

Übung: Woche 3 – Klassifikation und Clusteranalyse

1. Wenden Sie die Klassifikation mittels kNN beim Titanic-Datensatz an.
 1. Testen Sie verschiedene Werte für die Anzahl an Nachbarn k . Beschreiben Sie das Verhalten.
 2. Wenden Sie nun zuerst die PCA an, bevor Sie die Daten mit kNN klassifizieren. Testen Sie verschiedene Werte für die Anzahl an Hauptkomponenten n . Beschreiben Sie das Verhalten und vergleichen Sie die Ergebnisse mit Teilaufgabe 1.1.
2. Wenden Sie nun kMeans-Clustering und DBScan auf den Daten an.
 1. Berechnen Sie den Silhouette-Score für verschiedene k und vergleichen Sie die Ergebnisse.
 2. Plotten Sie die Ergebnisse für $k=2$ für kMeans und DBScan, indem Sie die Daten **nach** dem Clustering auf zwei Dimension reduzieren (also PC1 gegen PC2 plotten) und die Punkte entsprechend ihrer Clusterzugehörigkeit einfärben. Vergleichen Sie die Ergebnisse.