Einführung in das Wissenschaftliches Rechnen

Aufgabenblatt 1

Abgabe: 19.5.2022, 18:00

Sommer 2023

- Die Aufgaben sind von jeder/m Studierenden *einzeln* zu bearbeiten und abzugeben (insbesondere sollte jede/r ihren/seinen eigenen Code schreiben). Plagiate werden entsprechend der Studienordnung geahndet.
- Verwenden Sie den vorgegebenen Skelett-Code. Die zu implementierenden Funktionen befinden sich in der Datei **main.py**. Ihr Code ist an den mit `# TODO:` gekennzeichneten Stellen einzufügen. Wenn nicht in der Aufgabe angegeben sind die `NumPy`-Funktionen, welche Sie zur Lösung einer Aufgabe nicht verwenden dürfen, unter `Forbidden` in der Docstring Beschreibung der entsprechenden Funktion aufgelistet.
- Bitte reichen Sie `main.py` mit Ihrem Lösungen und ein PDF mit den Lösungen der theoretischen Aufgaben (wenn möglich mit Latex erstellt, eine geeignete Vorlage finden Sie im Elearning) bis zum Abgabedatum ein.

Aufgabe 1: Mittelwert und Varianz von Klimadaten (7 Punkte)

In den Naturwissenschaften entstehen bei Messungen und Experimenten oft sehr große Datenmengen. Eine zuverlässige Auswertung dieser Daten ist essentiell, um daraus verlässliche wissenschaftliche Erkenntnisse abzuleiten. Als Beispiel werden wir in dieser Aufgabe atmosphärische Messungen aus dem ERA5-Datensatz verwenden, welcher historische Wettermessungen in einem konsistenten Format zur Verfügung stellt.[1] Konkret werden wir mit Temperaturwerte an 5 ausgewählten Orten (Longitude (Längengrad) $0°$ und Latituden (Breitengrade) $\{84.4°N, 78.79°N, 67.59°N, 56.38°N, 45.18°N\}$) zwischen den Jahren 1979 und 2008 arbeiten. Die zeitliche Auflösung der Daten ist 3 Stunden. Die Daten sind in `temperature_locs.dat` als Binärdateien und in double precision gespeichert mit allen Daten für eine Latitude gefolgt bei denen für die nächste.

Seien $X = \{x_i\}_{i=0}^{n}$ die Temperaturwerte für einen Ort. Der empirische Erwartungswert $\bar{X} = X^{(0)}$ und die Varianz $X^{(1)}$ (d.h. der erste statistische Moment) von X sind gegeben durch:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1a}$$

$$X^{(1)} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2. \tag{1b}$$

Berechnet man die Varianz mit der obigen Definition, dann muss der Datensatz zwei Mal durchlaufen werden, d.h. es ist ein 2-Pass Algorithmus. Mit nur einem Durchlauf kann die Varianz mit folgender Formel berechnet werden:

$$X^{(1)} = \frac{1}{n}\left(\sum_{i=1}^{n}(x_i)^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right). \tag{2}$$

a) Zeigen Sie, dass die 1-Pass Formel mathematisch äquivalent zur 2-Pass Berechnung ist. (1 Punkte)

b) Implementieren Sie Gl. 1 und Gl. 2 zur Berechnung der Varianz in `compute_variance_1()`, `compute_variance_2()` und `compute_variance_2_Kahan()`. Verwenden Sie entsprechend für

[1] Die Daten sind aus dem ERA5-Datensatz extrahiert, Grundlage für Wetter- und Klimavorhersagen und Forschung zu diesem Thema dient. Nähere Informationen findet man z.B. unter `https://climate.copernicus.eu/climate-reanalysis`.

Gl. 2 sowohl eine naive Summierung als auch Kahans Algorithmus. Die Eingabe sollte ein 1-dimensionales Numpy array sein. (1.5 Punkte)

c) Implementieren Sie die Funktion `load_temperature_data()`, welche die Daten aus der Datei `./temperature_locs.dat` in ein Numpy array lädt, so dass die Zeilen (d.h. der erste Index) den verschiedenen Breitengrade entsprechen. (1 Punkt)

d) Berechnen Sie die Varianz der Temperatur an den 5 Orten mit den drei Implementierungen in single precision Gleitkommazahlen in `compute_variances()`. Bestimmen Sie jeweils den relativen Fehler in `compute_rel_errs()` und geben sie diese von der Funktion zurück. Verwenden Sie als Referenzlösung `np.var()` mit Berechnungen in double precision. (1.5 Punkte)

e) Berechnen Sie die jährlichen Durchschnittstemperatur für jede Latitude (sie können Schaltjahre ignorieren) in der Funktion `compute_mean_temperatures_yearly()`. Können Sie die globale Erderwärmung als Trend in den Daten erkennen? (1 Punkte)

- Zusatz: Wie können Sie Ihre Argumentation bzgl. dem Datentrend robuster machen? Implementieren Sie einen solchen Ansatz in `compute_mean_temperatures_yearly_improved()` und argumentieren sie im schriftlichen Teil, warum dieser geeignet ist. Verwenden Sie ggf. geeignete Plots. (2 Punkte)

f) Vergleichen Sie die Geschwindigkeit Ihrer Implementierungen zu Numpy's `np.var()`. Erklären Sie Ihre Beobachtung im schriftlichen Teil. (1 Punkt)

Aufgabe 2: Theorie (6 Punkte)

a) Berechnen Sie die Maschinengenauigkeit für das IEEE 754-2008 fp16 half precision Gleitkommazahl-Format. (2 Punkt)

b) Leiten Sie die Koordinatenform für das Skalarprodukt $\langle \vec{a}, \vec{b} \rangle$ mit $\vec{a}$, $\vec{b} \in \mathbb{R}^n$ anhand dessen Eigenschaften (Linearität, Symmetrie) her. (1 Punkt)

c) Leiten Sie eine Scherungsmatrix in $\mathbb{R}^2$ anhand der geometrischen Interpretation her (analog zur Herleitung einer Rotationsmatrix in der Übung). (1 Punkt)

d) Gegeben sei die Matrix

$$P = \begin{pmatrix} -2 & -4 \\ \frac{3}{2} & 3 \end{pmatrix}. \tag{3}$$

Zeigen Sie, dass für jeden Vektor $\vec{v} \in \mathbb{R}^2$, der Vektor $P\vec{v}$ ein Vielfaches von $\vec{u} = \left(1, -\frac{3}{4}\right)^T$ ist. Welche Eigenschaft muss $\vec{v}$ erfüllen damit $P\vec{v} = \vec{0}$? Welche geometrische Operation implementiert die Matrix P. (1 Punkt)

e) Bestimmen Sie die Matrix für die lineare Transformation $A : \mathbb{R}^2 \to \mathbb{R}^2$ so dass $A\vec{u}_i = \vec{v}_i$ für $u_1 = (1, 1)^T$, $v_1 = (2, 3)^T$ und $u_2 = (-1, 1)^T$ $v_2 = (4, 5)^T$. (1 Punkt)