# SVD

Let $S \in \mathbb{R}^{m \times n}$ be non-square, thus we cannot diagonalize it. However, we can look at

$$S^T S \in \mathbb{R}^{n \times n} \longrightarrow \text{square, symmetric, real-valued, pos. def.}$$

$\Downarrow$

eigen-decomposition exists!

$\lambda_i \geq 0$ since $S^T S$ pos. def

$$S^T S = \underbrace{V}_{(v_1, \ldots, v_n)} \Lambda V^T \underset{V^T = V^{-1}}{=} \sum_{i=1}^{n} \lambda_i \underbrace{v_i}_{\text{column vector of } V} v_i^T = \sum_{i=1}^{n} \sigma_i^2 v_i v_i^T$$

$$\Longrightarrow \quad S^T S v_i = \sigma_i^2 v_i$$

Note that $\text{rank}(S^T S) = \text{rank}(S)$

Why? Assume we have $z_1, \ldots, z_r \neq 0$ such that $S z_i = 0$ and $z_i^T z_j = \delta_{ij}$. Then $S^T S z_i = 0$.

Lets sort the eigenvalues and define $r \leq n$ such that $\underbrace{\sigma_{i+1} = 0}_{\text{nullspace of } S}$ $\forall i > r$.

Thus, we have

$$\boxed{\underbrace{v_1, \dots, v_r}_{\substack{\text{orthonormal} \\ \text{vectors with} \\ \sigma_i > 0,\, i \in [1, r]}} \quad , \quad \underbrace{v_{r+1}, \dots, v_n}_{\substack{\text{basis of null space} \\ \sigma_i = 0,\, i \in [r+1, n]}}}$$

Now lets look at $SS^T \in \mathbb{R}^{m \times m}$.

Define $u_i = \frac{S v_i}{\sigma_i}$. This is a $\sigma_i^2$-eigenvector of $SS^T$:

$$SS^T u_i = \frac{1}{\sigma_i} S S^T S v_i = \sigma_i S v_i = \sigma_i^2 u_i$$

Moreover, we have:

$$u_i^T u_j = \frac{1}{\sigma_i \sigma_j} v_i^T S^T S v_j = \frac{\sigma_j^2}{\sigma_i \sigma_j} \underbrace{v_i^T v_j}_{= \delta_{ij}} = \delta_{ij}$$

__Denote by:__

$U_r \in \mathbb{R}^{m\times r}$ the matrix where the $i^{th}$ column is $u_i = \frac{S v_i}{\sigma_i}$ $\quad(\sigma_i > 0\,!)$

$\Sigma_r \in \mathbb{R}^{r\times r}$ the diag. matrix $\operatorname{diag}(\sigma_1,\dots,\sigma_r)$

$V_r \in \mathbb{R}^{r\times n}$

Then we can write:
$$\underset{\mathbb{R}^{m\times r}}{\underset{\Uparrow}{U_r}} = \underset{\mathbb{R}^{m\times n}}{\underset{\Uparrow}{S}}\,\underset{\mathbb{R}^{n\times r}}{\underset{\Uparrow}{V_r}} \cdot \underset{\mathbb{R}^{r\times r}}{\underset{\Uparrow}{\Sigma_r^{-1}}} \checkmark$$

$$\Rightarrow \quad U_r \Sigma_r = S V_r$$

Lets write this out, complete $u_1,\dots,u_r$ to a full orthonormal basis in $\mathbb{R}^m$ by adding orth. vectors $u_{r+1},\dots,u_m$, (e.g. using Gram-Schmidt process) and add a few zeroes:

__Lefthand side__

$$\underbrace{(u_1,\dots,u_r,u_{r+1},\dots,u_m)}_{=U,\ \text{with } U^T=U^{-1}}
\begin{array}{c}\\ \\ \\ m-r\{\end{array}
\begin{pmatrix}
\begin{matrix}\sigma_1 & & 0\\ & \ddots & \\ 0 & & \sigma_r\end{matrix} & \overbrace{\quad 0 \quad}^{n-r} \\
0 &
\end{pmatrix}
=
\begin{pmatrix} U_r\Sigma_r + 0 \\ 0 \end{pmatrix}
\begin{matrix}\}\,r \\ \}\,m-r\end{matrix}$$

Right-hand side: Here we use $V$ instead of $V_r$

$$\rightarrow \quad S \cdot V = \begin{pmatrix} S \cdot V_r \\ \underbrace{0}_{\text{because } S \cdot v_i = 0 \text{ for } i > r} \end{pmatrix} \begin{matrix} \}\, r \\ \}\, n-r \end{matrix}$$

$\Rightarrow$ plug both sides together to get

$$U\Sigma = SV$$

and using $V^{-1} = V^T$, we get

$$\boxed{S = U \Sigma V^T}$$

with $V \in \mathbb{R}^{n \times n}$, $U \in \mathbb{R}^{m \times m}$, $\Sigma \in \mathbb{R}^{m \times n}$

and $V^T = V^{-1}$, $U^T = U^{-1}$, $\Sigma_{ii} = \sigma_i$

for $i < r$ and $\Sigma_{ij} = 0$ otherwise.

## Important property of SVD

Assume we have a matrix $A$ with SVD $A = U \Sigma V^T$. Let $A_s$ be the s-rank approximation of $A$, $A_s = \sum_{k=1}^{s} \sigma_k u_k v_k^T$

s largest singular components

**Then** among all matrices of rank $s$, $A_s$ best approximates $A$ w.r.t. the operator and Frobenius norm!

### Frobenius norm

This one is defined as $\|A\|_F = \sqrt{\sum_{ij} A_{ij}^2}$

$\Rightarrow$ equivalent to Euclidean norm for vectors!

We can write this using SVD:

$$\rightarrow \mathrm{Tr}(A^T A) = \sum_i \sum_k \underbrace{A_{ki}}_{= A^T_{ik}} A_{ki} = \sum_{ik} A_{ki}^2 = \|A\|_F^2$$

$$\rightarrow \|A\|_F^2 = \mathrm{Tr}(A^TA) \underset{A=U\Sigma V^T}{=} \mathrm{Tr}(V\Sigma \underbrace{U^TU}_{=\mathbb{1}}\Sigma V^T)$$

$$\underset{\substack{\mathrm{Tr}(ABC)\\=\mathrm{Tr}(CAB)}}{=} \mathrm{Tr}(\underbrace{V^TV}_{=\mathbb{1}}\Sigma^2)$$

$$= \mathrm{Tr}(\Sigma^2) = \sum_{k=1}^{r} \sigma_k^2$$

where $r = \mathrm{rank}(A)$

<u>Operator norm</u>

$$\|A\|_2 = \sup_{\|x\|=1} \|Ax\|$$ by how much does A dilate vectors?

This can also be represented using SVD.

$$\|Ax\|^2 = \|x^TA^TAx\| = \|x^TV\Sigma \overbrace{U^TU}^{=\mathbb{1}}\Sigma Vx\|$$

$$= \|\Sigma Vx\|^2 = \sum_{ijk}\Big(\sum_{ij} V_{jk}x_k\Big)^2 = \sum_{i=1}^{r}\sigma_i^2\Big(\sum_k V_{ik}x_k\Big)^2$$

$$\leq \sigma_1^2 \sum_{i=1}^{r} \Big(\sum_k V_{ik} x_k\Big)^2 \leq \sum_{i=1}^{n} \Big(\sum_k V_{ik} x_k\Big)^2 \cdot \sigma_1^2$$

largest singular component

$$= \sigma_1^2 \|Vx\|^2$$

$$= \sigma_1^2 \| x^T \underbrace{V^T V}_{= \mathbb{1}} x \|$$

$$= \sigma_1^2 \|x\|^2$$

Choosing $x = v_1$, we get $\|Av_1\|^2 = \sigma_1^2$

i.e., the identity

$$\Rightarrow \|A\|_2 = \sup_{\|x\|} \|Ax\| = \sigma_1$$

## Error of our s-rank approximation

Reminder: we had $A = U \Sigma V^T$

and $A_s = \sum_{k=1}^{s} \sigma_k u_k v_k^T$ = truncated SVD

$$\|A_s - A\|_F^2 = \sqrt{\sum_{k=s+1}^{r} \sigma_k^2}$$ ⇝ remaining singular values!

and $\|A_s - A\|_2 = \sigma_{s+1}$ → largest remaining singular value

## Non-negative matrix factorization

Before we derive the update rule, we have to introduce the concept of auxiliary functions.

$G(h,h')$ is an auxiliary fct. of $F(h)$ if
$$G(h,h') \geq F(h) \text{ and } G(h,h) = F(h)$$

These have useful properties! We will use the following:

If $G$ is an aux. fct. of $F$, then $F$ is non-increasing under the update rule $h^{t+1} = \underset{h}{\operatorname{argmin}}\, G(h, h^t)$.

**Proof:** $F(h^{t+1}) \underset{\uparrow \text{ Def.}}{\le} G(h^{t+1}, h^t) \underset{\uparrow \text{ upd. rule}}{\le} G(h^t, h^t) \underset{\uparrow \text{ Def.}}{=} F(h^t)$

From this follows: $F(h_{min}) \underbrace{\le \dots \le}_{\text{iteratively apply the update rule}} F(h_0)$

## Loss function

Let's introduce the following loss fct:

$$F(h) = \frac{1}{2}\Big(\sum_i \underbrace{v_i - \sum_a W_{ia} h_a\Big)^2}_{v = Wh \;\to\; \text{reconstr. for one column of } V}$$

Our main statement to prove is the following:

If $K_{ab}$ is the diag. matrix $K_{ab}(h^t) = \delta_{ab} \dfrac{(W^T W h^t)_a}{h^t_a}$,

then

$$G(h, h^t) = F(h^t) + (h - h^t)^T \nabla F(h^t) + \frac{1}{2}(h - h^t)^T K(h^t)(h - h^t) \qquad (A)$$

is an auxiliary function of $F(h)$.

**Proof:** Note first that $G(h,h) = F(h)$ by construction. So we have to show that

$G(h, h^t) \geq F(h)$.

First, we Taylor expand $F(h)$:

$$(B) \quad F(h) = F(h^t) + (h-h^t)^T \nabla F(h^t) + \frac{1}{2}(h-h^t)^T W^T W (h-h^t)$$

because
$$\frac{\partial F(h)}{\partial h_b} = -\left(\sum_i v_i - \sum_a W_{ia} h_a\right) W_{ib}$$

$$\frac{\partial F(h)}{\partial h_b \partial h_c} = \sum_i W_{ib} W_{ic} = \sum_i W_{bi}^T W_{ic} = (W^T W)_{bc}$$

and
$$\frac{\partial F(h)}{\partial h_b \partial h_c \partial h_e} = 0$$

Then we take the difference of $(A) - (B)$, which we want to show is $\geq 0$:

$$(h-h^t)^T (K(h^t) - W^T W)(h-h^t) \geq 0$$

So we have to show that $K(h^t) - W^TW$ is positive definite, i.e., $f^T(K(h^t) - W^TW)f \geq 0 \;\forall f$.

We will show this for the matrix $M = (h^t)^T[K - W^TW]h^t$ instead. If $M$ is pos. def., then $v^TMv \geq 0$

$$\Leftrightarrow (h^tv)^T[K - W^TW]h^tv \geq 0$$

$$\Leftrightarrow f^T[K - W^TW]f \geq 0$$

for $f = h^tv$

$$\rightarrow v^TMv = \sum_{a,b} v_a M_{ab} v_b$$

$$= \underbrace{\sum_{a,b} v_a h^t_a K(h^t)_{ab} h^t_b v_b}_{= N_1}$$

$$- \underbrace{\sum_{a,b} v_a h^t_a (W^TW)_{ab} h^t_b v_b}_{= N_2}$$

$$N_1 = \sum_{a,b} v_a \cancel{h^t_a} \delta_{ab} \frac{(W^TWh^t)_a}{\cancel{h^t_a}} h^t_b v_b$$

$$= \sum_a v_a^2 h^t_a \sum_b (W^TW)_{ab} h^t_b = \sum_{a,b} v_a^2 h^t_a h^t_b (W^TW)_{ab}$$

Since $(W^TW)_{ab} = (W^TW)_{ba}$, we can write

$$N_1 = \sum_{a,b} h_a^t h_b^t (W^TW)_{ab} \left[\frac{1}{2} v_a^2 + \frac{1}{2} v_b^2\right]$$

(it is the same sum if we swap $a \leftrightarrow b$)
$\Rightarrow$ factor $1/2$

Further, we have

$$N_2 = \sum_{a,b} h_a^t h_b^t (W^TW)_{ab} \left[v_a v_b\right]$$

$$\begin{aligned} \Rightarrow \quad v^T M v &= N_1 - N_2 \\ &= \sum_{a,b} h_a^t h_b^t (W^TW)_{ab} \left[\frac{1}{2} v_a^2 + \frac{1}{2} v_b^2 - v_a v_b\right] \\ &= \frac{1}{2} \sum_{a,b} \underbrace{h_a^t h_b^t}_{\geq 0} \underbrace{(W^TW)_{ab}}_{\geq 0} \underbrace{(v_a - v_b)^2}_{\geq 0} \\ &\geq 0 \end{aligned}$$

because we do non-neg. factorization!

$\square$

(note: we have to make sure $h^t$ stays positive $\forall t$. We will see that the update rule satisfies this!)

## Update rule

**Remember:** If $G(h,h')$ is an aux. fct. of $F(h)$, then $h^{t+1} = \underset{h}{\mathrm{argmin}}\, G(h,h^t)$ minimises $F$!

$\longrightarrow$ We apply this to our $G$ and $F$!

To get the argmin, calculate the minimum of $G(h,h^t)$:

$$\nabla_h G(h,h^t) = \nabla F(h^t) + K(h^t)(h-h^t) \overset{!}{=} 0$$

$$\Rightarrow h^{t+1} = h^t - K^{-1}(h^t)\nabla F(h^t)$$

Plug in: $K^{-1}_{ab}(h^t) = \delta_{ab} \cdot \dfrac{h^t_a}{(W^TWh^t)_a}$

and $\nabla_b F(h^t) = -\sum_i \left(v_i - \sum_c W_{ic} h^t_c\right) W_{ib}$

$$\Rightarrow h^{t+1}_a = h^t_a - \sum_b K^{-1}_{ab} \nabla_b F(h^t)$$

$$= h^t_a + \sum_{b,i} \frac{\left(v_i - \sum_c W_{ic} h^t_c\right) W_{ib}\, h^t_a\, \delta_{ab}}{(W^TWh^t)_a}$$

$$= h_a^t + \sum_{b,i} \frac{(v_i - \sum_c W_{ic} h_c^t) W_{ib} h_a^t \delta_{ab}}{(W^T W h^t)_a}$$

$$= h_a^t + \sum_i \frac{(v_i - \sum_c W_{ic} h_c^t) W_{ia} h_a^t}{(W^T W h^t)_a}$$

$$= \frac{\cancel{h_a^t (W^T W h^t)_a} + \overbrace{\sum_i v_i W_{ia}}^{(W^T v)_a} h_a^t - h_a^t \cancel{\overbrace{\left(\sum_{i,c} W^T_{ai} W_{ic} h_c^t\right)}^{= (W^T W h^t)_a}}}{(W^T W h^t)_a}$$

$$= h_a^t \frac{(W^T v)_a}{(W^T W h^t)_a}$$

$\Rightarrow$ for all matrix components

$$H_{ij}^{t+1} = H_{ij}^t \frac{(W^T V)_{ij}}{(W^T W H^t)_{ij}}$$

For the update rule of $W$, just do the same arguments but swap $W \leftrightarrow H$! $\square$