

Prof. Dr. Stefan Posch

Dr. Birgit Möller

(birgit.moeller@informatik.uni-halle.de)



Institut für Informatik
Universität Halle

Blatt 6

Aufgabe 6.1 (4 Punkte)

Beweisen Sie die folgenden Eigenschaften.

Für quadratische Matrizen B gilt:

- $\vec{x}^T B \vec{y} = \text{Sp}(B \vec{y} \vec{x}^T)$ mit $\text{Sp}(B) :=$ Summe über die Diagonalelemente
- $\vec{x}^T B \vec{y} = \text{Sp}(\vec{y} \vec{x}^T B^T)$
- $\sum_{\nu=1}^n \vec{a}_\nu^T B \vec{a}_\nu = \text{Sp}(A^T A B^T)$

Für eine $M \times n$ -Matrix A^T , $A^T = (\vec{a}_1, \dots, \vec{a}_n)$, gilt:

- $A^T A = \sum_{\nu=1}^n \vec{a}_\nu^T \vec{a}_\nu$

Aufgabe 6.2 (2 Punkte)

Zeigen Sie, dass für den Erwartungswert $E[\cdot]$ allgemein gilt: $\sum_{i=1}^n E[X_i] = E[\sum_{i=1}^n X_i]$.

Aufgabe 6.3 (6 Punkte)

Die Hauptkomponentenanalyse wird üblicherweise in Klassifikationssystemen zur Reduktion der Dimension von Merkmalsvektoren eingesetzt. Dabei hat die Auswahl einer geeigneten Anzahl von Eigenvektoren oftmals einen signifikanten Einfluss auf das spätere Ergebnis. Dieser Zusammenhang soll hier untersucht werden.

Die Grundlage der Experimente ist der Gen-Datensatz in der Datei "GenDaten.txt" auf der Webseite zur Vorlesung. Teilen Sie die darin enthaltenen Daten zunächst in einen Trainingsdatensatz mit 20 Elementen und eine Teststichprobe mit 5 Elementen auf, deren Klassenzugehörigkeit damit bekannt ist.

Nun soll der Einfluss einer Dimensionsreduktion auf das Klassifikationsergebnis untersucht werden, das mit einem Nearest-Neighbor-Klassifikator erzielt werden kann. Führen Sie dazu für den Trainingsdatensatz eine Hauptkomponentenanalyse durch und plotten Sie das Eigenspektrum. Ermitteln Sie die Klassifikationsraten des NN-Klassifikators auf dem Testdatensatz für verschiedene Anzahlen von Eigenvektoren.

Visualisieren Sie die Merkmalsvektoren der Trainings- und Testdaten bei der Verwendung von nur einem, den ersten beiden und den ersten drei Eigenvektoren. Mit welcher Anzahl von Eigenvektoren lässt sich das beste Klassifikationsergebnis erzielen?