

2. Übungszettel Mustererkennung SS15

Prof. Raúl Rojas, Daniel Göhring, Fritz Ulbrich
Institut für Informatik, Freie Universität Berlin
Abgabe Online bis Montag, 04.05.15, 24 Uhr

Laden Sie die Dateien **chickwts_testing.csv** und **chickwts_training.csv** aus dem Resources Ordner der KVV-Seite herunter. Jede Zeile dieser Dateien ist ein Datensatz über ein Huhn mit folgenden Werten: **id**, **Gewicht**, **Futterart**. (Die Futterart ist von 0 bis 5 indiziert und entspricht: 0=horsebean, 1=linseed, 2=soybean, 3=sunflower, 4=meatmeal, 5=casein)

1. Aufgabe (4 Punkte): K-NN

Schreiben Sie in matlab/octave ein Script, das die Datensätze aus **chickwts_testing.csv** anhand der Datensätze aus **chickwts_training.csv** mit dem K-NN-Algorithmus klassifiziert und die Konfusionsmatrix und Klassifikationsgüte (korrekt klassifizierte Individuen geteilt durch Gesamtzahl der Individuen) ausgibt. Geben Sie die Konfusionsmatrix und Klassifikationsgüte jeweils für $K = 1, 3$ und 5 an.

2. Aufgabe (6 Punkte): Normalverteilung

- (2 Punkte) Berechnen Sie die **Normalverteilung** (Erwartungswert und Varianz) über den Gewichten jeweils für alle 6 Futterklassen sowie die **A-priori-Wahrscheinlichkeit** für jede der 6 Futterklassen anhand der Werte aus **chickwts_training.csv**.
- (1 Punkte) Plotten Sie die A-posteriori-Wahrscheinlichkeitsdichtefunktion $p(\text{Futterklasse} \mid \text{Gewicht})$ für alle 6 Futterklassen zusammen in einem Diagramm. Verwenden Sie dabei die Werte aus a).
- (3 Punkte) Klassifizieren Sie die Hühner in **chickwts_testing.csv** mit einem Bayes-Klassifikator anhand ihres Gewichtes in die 6 Futterklassen. Verwenden Sie dabei die Werte bzw. Wahrscheinlichkeitsdichtefunktionen aus a) und b). Geben Sie die Konfusionsmatrix und Klassifikationsgüte aus.