

# 1. Übungszettel *Mustererkennung* WS11/12

Prof. Raúl Rojas, Fabian Wiesel  
Institut für Informatik, Freie Universität Berlin

## 1. Aufgabe (5 Punkte)

Schreiben Sie für die Trainings- und Testdaten ein Programm, dass ein Datensatz einlesen und visuell ausgegeben kann (inklusive des entsprechenden Labels):

<http://www.inf.fu-berlin.de/lehre/WS11/ME/uebungen/uebung01/pendigits-training.txt>

<http://www.inf.fu-berlin.de/lehre/WS11/ME/uebungen/uebung01/pendigits-testing.txt>

Das Format wie folgt: Jede Zeile ist ein Datensatz für einen Linienzug einer Ziffer bestehend aus 17 Zahlen, die durch Leerzeichen getrennt sind. Die ersten 16 Zahlen sind 8 x-y Koordinatepaare, die letzte Zahl ist die Ziffer, die der Linienzug darstellen soll.

Die Datensätze sind auch auf der Veranstaltungsw Webseite verlinkt.

## 2. Aufgabe (10 Punkte)

Implementieren Sie den *k-nächste-Nachbar*-Klassifikator. Die Datei *pendigits-training.txt* dient Ihnen als Referenzdatenbank für die Methode. Ermitteln Sie die Erkennungsraten für  $k=1,2,3,\dots,7$  für den Datensatz in der Datei *pendigits-testing.txt*. Stellen Sie Erkennungsrate als x-y Diagramm mit einem beliebigen Programm dar.

## 3. Aufgabe (10 Punkte)

Generieren Sie für zwei normalverteilte Klassen jeweils gleichviele zufällige Repräsentanten.

$$\mu_1 = 1, \sigma_1 = 1$$

$$\mu_2 = 2, \sigma_2 = 2$$

Wie gut ist die Erkennungsrate mittels *k-Nearest-Neighbours* in Abhängigkeit von

- der Anzahl der Repräsentanten?
- der Anzahl der nächsten Nachbarn  $k$ ?

(2.5D-Diagramm)

x,y: Anzahl der Repräsentanten und Anzahl der Nachbarn

z: Erkennungsrate

**Abgabe:** Do. 03.11.11, 14:00 (vor Beginn des Tutoriums). Verspätete Abgaben werden nicht entgegengenommen