

Interaktive Visuelle Datenanalyse I
—
Praktikum 2

Abgabetermin: 10.06.2025, 23:00 Uhr
Abnahme: ab dem 11.06.2025
Abgabeformat: zip für die Programme
Maximale Anzahl an Punkten: 15

Lineare Regression, K-Means Clustering, PCA, Scatterplot, Interpretation der Ergebnisse

Datensatz: Wein.csv

Beachten Sie die Anforderungen aus der Einführung in das Praktikum!

1. Datenbereinigung (1 Punkt)

- Eingabe: Daten aus dem Datensatz
- Ausgabe: Bereinigte Daten
- Untersuchen Sie die Datenqualität.
- Identifizieren Sie mögliche Probleme in den Daten und beheben Sie die Fehler, falls möglich.
- Dokumentieren Sie Ihr Vorgehen.

2. Lineare Regression (4 Punkte)

- Eingabe: Bereinigte Daten aus dem Datensatz
- Berechnen Sie die paarweisen linearen Regressionen der Attribute.
- Stellen Sie diese visuell dar.
- Geben Sie die Qualitätswerte der Regressionen aus.

3. K-Means Clustering (7 Punkte)

- Eingabe: Bereinigte Daten aus dem Datensatz
- Clustern Sie die Daten mittels k-means.
 - Wählen Sie eine geeignete Distanzmetrik für die Daten.
 - Transformieren Sie die Daten, falls notwendig.
 - Berechnen Sie k-means Clusterings mit 2, 3, 4 und 5 Clustern.
- Projizieren Sie die Daten mittels der Dimensionsreduktionsmethode PCA auf 2 Dimensionen.
- Stellen Sie die Ergebnisse des Clusterings visuell als Scatterplot unter Verwendung der Projektion dar. Nutzen Sie für die Darstellung der Datenpunkte als Einfärbung die Zugehörigkeit zum berechneten Cluster.
- Stellen Sie den Scree Plot der PCA dar.
- Berechnen Sie
 - den Silhoutte Coefficient,

-
- den Davies Bouldin Index.
 - Stellen Sie die Kurve
 - des Silhoutte Coefficient,
 - des Davies Bouldin Indexder verschiedenen Läufe visuell dar.
 - Bestimmen Sie anhand der Kurven, welche Clusteranzahl jeweils am Besten für diese Daten geeignet ist.
4. Analysieren/Interpretieren Sie die Ergebnisse. (*3 Punkte*)