

## 6. Übung zur Veranstaltung *Künstliche Intelligenz*

Institut für Mathematik und Informatik  
Freie Universität Berlin, SoSe 2009  
Marco Block, Miao Wang, Prof.Dr.Raúl Rojas

Auf der Veranstaltungswebseite finden Sie ein Archiv mit Datensätzen für handschriftliche Ziffern. Die Datei *Readme.txt* beschreibt das verwendete Datenformat der beiden Dateien *digits.trn* und *digits.tst* mit Einträgen, die Grauwertbilder der Ziffern 0 bis 9 darstellen. Bitte beachten Sie, dass unter den Einträgen sich auch einige Datensätze befinden, denen keine Ziffern zugeordnet werden kann. Sie können entweder Java oder MatLab für die Programmierung nutzen.

### 1. Aufgabe (10 Punkte) kNN für Ziffernerkennung

Implementieren Sie das kNN-Verfahren und testen Sie mit der Trainingsmenge *digits.trn* und der Testmenge *digits.tst*. Geben Sie die Fehlklassifizierungen als Anzahl und als prozentuale Fehler-rate sowohl aufgeschlüsselt nach Zifferntypen als auch summarisch für alle Testdaten aus. Spielen Sie mit verschiedenen Werten für  $k$  rum.

### 2. Aufgabe (10 Punkte) kMeans für Ziffernerkennung

Implementieren Sie das kMeans-Verfahren und testen Sie mit der Trainingsmenge *digits.trn* und der Testmenge *digits.tst*. Geben Sie für jeden Cluster den Repräsentanten und die Anzahl an zugeordneten Datenvektoren aus. Geben Sie die Fehlklassifizierungen als Anzahl und als prozentuale Fehlerrate sowohl aufgeschlüsselt nach Zifferntypen als auch summarisch für alle Testdaten aus. Spielen Sie mit verschiedenen Werten für  $k$  rum.

### 3. Aufgabe (KI-Skills und ME-Skills) Gauß-Klassifikator für Ziffernerkennung

Implementieren Sie den Gauß-Klassifikator und testen Sie mit der Trainingsmenge *digits.trn* und der Testmenge *digits.tst*. Geben Sie die Fehlklassifizierungen als Anzahl und als prozentuale Fehlerrate sowohl aufgeschlüsselt nach Zifferntypen als auch summarisch für alle Testdaten aus.

**Denken Sie an die schriftliche und ggf. Email-Abgabe bis Freitag, den 05.06.2009 (12 Uhr s.t.). Verspätete Abgaben werden nicht mehr entgegengenommen. Denken Sie an ausführbare Quelldateien und Testläufe.**