S. Riedel
L.V. Brinker

Mathematische Grundlagen von Data Science

FernUniversität in Hagen
WS 23/24

EA 1

1. Orthogonale Diagonalisierbarkeit (10 Punkte). Es sei

$$A = \begin{pmatrix} 2 & 0 & 2 \\ 0 & 6 & 0 \\ 2 & 0 & 4 \end{pmatrix}.$$

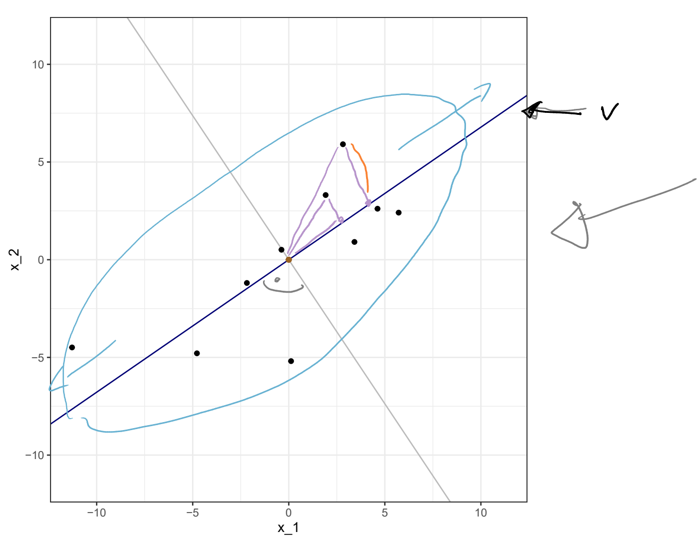
a) (2 Punkte) Bestimmen Sie, ob die Matrix orthogonal diagonalisierbar ist.

b) (2 Punkte) Berechnen Sie die Eigenwerte von A.

c) (2 Punkte) Berechnen Sie eine Matrix M und eine Diagonalmatrix D, sodass $D = M^{-1}AM$ gilt.

d) (4 Punkte) Berechnen Sie eine orthogonale Matrix S, sodass $A = S^{-1}DS$ gilt.

2. Singulärwerte I (10 Punkte). Wir nehmen an, wir haben einen Datensatz, der aus 10 zweidimensionalen Punkten besteht. Dieser ist in der Abbildung dargestellt. Wir wollen nun die Singulärwertzerlegung nutzen, um die so genannten Hauptkomponenten des Datensatzes herauszufinden.



x_1	-2.18	1.92	5.72	-11.28	4.62	-0.38	2.82	-4.78	3.42	0.12
x_2	-1.19	3.31	2.41	-4.49	2.61	0.51	5.91	-4.79	0.91	-5.19

./.

Sie dürfen ohne Beweis verwenden, dass der Mittelpunkt (Durchschnitt) der Punktewolke im Punkt $(0,0)$ liegt.

a) (3 Punkte) Wir suchen eine Koordinatenachse (bzw. Regressionsgerade) die durch den Ursprung verläuft, sodass beim Projizieren der Punkte $x^i = (x_1^i, x_2^i)$ auf die Gerade ein möglichst kleiner quadratischer Fehler entsteht. Die möglichen Achsen können wir als Vektoren $z = (z_1, z_2) \in \mathbb{R}^2$ mit der Länge 1 verstehen. Begründen Sie, warum

$$\boxed{z^* = \arg\max_{\|z\|=1} \sum_{i=1}^{10} \langle x^i, z \rangle^2}$$

die gesuchte Koordinatenachse ist und erklären Sie geometrisch, was hier maximiert wird.
Hinweis: Die orthogonale Projektion von x^i auf einen Vektor u ist der Punkt $\tilde{x}^i = \langle x^i, u \rangle \|u\|^{-2} u$.

b) (2 Punkte) Es sei X die Matrix, die die Datenpunkte x^i als Zeilen enthält. Zeigen Sie:

$$z^* = \arg\max_{\|z\|=1} \|Xz\|^2 \qquad \text{und} \qquad \boxed{\max_{\|z\|=1} \|Xz\|^2 = \sigma^2}.$$

σ ist hier der größte Singulärwert von X.

c) (2 Punkte) Es sei $X = U\Sigma V^T$ die Singulärwertzerlegung von X. Zeigen Sie, dass die erste Spalte von V das obige Optimierungsproblem löst. $\|Xv\|^2 = \sigma^2$

d) (3 Punkte) Berechnen Sie (numerisch) die Hauptachse und die Singulärwerte für das obige Beispiel. Erklären Sie, wo die Hauptachse in der Abbildung zu finden ist.

Die Singulärwertzerlegung kann (auch für $d > 2$) als Hauptachsentransformation verstanden werden: wir suchen Projektionen in q-dimensionale Unterräume, sodass die Abstände der Punkte zu ihren Projektionen möglichst klein sind. Dafür werden auch in höheren Dimensionen Koordinatenachsen berechnet. Die erste Koordinatenachse kann wie oben berechnet werden. Die zweite Koordinatenachse wird dann so gewählt, dass sie senkrecht zu der ersten steht und wieder die quadratischen Abstände minimiert und so weiter. Die Interpretation ist, dass bei der Projektion möglichst wenig Informationen verloren gehen sollen. Das Verfahren wird auch *Principal Component Analysis* genannt.

3. Singulärwerte II (10 Punkte). Es sei $A \in \mathbb{R}^{m\times n}$ und σ_r sei der kleinste Singulärwert von A. Zeigen Sie

$$\sigma_r = \min_{\substack{x\in \mathrm{Im}(A^T) \\ \|x\|=1}} \|Ax\|$$

Hinweis. Ist $A = U\Sigma V^T$ eine reduzierte Singulärwertzerlegung, dann gilt $\mathrm{Im}(A^T) = \mathrm{Span}\{v_1, \dots v_r\}$. Dabei sind $v_1, \dots v_r$ die orthogonalen Spalten von V.

4. Rechnen mit der Pseudo-Inversen (10 Punkte). Sei $A \in \mathbb{R}^{m\times n}$. Zeigen Sie die folgenden Aussagen:

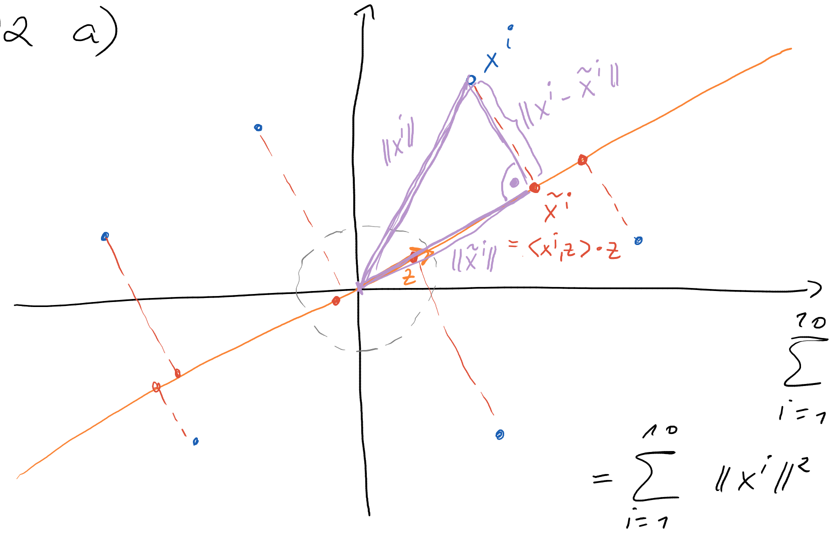
a) (4 Punkte) Wir haben $A^+ = A^T(AA^T)^{-1}$ wenn $\operatorname{rank}(A) = m$ und $A^+ = (A^TA)^{-1}A^T$ wenn $\operatorname{rank}(A) = n$.

b) (3 Punkte) Es gilt $(PAQ)^+ = Q^TA^+P^T$, wobei $P \in \mathbb{R}^{m\times m}$ und $Q \in \mathbb{R}^{n\times n}$ orthogonale Matrizen sind.

c) (3 Punkte) $A^T = A^TAA^+ = A^+AA^T$.

Die Abgabe erfolgt bis zum 15.10.23 über das Online-Übungssystem. Fragen zu den Aufgaben können Sie auch schon während der Bearbeitungszeit im Forum oder in der Online-Übung stellen.

A2 a)



gesucht: $\min\limits_{\substack{z \\ \|z\|=1}} \sum_{i=1}^{10} \underbrace{\|x^i - \tilde{x}^i\|^2}_{= \|x^i\|^2 - \|\tilde{x}^i\|^2}$

Umschreiben:

$$\sum_{i=1}^{10} \|x^i\|^2 - \|\tilde{x}^i\|^2 = \sum_{i=1}^{10} \|x^i\|^2 - |\langle z, x^i\rangle|^2 \underbrace{\|z\|^2}_{=1}$$

$$= \sum_{i=1}^{10} \|x^i\|^2 - \langle z, x^i\rangle^2$$

$$= \sum_{i=1}^{10} \|x^i\|^2 - \sum_{i=1}^{10} \langle z, x^i\rangle^2 \leftarrow \min_z !$$

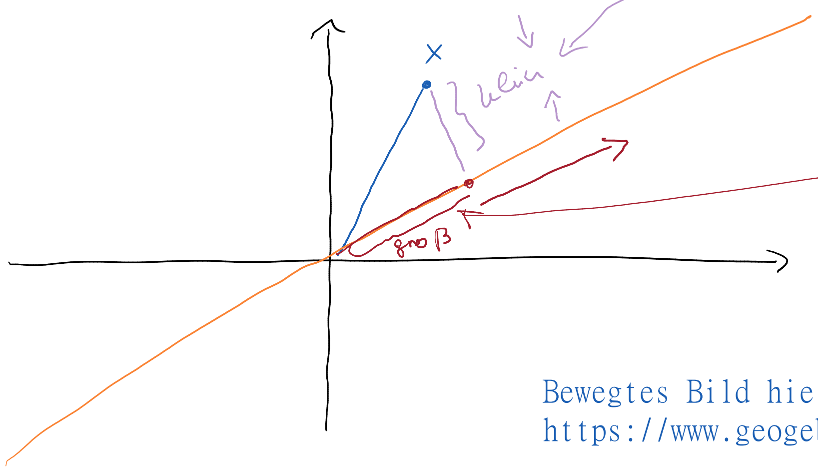
„diejenige z mit $\|z\|=1$, sodass der Term dahinter minimal wird"

$$z^* = \overbrace{\arg\min_{\|z\|=1}} \underbrace{\sum_{i=1}^{10} \|x^i\|^2}_{\text{unabh. von } z} - \sum_{i=1}^{10} \langle x^i, z\rangle^2 = \arg\min_{\|z\|=1} \left(- \sum_{i=1}^{10} \langle x^i, z\rangle^2 \right)$$

$$= \arg\max_{\|z\|=1} \sum_{i=1}^{10} \langle x^i, z\rangle^2$$

geometrische Interpretation:

$$\|x^i\|^2 = \|x^i - \tilde{x}^i\|^2 + \|\tilde{x}^i\|^2 = \underbrace{\|x^i - \tilde{x}^i\|^2}_{\text{möglichst kleiner quad. Abstand zur gerade}} + \underbrace{\langle x^i, z\rangle^2}_{\max!}$$



Bewegtes Bild hier: https://www.geogebra.org/calculator/pfrqwqev

b) Erste Gleichung folgt aus Def. von X. Zweite Gleichung:

$$\sigma = \|X\| = \sup_{z \in \mathbb{R}^n \setminus \{0\}} \frac{\|Xz\|}{\|z\|} = \sup_{z \in \mathbb{R}^n \setminus \{0\}} \|X \overbrace{\tfrac{z}{\|z\|}}^{\text{Länge } 1}\| = \underline{\sup_{\|z\|=1} \|Xz\|}$$

Koroller 1.3.8 $\Rightarrow$ $\|X\| = \sigma$, wobei σ der größte Singulärwert von X ist. Also $\underline{\max_{\|z\|=1} \|Xz\|^2 = \sigma^2}$.

c) Wir wollen zeigen, dass der Wert σ angenommen wird, wenn wir $z = v$ einsetzen.

Wir wissen $v^T v = 1 = \langle v, v\rangle$ und $u^T v = 0$ für alle anderen Spalten u der Matrix V. D.h. $V^T v = e_1 = (1, 0, \dots, 0)^T$. Wir schreiben $\tilde{u}$ für die erste Zeile des Matrix U. Nun gilt

$$\|Xv\|^2 = \|\underbrace{U\Sigma V^T}_{=X} v\|^2 = \|U\underbrace{\Sigma e_1}\|^2 = \|U(\sigma, 0, \dots, 0)^T\|^2$$

$$= \|u\sigma\|^2 = \underbrace{\|u\|^2}_{=1}\sigma^2 = \sigma^2. \qquad \Rightarrow c)$$

A3) $A \in \mathbb{R}^{m\times n}$, σ_r der kleinste Singulärwert

Zeigen:
$$\sigma_r = \min_{\substack{x \in \mathrm{Im}(A^T)\\ \|x\|=1}} \|Ax\|$$
$U\Sigma V^T$

$\underline{\sum \alpha_i v^i}$

Hinweis: $A = U\Sigma V^T$ red. Singulärwertz., dann gilt $\underline{\mathrm{Im}(A^T)} = \underline{\mathrm{span}\{v_1, \dots, v_r\}}$

$v_1, \dots, v_r$ die Spalten von V.

Lösung: Sei $x \in \mathrm{Im}(A^T)$ $(\exists y : A^T y = x)$ und $\|x\| = 1$.

Mit dem Hinweis folgt $x = \sum_{i=1}^r \alpha_i v_i$, wobei $\alpha_i \in \mathbb{R}$. Da $\|x\|^2 = 1$, haben wir:

$\langle \lambda a, b\rangle = \lambda\langle a, b\rangle$

$\langle a+b, c\rangle = \langle a, c\rangle + \langle b, c\rangle$

$$1 = \|x\|^2 = \Big\langle \underbrace{\sum_{i=1}^r \alpha_i v^i}_{x}, \sum_{i=1}^r \alpha_i v^i \Big\rangle \overset{v^i \text{ orth.}}{=} \sum_{i=1}^r \langle \alpha_i v^i, \alpha_i v^i\rangle = \sum_{i=1}^r \alpha_i^2 \underbrace{\langle v^i, v^i\rangle}_{=1}$$

$$\Rightarrow \sum_{i=1}^r \alpha_i^2 = 1$$

Wir haben $V^T x = \sum_{i=1}^r \alpha_i \underbrace{V^T v^i} = \sum_{i=1}^r \alpha_i e_i$. Darum $\Sigma V^T v^i = \sum_{i=1}^r \sigma_i \alpha_i e_i$ und $U \sum_{i=1}^r \sigma_i \alpha_i e_i = \sum_{i=1}^r \sigma_i \alpha_i u_i$ ← Zeilen von U. Also

$$Ax = (U\Sigma V^T)\Big(\sum_{i=1}^r \alpha_i v^i\Big) = \sum_{i=1}^r \sigma_i \alpha_i u_i \quad \text{und} \quad \|Ax\|^2 = \sum_{i=1}^r \sigma_i^2 \alpha_i^2.$$

$$\Rightarrow \min_{\substack{x \in \mathrm{Im}(A^T) \\ \|x\| = 1}} \|Ax\| = \min_{\substack{\alpha \in \mathbb{R}^r \\ \sum \alpha_i^2 = 1}} \sqrt{\sum_{i=1}^{r} \underbrace{\sigma_i^2}_{\geq \sigma_r^2} \underbrace{\alpha_i^2}_{\substack{=1 \quad i=r \\ 0 \quad \text{sonst}}}} = \sigma_r .$$

A4) Rechenregeln: i) $(AB)^{-1} = B^{-1} A^{-1}$

ii) $(AB)^T = B^T A^T$

+ Singulärwertzerl. + $V^T V = I$, $U^T U = I$