

## Maschinelles Lernen und Data Mining

Sommersemester 2011

### Übungsblatt 4

Besprechung des Übungsblattes am 16./17.06.2011

#### Aufgabe 4-1 Interpretation der Standardabweichung

Skizzieren Sie den Graph der Standardnormalverteilung  $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$ ; mit  $\mu = 0$  und  $\sigma = 1$  im Intervall  $x \in [-4, 4]$ . Markieren und interpretieren Sie die Bereiche  $0 \pm \sigma$ ;  $0 \pm 2\sigma$ ;  $0 \pm 3\sigma$

#### Aufgabe 4-2 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:

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

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.

#### Aufgabe 4-3 Handschrifterkennung

Laden Sie den Datensatz von der Homepage und testen Sie die folgenden Classifier mittels WEKA. Für nominale Klassenlabels: pendigits.[tra—tes].arff. Für numerische Klassenlabels: pendigits.[tra—tes].regression.arff.

- Naive Bayes (nominale Labels erforderlich)
- LeastMedSq
- RBFNetwork mit 2, 10, 50 clustern
- LinearRegression

Diskutieren sie die Ergebnisse anhand der Fehlerraten und anhand des visuellen Fehlers.

Klassifizieren sie dazu zunächst den Datensatz, danach öffnen Sie per Rechtsklick auf das Modell und “visualize classifier Errors” das visuelle Modell.