Abteilung für
Angewandte Mathematik

WiSe 2022/2023
JProf. Dr. Diyora Salimova, Oliver Suchan

Praktische Übungen zu Numerik I

Projekt 5 – 12.12.2022

Abgabe: per E-Mail bis Freitag, den 23.12.2022, 10:00 Uhr

Homepage zur Vorlesung:

`https://aam.uni-freiburg.de/agsa/lehre/ws22/num`

Projekt 1 (6 Punkte)**.** Ein digitalisiertes Graustufenbild lässt sich mit einer $(m \times n)$-Matix beschreiben, in der der Eintrag (i, j) den Grauwert des Pixels (i, j) kodiert.
Sei nun $A \in \mathbb{R}^{m\times n}$ ein digitalisiertes Bild mit Singulärwertzerlegung $A = U\Sigma V^\top$ und den Singulärwerten $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$, $r = \operatorname{rang}(A)$.
Die beste Rang-k-Approximation A_k an A ist dann gegeben durch

$$A_k = U \begin{pmatrix} \Sigma_k & 0 \\ 0 & 0 \end{pmatrix} V^\top \quad \text{mit} \quad \Sigma_k = \operatorname{diag}(\sigma_1, \ldots, \sigma_k) \in \mathbb{R}^{k\times k}, \quad 1 \leq k \leq r.$$

Wir fragen uns jetzt, ob A_k auch eine gute visuelle Approximation an das durch A gegebene Bild liefert. Man beachte, dass sich A_k alternativ auch als $A_k = U_k \Sigma_k V_k^\top$ schreiben lässt, wobei $U_k \in \mathbb{R}^{m\times k}$ die Matrix bezeichne, die entsteht, wenn man aus der Matrix U die Spalten $k+1$ bis m entfernt. Analog bezeichne $V_k \in \mathbb{R}^{n\times k}$ die Matrix V bis auf die Spalten $k+1$ bis n.
Erweitern Sie obige Idee auf ein RGB-Farbbild, also $A \in \mathbb{R}^{m\times n\times 3}$. Implementieren Sie diese Idee für verschiedene Werte von k und beurteilen sie das Verhältnis von Qualitätsverlust zur Reduktion des Speicheraufwands für verschiedene Werte von k und berechnen Sie außerdem den Fehler $\|A - A_k\|_{\mathcal{F}}$.
Verwenden Sie zum Laden des Bildes `weihnachtsbaum.jpg` die Vorlagen auf der Vorlesungswebseite. Zur Berechnung der Singulärwertzerlegung können Sie die Funktionen `svd` (Matlab) und `numpy.linalg.svd` (Python) verwenden.

Projekt 2 (6 Punkte)**.** Eine Firma stellt m verschiedene Produkte her, für deren Fertigung n Maschinen benötigt werden. Die j-te Maschine hat eine maximale monatliche Laufzeit von l_j Stunden. Das k-te Produkt generiert pro Mengeneinheit einen Ertrag von e_k Euro und belegt die j-te Maschine mit t_{jk} Stunden pro Mengeneinheit. Der monatliche Gesamtertrag soll ohne Überschreitung der Maximallaufzeiten optimiert werden.

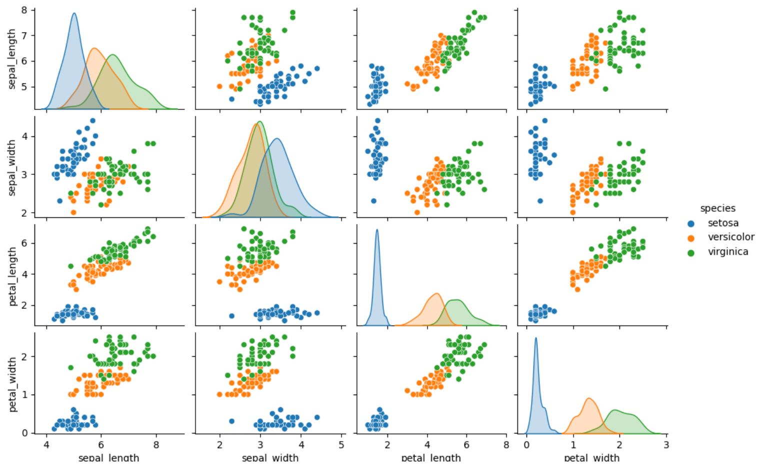
(a) Formulieren Sie den beschriebenen Sachverhalt als Maximierungsproblem mit Nebenbedingungen in der Form

$$\text{Maximiere } f(x) = c \cdot x \text{ unter den Bedingungen } Ax \leq b, x \geq 0$$

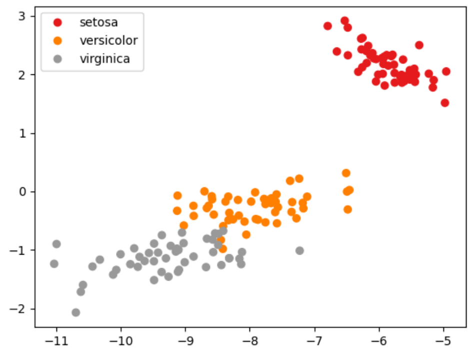
wobei $x = (x_1, x_2, \ldots, x_m)$ die monatlichen Mengeneinheiten der verschiedenen Produkte seien und die Ungleichungen komponentenweise zu verstehen sind.

(b) Verwenden Sie `linprog` (Matlab) bzw. `scipy.optimize.linprog` (Python), um das Problem für die Daten $m = 2, n = 3, e_1 = 200, e_2 = 600$, und $t_{11} = 1, t_{21} = 1, t_{31} = 0, t_{12} = 3, t_{22} = 1, t_{32} = 2$ sowie $l_1 = 150, l_2 = 180, l_3 = 140$ zu lösen. Wie groß ist der optimale monatliche Ertrag?

Bonusprojekt (6 Bonuspunkte)**.** Im Folgenden wollen wir die *Hauptkomponentenanalyse (Principal Component Analysis, PCA)* als Anwendung der Singulärwertzerlegung kennenlernen. Die PCA kann zur Dimensionsreduktion verwendet werden. Betrachten Sie den folgenden Plot des 'Iris' Datensatzes:



In diesem Datensatz wurden für drei Schwertlilien ('setosa', 'versicolor' und 'virginica') jeweils vier Merkmale gemessen und notiert. Der Plot visualisiert die verschiedenen Kombinationen der Ausprägungen. Mit der PCA können wir nun eine sogenannte *Feature Reduktion* durchführen, um die Pflanzengattung anhand weniger Merkmale klassifizieren zu können. Dabei erhält die PCA so viel Varianz wie möglich in dem Datensatz. Wir wollen nun konkret die $2d$-Hyperebene im $4d$-Iris-Merkmalsraum finden, in welcher die Daten am stärksten verteilt sind, um so potentiell 'Cluster' zur Identifzierung der Gattung zu erhalten (siehe folgender Plot):



Laden Sie dazu den Iris-Datensatz (siehe Vorlesungshomepage) und berechnen Sie die Singulärwertzerlegung der Daten $X = U\Sigma V^\top$. Die reduzierten Daten können Sie dann via $T_k = U_k\Sigma_k = XV_k$ (siehe Aufgabe 1 für Notation) berechnen, wobei in unserem Fall $k = 2$ gewählt wurde. Plotten Sie nun die reduzierten Daten T_k.