

2. Praktikumsaufgabe IVDA

- a) Lesen Sie den Datensatz „wein.csv“ ein.
- b) Berechnen Sie die paarweisen linearen Regressionen der Attribute und stellen Sie diese visuell dar. Geben Sie zudem die Qualitätswerte der Regressionen aus.
- c) Clustern Sie die Daten mittels k-means. Wählen Sie eine geeignete Distanzmetrik für die Daten und transformieren Sie die Daten, falls notwendig.
Berechnen Sie k-means Clusterings mit 2, 3, 4 und 5 Clustern und evaluieren Sie mittels des Silhouette Coefficient und des Davies Bouldin Index, welche Clusteranzahl am Besten für diese Daten geeignet sind. Stellen Sie hierfür die Kurve des Davies Bouldin Index und die Silhouette Coefficients der verschiedenen Läufe visuell dar.
- d) Stellen Sie die Ergebnisse des Clusterings visuell als Scatterplot dar. Nutzen Sie hierfür die Dimensionsreduktionsmethode PCA. Nutzen Sie für die Darstellung der Datenpunkte als Einfärbung die Zugehörigkeit zum berechneten Cluster. Stellen Sie zudem den Scree Plot der PCA dar.