

1. Übung zur Vorlesung

Mustererkennung

Institut für Mathematik und Informatik
Freie Universität Berlin, WiSe 2008/2009
Prof. Dr. Raúl Rojas, Marco Block, Dr. Ernesto Tapia

1. Aufgabe (14 Punkte)

Sei x_i eines der N Beispiele aus der gegebenen Ziffern-Offlinedatenbank (wird im Tutorium vorgestellt) und x_{ij} mit $j = 1, \dots, 192$ die gegebenen Merkmale. Berechnen Sie für alle Merkmale die Varianz s_j mit

$$s_j = \frac{1}{N} \sum_{i=1}^N (x_{ij} - \mu_j)^2 \quad (1)$$

und

$$\mu_j = \frac{1}{N} \sum_{i=1}^N x_{ij}$$

wobei

$$s_1 \geq s_2 \geq \dots \geq s_{192}. \quad (2)$$

Sei weiter $M(\alpha)$ die maximale Dimension, so dass für ein gegebenes α gilt

$$\frac{s_1 + s_2 + \dots + s_{M(\alpha)}}{s_1 + s_2 + \dots + s_{192}} \leq \alpha \quad (3)$$

- a) (4 Punkte) Wie interpretieren Sie die Anordnung (2) und die Formel (3)?
b) (10 Punkte) Berechnen Sie $M(1.0)$, $M(0.99)$, $M(0.95)$, $M(0.8)$, $M(0.75)$, $M(0.70)$ und $M(0.50)$. Für jede dieser Dimensionsreduktionen sollen Sie die Erkennungsraten für den in der Vorlesung vorgestellten k -nn Algorithmus ermitteln und auswerten, wobei beispielsweise für ein $M(\alpha)$ die Merkmale $\{s_1, \dots, s_{M(\alpha)}\}$ verwendet werden.

Abgabe: Freitag 31.10.2008, 14 Uhr