

Wissenschaftliches Rechnen - Großübung 7

Klausurvorbereitung

Ugo & Gabriel

14. Februar 2023

Aufgabe 1: Hauptkomponentenanalyse

1. Ein Beispiel für eine Hauptkomponentenanalyse durchführen, zum Beispiel mit dem Datensatz $\{(0, 2), (1, -1), (-1, -1)\}$.
2. Was ist der Zusammenhang zwischen der Hauptkomponentenanalyse (PCA) und der Singulärwertzerlegung (SVD)?
3. Wie führe ich eine Hauptkomponentenanalyse durch?
4. Sowohl bei PCA als auch bei der Fourier-Transformation führen wir einen Basiswechsel durch. Welche Gemeinsamkeiten und Unterschiede gibt es bei diesen Verfahren?

Lösung

1. Im Folgenden nutzen wir den Algorithmus aus der praktischen Hausaufgabe:
 - a) Wir berechnen den Mittelwert und ziehen diesen von jedem Punkt ab. Es ist hier nichts zu tun, da das arithmetische Mittel der Punkte $(0, 0)$ ist.
 - b) Wir schreiben die Daten in eine Matrix:

$$\mathbf{X} = \begin{bmatrix} 0 & 2 \\ 1 & -1 \\ -1 & -1 \end{bmatrix}$$

- c) Wir berechnen eine SVD der Datenmatrix, also $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$.
- d) Uns interessiert die erste Hauptkomponente, was die erste Spalte von \mathbf{V} ist.

$$\mathbf{V} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Damit ist $\mathbf{v}_1 = (0, 1)^T$.

2. Die Singulärwertzerlegung bietet uns eine Möglichkeit, die Hauptkomponenten eines Datensatzes zu ermitteln.
3. Dazu gibt es mehrere Möglichkeiten:
 - a) Der Algorithmus aus Teilaufgabe 1.

- b) Statt die SVD von \mathbf{X} zu berechnen, können wir auch die Hauptkomponenten mittels einer Diagonalisierung von $\mathbf{X}^T \mathbf{X}$ berechnen. Schließlich sind die rechten Singulärvektoren die Eigenvektoren jener Matrix. Alternativ kann man auch die Potenzmethode verwenden.
- c) Iterative Algorithmen wie Gradient Descent oder Newton-Verfahren funktionieren ebenfalls.

4. Folgende Gemeinsamkeiten gibt es:

- a) Bei beiden führen wir den Basiswechsel für effizientere Berechnungen durch (bei PCA geht die (k)-Nearest-Neighbor Klassifikation schneller, da wir in eine reduzierte Basis projizieren und die Vektoren weniger Einträge haben; in der Fourier-Basis geht die Faltung in asymptotisch linearer Laufzeit)
- b) Bei beiden Basen handelt es sich um Eigenbasen (bei PCA die Eigenbasis von $\mathbf{X}^T \mathbf{X}$; bei der DFT ist es die Eigenbasis aller Faltungsmatrizen bzw. der zyklischen Verschiebungsmatrizen)
- c) Beides sind orthonormale Basen.

Folgende Unterschiede gibt es jedoch:

- a) Die Basisvektoren sind unterschiedlich. Die PCA Basisvektoren haben die Eigenschaft, möglichst viel Informationen (Varianz) der Daten in möglichst wenigen Vektoren zu rekonstruieren. Die Fourier-Basisvektoren zerlegen das (per Annahme periodische) Signal (mit beschränkter Bandbreite) in die Anteile zu verschiedenen Frequenzen.
- b) Bei PCA zielen wir darauf ab, die Daten auf deutlich weniger Basisvektoren zu projizieren als wir Dimensionen (oder Datenpunkte) haben. Falls die Anzahl der Hauptkomponenten k kleiner als das Minimum von n und d ist (Anzahl an Punkten bzw. Dimensionalität), so ist der Basiswechsel nicht verlustfrei. Die Fourier-Transformation hingegen ist verlustfrei.
- c) Die Fourier-Basis ist, selbst bei reellen Daten, nicht reell (für $n > 2$). Die Basis der Hauptkomponenten ist hingegen (bei reellen Daten) immer reell.

Lösung Ende

Aufgabe 2: Cholesky-Zerlegung

1. Wie kann man erkennen, ob eine Matrix die Kriterien für die Existenz einer Cholesky-Zerlegung (symmetrisch und positiv (semi)definit erfüllt)?
2. Wie kann man dies effizient für größere Matrizen tun (3×3 , 4×4 Matrizen)?

Lösung

1. Im Folgenden fordern wir, dass die Matrix strikt positiv definit ist, da lineare Gleichungssysteme mit singulärer Matrix 0 oder ∞ Lösungen haben. Es gibt zwei notwendige sowie hinreichende Kriterien für symmetrische Matrizen:

- a) Eigenwertkriterium: $\mathbf{A} \in \mathbb{R}^{n \times n}$ ist genau dann positiv definit, wenn alle Eigenwerte größer als 0 sind.

Hinweis: Symmetrische Matrizen haben ausschließlich reelle Eigenwerte.

- b) Hauptminorenkriterium: $\mathbf{A} \in \mathbb{R}^{n \times n}$ ist genau dann positiv definit, wenn alle führenden Hauptminoren (das sind die Determinanten aller quadratischen Untermatrizen beginnend oben links) größer als 0 sind.
2. Zwar kann man Eigenwerte für 2×2 -Matrizen noch ziemlich einfach ermitteln, ab 3×3 -Matrizen wird es sehr aufwendig. Prinzipiell ist es in jedem Fall effizienter das Hauptminorenkriterium anzuwenden. Mithilfe der Regel von Sarrus lässt sich die Determinante einer 3×3 -Matrix effizient bestimmen, jedoch wird es ab 4×4 -Matrizen schwierig. Falls eine Matrix ein nicht-positives Element auf der Diagonale zu stehen hat, so kann sie nicht positiv definit sein.
- Eine alternative Möglichkeit wäre es die Cholesky-Zerlegung zu berechnen und zu schauen, ob es an einer Stelle schief geht.

Lösung Ende

Aufgabe 3: Hessematrix

1. Welche Informationen liefert die Hessematrix und welche geometrische Bedeutung hat sie?
2. Welchen Zusammenhang zum Gradienten hat sie?
3. Was passiert mit der Hessematrix und ihren Eigenschaften (insbesondere Definitheit), wenn man eine Fourier-Transformation auf den Punkt durchführt?

Lösung

1. Ähnlich zur zweiten Ableitung im Eindimensionalen liefert sie uns Aussagen über die Krümmung der Funktion. Wir betrachten beispielsweise eine Funktion $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, welche eine 2×2 -Hessematrix hat:

$$\mathbf{H}_f(x, y) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2}(x, y) & \frac{\partial^2 f}{\partial y \partial x}(x, y) \\ \frac{\partial^2 f}{\partial x \partial y}(x, y) & \frac{\partial^2 f}{\partial y^2}(x, y) \end{bmatrix}$$

Wir nehmen an, dass f zweimal stetig differenzierbar ist. Anders als im eindimensionalen Fall, haben wir hier nicht eine sondern vier bzw. drei zweite Ableitungen. Die zwei Terme auf der Diagonalen kann man sich einfach vorstellen. Diese entsprechen den regulären zweiten Ableitungen der Parameterkurven, also eindimensionalen Funktionen, wenn man einen der beiden Parameter fest wählt. Die Terme außerhalb der Diagonalen sind eher schwierig zu interpretieren, deshalb schauen wir uns die Hessematrix als Ganzes an.

Mithilfe der quadratischen Form $\mathbf{u}^T \mathbf{H}_f \mathbf{u}$ lässt sich die Krümmung der Funktion in jede Richtung berechnen (zweite Richtungsableitung). Dies ist so ähnlich wie mit dem Gradienten, mit dem man die Richtungsableitung in jede Richtung ermitteln kann $\frac{\partial f}{\partial \mathbf{u}}(\mathbf{x}) = \mathbf{u}^T \nabla f(\mathbf{x})$. Insbesondere bedeutet eine positiv definite Hessematrix ($\mathbf{u}^T \mathbf{H}_f \mathbf{u} > 0$ für alle $\mathbf{u} \neq \mathbf{0}$), dass die Funktion in jede Richtung linksgekrümmt ist.

2. Die Hessematrix ist die Ableitung (Jacobi-Matrix) des Gradienten. Ihre Einträge zeigen, wie stark welche Komponente des Gradienten entlang der Basisvektoren wachsen, d.h. wie sich der Gradient verändert. Da dieser orthogonal zu den Niveaulinien steht, kann man mithilfe der Hessematrix auch deren Krümmung bestimmen.

3. Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}$ eine skalare Funktion. Bevor wir die Funktion auf den Raum ausführen, führen wir eine Fourier-Transformation $\mathbf{x} \mapsto \Omega \mathbf{x}$ durch. Aus dem letzten Aufgabenblatt weiß man, dass die Hessematrix von $g(\mathbf{x}) = f(\mathbf{A}\mathbf{x})$ dem Term $\mathbf{H}_g(\mathbf{x}) = \mathbf{A}^T \mathbf{H}_f(\mathbf{A}\mathbf{x}) \mathbf{A}$ entspricht (bei einer linearen Transformation wird die Hessematrix zwei mal transformiert, einmal von vorne und einmal von hinten).

Zunächst ergibt sich aber das Problem, dass die Fourier-Transformation in einen komplexen Vektorraum transformiert. Die Regel ist für komplexe Funktionen analog zum Reellen, statt der transponierten Matrix \mathbf{A}^T nimmt man aber die adjungierte \mathbf{A}^H . Außerdem ist die Definitheit von komplexen Matrizen über eine leicht andere quadratische Form $\mathbf{z}^H \mathbf{A} \mathbf{z}$ definiert. Dann gilt aber für $g(\mathbf{x}) = f(\Omega \mathbf{x})$:

$$\mathbf{z}^H \mathbf{H}_g(\mathbf{x}) \mathbf{z} = \mathbf{z}^H \Omega^H \mathbf{H}_f(\Omega \mathbf{x}) \Omega \mathbf{z} = (\Omega \mathbf{z})^H \mathbf{H}_f(\Omega \mathbf{x}) (\Omega \mathbf{z}) = \mathbf{y}^H \mathbf{H}_f(\Omega \mathbf{x}) \mathbf{y}$$

Also ändert sich die Definitheit der Hessematrix durch die Transformation nicht. Falls $\Omega^{-1} \mathbf{z}^*$ ein Extremum von f ist, ist \mathbf{z}^* ein Extremum von g selbiger Art. Dies ergibt Sinn, da lediglich der Raum transformiert wird und f im transformierten Raum genauso aussieht wie im ursprünglichen Raum.

Ein paar zusätzliche Bemerkungen:

- Die Hessematrix ist für 2 mal stetig differenzierbare Funktionen nicht mehr symmetrisch, sondern hermitesch.
- Differenzierbarkeit hat im Komplexen eine andere Definition als im Reellen.
- Im Komplexen können Funktionen mehr Extrema haben als im Reellen.

Lösung Ende