\|z\| = 1$ $\quad z = \alpha_1 v_1 + \cdots + \alpha_k v_{k+1}$

$$\begin{array}{ccccc} \|(A-B)z\| & \leq & \|A-B\|_2 & < & \sigma_{k+1} \\ \| & & & & \\ \|Az\| & = & \sqrt{\lambda_1 \alpha_1^2 + \cdots + \lambda_{k+1}\alpha_{k+1}^2} & \geq & \sqrt{\lambda_{k+1}} = \sigma_{k+1} \end{array}$$

contradiction.

Q.E.D.