

Maschinelles Lernen und Data Mining
Sommersemester 2009
Übungsblatt 2

Besprechung des Übungsblattes am 19.5.2009

Aufgabe 2-1 Distanzmaße
schriftlich bearbeiten

Gegeben seien folgende Vektoren:

$$A = (2, 3)$$

$$B = (1, 0)$$

$$C = (x_1, \dots, x_{100}), x_i = \begin{cases} 1 & i = 1 \\ 0 & \text{sonst} \end{cases}$$

$$D = (x_1, \dots, x_{100}), x_i = \begin{cases} 2 & i = 2 \\ 0 & \text{sonst} \end{cases}$$

$$E = (x_1, \dots, x_{100}), x_i = \begin{cases} 1 & i \in [1, 10] \\ 0 & \text{sonst} \end{cases}$$

$$F = (x_1, \dots, x_{100}), x_i = \begin{cases} 2 & i \in [6, 15] \\ 0 & \text{sonst} \end{cases}$$

- a) Berechnen Sie die Abstände der Vektorenpaare A, B sowie C, D und E, F mit Hilfe der Distanzmaße $dist_{euklid}(x, y)$, $dist_{simple}(x, y)$, $dist_{simple00}(x, y)$, $dist_{cos}(x, y)$, $dist_{pearson}(x, y)$
- b) Welches Problem kann beim Pearson-Distanzmaß auftreten?

Aufgabe 2-2 Vergleich: Nächster Nachbar Schätzer und das Perceptron
schriftlich bearbeiten

Vergleichen Sie den Nächste Nachbar Schätzer mit dem Perceptron. Wie lassen sich die beiden Klassifikatoren visualisieren?

Aufgabe 2-3 Kernglätter

- a) Skizzieren Sie den Verlauf der des Kerndichteschätzers im Bereich $x = [-2,5; 2,5]$ mit $z = 0$:

Gauss Kern: $K_\lambda(z, x_i) = \exp\left(-\frac{|z-x_i|^2}{2\lambda^2}\right)$, $\lambda = 0,17$

Epanechnikov Kern:

$$K(z, x_i) = \begin{cases} 3/4 \cdot (1 - (z - x)^2) & , |z - x| < 0 \\ 0 & \text{sonst} \end{cases}$$

- b) Glätten Sie die gegebenen Punkte $(1;5)$, $(2;8)$, $(3;6)$, $(4;5)$, $(5;4)$, $(6;5,5)$ mit den oben angegebenen Kernen. Berechnen Sie dazu die Werte an den Stellen $x = \{0,5; 1; \dots; 6,5\}$ und erläutern sie kurz die Vor- und Nachteile der Kerne.

Hinweis: zur Lösung dürfen auch Computerprogramme verwendet werden. Im CIP steht z.B Maple zur Verfügung (Kommando: 'xmaple'). Abgebener Code muss lesbar und ausreichend dokumentiert sein.