ALS1 – Přednáška 6

Rozklad na singulární hodnoty (SVD; Singular values decomposition)

1 Rozklad

Tato část rozebírá vlastnosti SVD; především to, že SVD rozklad existuje pro každou matici (věta 1).
Budeme potřebovat následující tvrzení:

Proposition 1. *Pro matici $Q_1 \in \mathbb{R}^{m\times k}$ s ortonormálními sloupci existuje matice $Q_2 \in R^{m\times(m-k)}$ tak že $Q = (Q_1\ Q_2)$ je ortogonální matice.*

Tvrzení 1 říká, že ortonormální báze podprostoru prostoru $\mathbb{R}^m$ může být rozšířena na ortonormální bázi celého prostoru. Toto tvrzení je známý výsledek z lineární algebry.

Doplnění 1 *2-norma matice:*

$$\|A\|_2 = \max_x \frac{\|Ax\|_2}{\|x\|_2}$$

$$\|A\|_2 = \sqrt{\max_{1\le i\le n} \lambda_i(A^TA)} =: \sigma_{\max}(A)$$

Theorem 1. *Jakákoli $m \times n$ matice A s $m \ge n$ může být rozložena na*

$$A = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^T, \tag{1}$$

kde $U \in \mathbb{R}^{m\times m}$ a $V \in \mathbb{R}^{n\times n}$ jsou ortogonální a $\Sigma \in \mathbb{R}^{n\times n}$ je diagonální

$$\Sigma = \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n),$$

$$\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_n \ge 0.$$

Důkaz. Předpoklad $m \ge n$ není nijak omezující: v opačném případě lze větu aplikovat na A^T.
Uvažujme maximalizační problém

$$\sup_{\|x\|_2=1} \|Ax\|_2$$

Protože hledáme supremum spojité funkce nad uzavřenou množinou, je tohoto suprema dosaženo pro nějaký vektor x. Položme $Ax = \sigma_1 y$, kde $\|y\|_2 = 1$ a $\sigma_1 = \|A\|_2$ (podle definice).
Podle Tvrzení 1 můžeme zkonstruovat ortogonální matice

$$Z_1 = (y\overline{Z}_2) \in \mathbb{R}^{m\times m}, \qquad W_1 = (x\overline{W}_2) \in \mathbb{R}^{m\times m}.$$

Pak

$$Z_1^T A W_1 = \begin{pmatrix} \sigma_1 & y^T A\overline{W}_2 \\ 0 & \overline{Z}_2^T A\overline{W}_2 \end{pmatrix},$$

protože $y^TAx = \sigma_1$ a $Z_2^TAx = \sigma_1\overline{Z}_2^T y = 0$.
Položme

$$A_1 = Z_1^T A W_1 = \begin{pmatrix} \sigma_1 & w^T \\ 0 & B \end{pmatrix}.$$

Pak

$$\frac{1}{\sigma_1^2 + w^Tw}\|A_1(\sigma_1 w)\|_2^{\,2} = \frac{1}{\sigma_1^2 + w^Tw}\left\|\begin{pmatrix} \sigma_1^2 + w^Tw \\ Bw \end{pmatrix}\right\|_2^2 \ge \sigma_1^2 + w^Tw.$$

Ale $\|A_1\|_2^{\,2} = \left\|Z_1^TAW_1\right\|_2^{\,2} = \sigma_1^2$. Takže musí platit $w = 0$. A tedy jsme provedli jeden krok v diagonalizaci A. Důkaz je pak dokončen indukcí:

Předpokládejme, že

$$B = Z_2 \begin{pmatrix} \Sigma_2 \\ 0 \end{pmatrix} W_2 \qquad \Sigma_2 = \mathrm{diag}(\sigma_2, \dots, \sigma_n).$$

Pak máme

$$A = Z_1 \begin{pmatrix} \sigma_1 & 0 \\ 0 & B \end{pmatrix} W_1^T = Z_1 \begin{pmatrix} 1 & 0 \\ 0 & Z_2 \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & W_2^T \end{pmatrix} W_1^T.$$

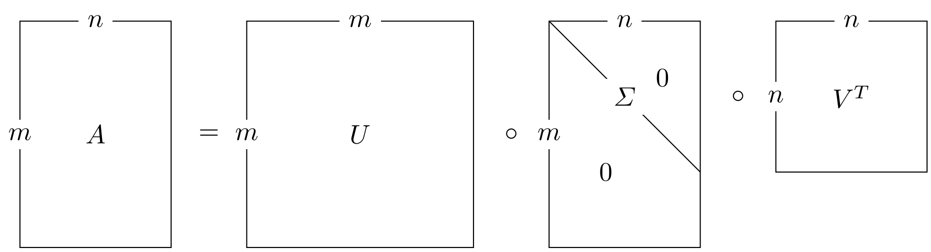
Definujeme U, Σ, V jako

$$U = Z_1 \begin{pmatrix} 1 & 0 \\ 0 & Z_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix}, \quad V = W_1 \begin{pmatrix} 1 & 0 \\ 0 & W_2^T \end{pmatrix}$$

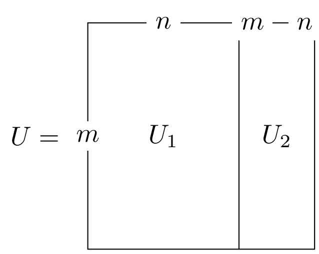
Věta je tímto dokázána.

Sloupce U a V se nazývají **singulární vektory**, diagonální prvky σ_i matice Σ se nazývají **singulární hodnoty**.

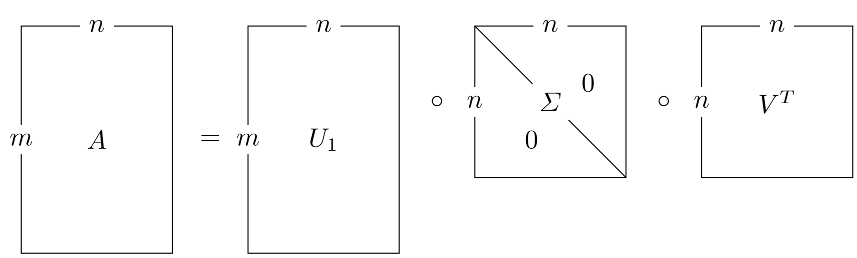
1.1 SVD symbolicky



Matici U můžeme vertikálně rozseknout na $m \times n$ matici U_1 a $m \times (m-n)$ matici U_2.



Při násobení $U\Sigma$ se hodnoty z U_2 násobí pouze nulami. Můžeme tedy psát:



2 Fundamentální podprostory dané SVD rozkladem

SVD dává ortogonální báze pro čtyři fundamentáln podprostory matice.

Obor hodnot matice A je lineární podprostor

$$\mathcal{R}(A) = \{ y \mid y = Ax \text{ pro lib. } x \}$$

Uvažujme, že A má hodnost r:

$$\sigma_1 \geq \dots \geq \sigma_r > \sigma_{r+1} = \dots = \sigma_n = 0.$$

Pak máme

$$y = Ax = \sum_{i=1}^{r} \sigma_i u_i v_i^T x = \sum_{i=1}^{r} (\sigma_i v_i^T x) u_i == \sum_{i=1}^{r} \alpha_i u_i.$$

Nulový prostor matice A je lineární podprostor

$$\mathcal{N}(A) = \{x \mid Ax = 0\}.$$

Protože $Ax = \sum_{i=1}^{r} \sigma_i u_i v_i^T x$, dostáváme, že pro libovolný vektor $z = \sum_{i=r+1}^{n} \beta_i v_i$ je v tom nulovém prostoru

$$Az = \left(\sum_{i=1}^{r} \sigma_i u_i v_i^T\right)\left(\sum_{i=r+1}^{n} \beta_i v_i\right) = 0.$$

Totéž můžeme ukázat pro A^T.

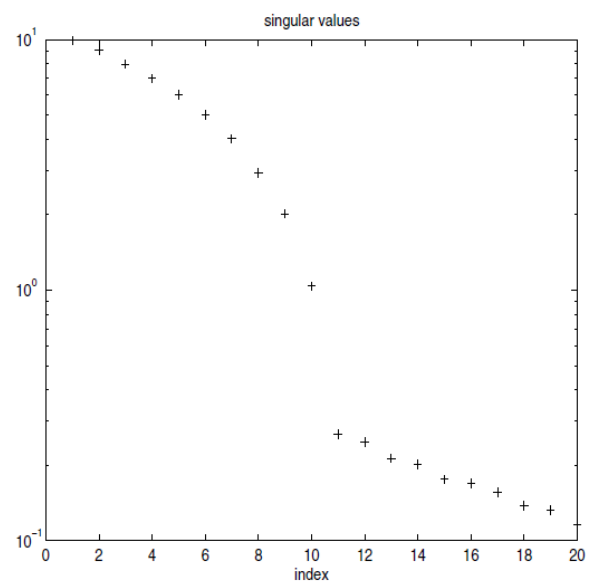
Theorem 2.

1. *Singulární vektory* $u_1, u_2, \ldots, u_r$ *jsou ortogonální báze v* $\mathcal{R}(A)$ *a* $h(A) = \dim(\mathcal{R}(A)) = r$.
2. *Singulární vektory* $v_{r+1}, v_{r+2}, \ldots, v_n$ *jsou ortogonální báze v* $\mathcal{N}(A)$ *a* $\dim(\mathcal{N}(A)) = n - r$.
3. *Singulární vektory* $v_1, v_2, \ldots, v_r$ *jsou ortogonální báze v* $\mathcal{R}(A^T)$.
4. *Singulární vektory* $u_{r+1}, u_{r+2}, \ldots, u_m$ *jsou ortogonální báze v* $\mathcal{N}(A^T)$.

3 Aplikace SVD

3.1 Aproximace matic

Uvažujme, že matice A matice s nízkou hodnosti a přidaným šumem: $A = A_0 + N$, kde šum je malý v porovnání s A_0. Singulární hodnoty matice A se pak typicky chovají jako ty na obrázku 1.



Obrázek 1. Singulární hodnoty matice hodnosti 10 s přidaným šumem.

Pokud je šum dostatečně malý, můžeme uvažovat pouze velké vlastní hodnoty, rekonstruovat matici A pouze pomocí nich (a odpovídajících singulárních vektorů) a aproximovat tak matici A maticí s nižší hodnosti. Předpokládejme, že počet velkých vlastních hodnot je k. Pak aproximujeme

$$A = \Sigma_{i=1}^{n} \sigma_i u_i v_i^T = A_k.$$

Tomuto se říká ořezaná SVD (*truncated SVD*). Ořezaná SVD je řešením aproximačního problému, kde se chce aproximovat danou matici maticí nižší hodnosti.

Theorem 3. *Předpokládejme, že matice* $A \in \mathbb{R}^{m \times n}$ *má hodnost* $r > k$. *Problém aproximace matice*

$$\min_{h(Z)=k} \|A - Z\|_2$$

má řešení

$$Z = A_k = U_k \Sigma_k V_k^T,$$

kde $U_k = (u_1, \ldots, u_k)$, $V_k = (v_1, \ldots, v_k)$, *and* $\Sigma_k = \operatorname{diag}(\sigma_1, \ldots, \sigma_k)$. *To minimum je*

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

Dá se ukázat, že totéž platí i pro *Frobeniovu normu*

$$\|A\|_F = \sqrt{\sum_{i,j} a_{ij}^2}.$$

Theorem 4. *Předpokládejme, že matice* $A \in \mathbb{R}^{m \times n}$ *má hodnost* $r > k$. *Problém aproximace matice*

$$\min_{h(Z)=k} \|A - Z\|_F$$

má řešení

$$Z = A_k = U_k \Sigma_k V_k^T,$$

kde $U_k = (u_1, \ldots, u_k)$, $V_k = (v_1, \ldots, v_k)$, *and* $\Sigma_k = \operatorname{diag}(\sigma_1, \ldots, \sigma_k)$. *To minimum je*

$$\|A - A_k\|_F = \left(\sum_{i=k+1}^{p} \sigma_i^2 \right)^{\frac{1}{2}},$$

kde $p = \min(m, n)$.

3.2 Analýza hlavních komponent (principal component analysis (PCA))

Aproximační vlastnosti SVD mohou být použity k vysvětlení ekvivalence SVD a PCA. Uvažujme že $X \in \mathbb{R}^{m \times n}$ je datová matice, kde každý sloupec je pozorování reálně-hodnotového náhodného vektoru s nulovým průměrem. Nechť SVD matice X je $X = U \Sigma V^T$. Pravé singulární vektory v_i se nazývají směry hlavních komponent v X. Vektor

$$z_1 = X v_1 = \sigma_1 u_1$$

má největší varianci ve vzorku mezi všemi normalizovanými lineárními kombinacemi sloupců z X:

$$\operatorname{Var}(z_1) = \operatorname{Var}(X v_1) = \frac{\sigma_1^2}{m}.$$

Nalezení vektoru s maximální variancí je ekvivalentní maximalizaci tzv. Rayleighova kvocientu:

$$\sigma_1^2 = \max_{v \neq 0} \frac{v^T X^T X v}{v^T v}, \qquad \operatorname{argmax}_{v \neq 0} \frac{v^T X^T X v}{v^T v}.$$

Normalizovaná proměnná $u_1 = (1/\sigma_1) X v_1$ se nazývá normalizovaná hlavní komponenta v X.

Pokud máme nalezený vektor největší variance ve vzorku, obvykle chceme pokračovat a najít vektor druhé největší variance ve vzorku, který je kolmý k tomu prvnímu. Toto je provedeno výpočtem vektoru redukované matice $X - \sigma_1 u_1 v_1^T$. Pokračováním tohoto procesu můžeme najít všechny hlavní komponenty, tedy vypočítáme singulární vektory.

3.3 Řešení problému nejmenších čtverců

Problém nejmenších čtverců může být vyřešen použitím SVD. Uvažujme, že máme systém $Ax \approx b$, kde matice $A \in \mathbb{R}^{m \times n}$ má hodnost $h(A) = n$ a SVD je

$$A = (U_1 U_2) \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^T,$$

kde $U_1 \in \mathbb{R}^{m \times n}$. Použitím SVD a faktu, že norma je invariantní k ortogonálním transformacím, dostáváme

$$\|r\|^2 = \|b - Ax\|^2 = \|b - U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^T x\|^2 - \| \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} - \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} y\|^2,$$

kde $b_i = U_i^T b$ a $y = V^T x$. Tedy

$$\|r\|^2 = \|b_1 - \Sigma y\|^2 + \|b_2\|^2.$$

Nyní můžeme snadno minimalizovat $\|r\|^2$ položením $y = \Sigma^{-1} b_1$. Řešení nejmenších čtverců je dánp

$$x = Vy = V\Sigma^{-1} b_1 = V\Sigma^{-1} U_1^T b.$$

Protože Σ je diagonální a diagonální hodnoty jsou nenulové (jelikož A má hodnost n), můžeme jednoduše najít Σ^{-1}:

$$\Sigma^{-1} = \mathrm{diag}\left(\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \ldots \frac{1}{\sigma_n}\right).$$

Takže minimální řešení může být zapsáno jako

$$x = \Sigma_{i=1}^{n} \frac{u_i^T b}{\sigma_i} v_i.$$

Theorem 5. *Mějme matici $A \in \mathbb{R}^{m \times n}$ s hodností n, a úzkou SVD $A = U_1 \Sigma V^T$. Pak problém nejmenších čtverců $\min_x \|Ax - b\|_2$ má řešení*

$$x = V\Sigma^{-1} U_1^T b = \Sigma_{i=1}^{n} \frac{u_i^T b}{\sigma_i} v_i.$$