

# 1. Übungsblatt

(Statistische Klassifikatoren)

Zu Beginn werden auf einem Beispieldatensatz Klassifikatoren implementiert. Der data2d Datensatz enthält die zwei-dimensionalen Trainingsmerkmale der drei Musterklassen. Für diese Daten soll eine Klassifikation über Nächster-Nachbar und Bayesklassifikatoren realisiert werden.

Sie finden die Aufgaben in den Modulen *blatt1.aufg1* und *blatt1.aufg2*. Folgen Sie den Beschreibungen im Quelltext.

**Achtung:** Gestalten Sie Ihre Implementierung so, dass Sie die Klassifikatoren für zukünftige Aufgaben wiederverwenden können.

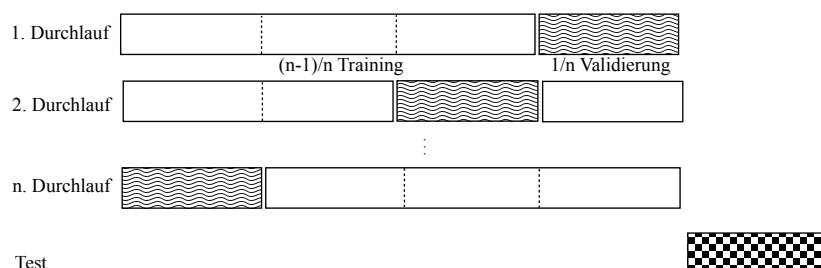
**1.Aufgabe:** Zunächst sollen Nächste-Nachbar-Klassifikatoren (NN) implementiert werden.

1. Laden Sie die Trainingsdaten und visualisieren Sie sie.
2. Implementieren Sie einen NN-Klassifikator (*Nearest-Neighbour*) welcher als Metrik eine  $p$ -Norm verwendet (z. B. 1-Norm: Cityblock-Metrik, 2-Norm: Euklidischer Abstand und  $\infty$ -Norm: Maximumnorm).
3. Ein NN-Klassifikator alleine ist meist nicht ausreichend. Erweitern Sie den Klassifikator zum  $k$ -NN Klassifikator.

Für einen  $k$ -NN Klassifikator muss der Parameter  $k$  gewählt werden. Bestimmen Sie den optimalen Parameter über eine **Kreuzvalidierung**.

Bei der Kreuzvalidierung werden die Trainingsdaten in  $n$  gleich grosse Teilmengen geteilt. In  $n$  Durchläufen wird dabei jeweils eine Teilmenge für die Validierung verwendet und die jeweils verbleibenden Teilmengen zum Training. Die Ergebnisse der Kreuzvalidierung ergeben sich durch Mittelung über die Durchläufe. Die Testdaten bleiben davon unabhängig.

- Warum sollen Sie hier den Parameter  $k$  über eine Kreuzvalidierung bestimmen? Was wäre die Alternative dazu und warum ist das problematisch?
- Bei welchen Parametern von Klassifikatoren ist eine Kreuzvalidierung allgemein von Bedeutung? Schauen Sie sich in diesem Zusammenhang auch die Parameter des Normalverteilungsklassifikators (Aufgabe 2) an.



**2. Aufgabe:** Realisieren Sie einen Bayes'schen Normalverteilungs-Klassifikator mit drei Klassen.

1. Berechnen Sie die Normalverteilungen der drei Klassen und visualisieren Sie diese.
2. Implementieren sie einen Bayes'schen Normalverteilungs-Klassifikator (ohne Rückweisung), der die soeben berechneten Verteilungen als Modellparameter verwendet. Evaluieren Sie den erzeugten Klassifikator: Bestimmen Sie dazu die Klassifikationsfehlerrate sowohl für den gesamten Datensatz als auch klassenspezifisch.
3. Ist der erstellte Klassifikator für diese Daten geeignet? Vergleichen Sie die Ergebnisse mit dem ( $k$ )-NN Klassifikator.
4. Diskutieren Sie Möglichkeiten für eine automatische Rückweisung.