

**Blatt 2: Parametrische Klassifikatoren****Aufgabe 1: Bayes-Klassifikator**

Diese Teilaufgabe ist handschriftlich zu bearbeiten.

**Bayes-Klassifikator:**

In einem Topf liegen 3000 normale Würfel (Klasse 1) und 2000 gezinkte Würfel (Klasse 2). Es wird jeweils zufällig ein Würfel gezogen und eine Augenzahl gewürfelt, und dann wird der Würfel zurückgelegt. Der Klassifikator soll dann entscheiden, ob es ein normaler oder gezinkter Würfel war.

Die klassenbedingten Wahrscheinlichkeiten der Muster (= Augenzahl) sind:

$$\text{Klasse 1: } p(x|\omega_1) = \frac{1}{6}, x \in 1, \dots, 6,$$

und

$$\text{Klasse 2: } p(x|\omega_2) = \begin{cases} 0 & \text{für } x = 1, \\ \frac{1}{6} & \text{für } x \in \{2, 3, 4, 5\}, \\ \frac{1}{3} & \text{für } x = 6 \end{cases}$$

1. Berechnen Sie die *a-priori* Wahrscheinlichkeiten  $P(\omega_i), i = 1, 2$ .
2. Entwerfen Sie den Bayes-Klassifikator anhand des MAP-Kriteriums zur optimalen Klassifizierung eines beliebig gezogenen Würfels anhand der Augenzahl eines Wurfes.
3. Berechnen Sie die Fehlerwahrscheinlichkeit des Bayesklassifikators und vergleichen Sie diese mit der Fehlerwahrscheinlichkeit des (naiven) Klassifikators:

$$g(x) = 1 \text{ für } x \in \{1, 2, 3, 4, 5, 6\} \text{ beliebig.}$$

**Aufgabe 2: Polynom-Klassifikator****Datenerzeugung:**

- Erzeugen Sie jeweils 1000 Samples zweier ( $c = 2$ ) teilweise überlappender normalverteilter Klassen für  $d = 2$  und visualisieren Sie diese.
- Erzeugen Sie einen zweiten Datensatz, in dem eine der beiden Klassen **nicht unimodal** verteilt ist (also z.B. zwei oder mehr deutlich abgesetzte Klassenzentren aufweist).

**Klassifikatortraining:**

Cl. Schnörr

Hochschule München, FK Informatik und Mathematik

Trainieren bzw. berechnen Sie für beide Datensätze jeweils einen **quadratischen** Polynomklassifikator wie im Skript (eigene Implementierung) mit 70% der Daten und visualisieren Sie die Trennfunktion.

Bestimmen Sie  $P_F$  mit den restlichen 30% der Daten.

**Aufgabe 3: Logistic-Regression-Klassifikator****Klassifikatortraining:**

Trainieren Sie für beide Datensätze aus Aufgabe 2 jeweils einen Logistic-Regression-Klassifikator mittels einer Bibliothek in Matlab/Python mit 70% der Daten und visualisieren Sie die Trennfunktion.

Wie werden die Parameter über eine Optimierung bestimmt?

Bestimmen Sie  $P_F$  mit den restlichen 30% der Daten.

Visualisieren Sie die Trennfunktion und bestimmen Sie  $P_F$  nun für quadratisch nichtlinear transformierte Merkmale und vergleichen Sie das Ergebnis mit dem obigen quadratischen Polynomklassifikator.

Cl. Schnörr

Hochschule München, FK Informatik und Mathematik