Ludwig-Maximilians-Universität München 23.05.2008
Institut für Informatik
Dr. Volker Tresp
Franz Graf, Marisa Thoma

Maschinelles Lernen und Data Mining

Sommersemester 2008

Übungsblatt 5

Besprechung des Übungsblattes am 16.06.2008

Aufgabe 5-1 Basisfunktionen von Neuronalen Netzen

Wdh.: Die Ausgabe eines neuronalen Netzes für einen Testvektor $\mathbf{x}_i$ ist definiert durch

$$f(\mathbf{x}_i) = \sum_{h=0}^{M_\phi - 1} w_h \phi_h(\mathbf{x}_i, \mathbf{v}_h) \ . \qquad (1)$$

Die Gewichte der einzelnen Neuronen können über die Backpropagation-Regel mit musterbasiertem Gradientenabstieg gelernt werden.

In der Vorlesung wurden die neuronalen Netze mit sigmoiden Neuronen vorgestellt. Natürlich können auch andere Basisfunktionen verwendet werden.

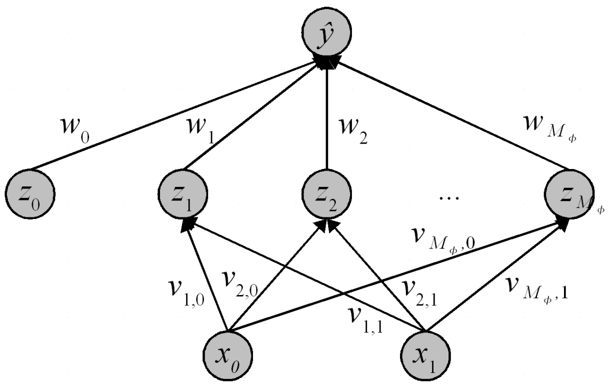
a) Welche Eigenschaften müssen diese Basisfunktionen erfüllen?

b) Ist eine Linearkombination $\phi(\mathbf{x}_i, \mathbf{v}_h) = z_h = \sum_{j=0}^{M} v_{h,j} x_{i,j}$ hierfür geeignet? Begründen Sie.

c) Ist die Anzahl der Parameter für $\phi(\mathbf{x}_i, \mathbf{v}_h)$ beschränkt? Können mehrere verschiedene Basisfunktionen für ein neuronales Netz verwendet werden?

Aufgabe 5-2 Ein einfaches neuronales Netz
schriftlich bearbeiten

Unten abgebildet sehen Sie ein zweischichtiges neuronales Netz mit einem Eingabeneuron $x \in \mathbb{R}$ und je einem Biasneuron $x_0 = z_0 = 1$ (d.h. $\mathbf{x}_i = (1, x_{i,1})^T$) in der Eingabeschicht und der versteckten Schicht.



b.w.

Als Aktivierungsfunktion der versteckten Neuronen verwenden wir einen Sigmoiden, also

$$z_h = \phi(\mathbf{x}_i,\mathbf{v}_h) = \frac{1}{1+\exp\left(-\sum_{j=0}^{M} v_{h,j}x_{i,j}\right)} ,$$

das Ausgangsneuron $\hat{y}$ wird wie üblich über eine Linearkombination gebildet.

a) Zeigen Sie, dass gilt: $\frac{\partial z_h}{\partial v_{h,j}} = x_{i,j} \cdot z_h \cdot (1 - z_h)$

b) Drücken Sie den maximalen Wert von $\hat{y}$ in Abhängigkeit von $\mathbf{w}$ aus, wenn alle Ausgangsgewichte $w_h (h \in \{0, \dots, M_\phi\})$ positiv sind. Was ist der minimale Wert?

c) Wie sieht $\hat{y}$ aus, wenn $v_{h,j} = 0$ für alle $j \in \{0, \dots, M\}, h \in \{1, \dots, M_\phi\}$? Welche Funktion erhalten Sie für $\hat{y}$, wenn alle $v_{h,j} = c, c \neq 0$?

Aufgabe 5-3 Wdh. Vektor Calculus
schriftlich bearbeiten

Eine Funktion $g(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$ über einen Vektor $\mathbf{x} \in \mathbb{R}^n$, die nach $\mathbf{x}$ abgeleitet wird ist:

$$\frac{\partial g(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}_1} \\ \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}_2} \\ \vdots \\ \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}_n} \end{bmatrix} .$$

Berechnen Sie $\frac{\partial \langle \mathbf{x}, \mathbf{x} \rangle}{\partial \mathbf{x}}$, wobei $\langle \mathbf{x}, \mathbf{x} \rangle$ das Standard-Skalarprodukt von $\mathbf{x}$ mit sich selbst ist.

Aufgabe 5-4 Kernkombinationen
schriftlich bearbeiten

Um einen selbstdefinierten Kern $k(\mathbf{x}_i, \mathbf{x}_j)$ für $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^n$ anzuwenden, muss für gewöhnlich gezeigt werden, dass es sich auch tatsächlich um einen legitimen Kern handelt. Da es recht aufwendig sein kann zu zeigen, dass für k das *Mercer Theorem* zutrifft, wird häufig explizit das Mapping der impliziten Basistransformationen angegeben: $k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$.

Eine weitere beliebte Variante die Gültigkeit einer Kernfunktion zu zeigen ist die Rückführung auf eine Kombination aus Kernels, da für einige Operationen $\circ$ gilt, dass $k(\mathbf{x}_i, \mathbf{x}_j) = k_1(\mathbf{x}_i, \mathbf{x}_j) \circ k_2(\mathbf{x}_i, \mathbf{x}_j)$ ein legitimer Kernel ist.

Zeigen Sie dass für valide Kernelfunktionen $k_l(\mathbf{x}_i,\mathbf{x}_j) (l \in 1, \dots ,n)$ gilt:

a) **Skalierung**: Für $a > 0$ ist $k(\mathbf{x}_i,\mathbf{x}_j) := a\, k_1(\mathbf{x}_i,\mathbf{x}_j)$ ein Kernel.

b) **Summe**: $k(\mathbf{x}_i,\mathbf{x}_j) := k_1(\mathbf{x}_i,\mathbf{x}_j) + k_2(\mathbf{x}_i,\mathbf{x}_j)$ ist ein Kernel.

c) **Linearkombination**: Für $w \in \mathbb{R}_+^n$ ist $k(\mathbf{x}_i,\mathbf{x}_j) := \sum_{l=0}^{n-1} w_l\, k_l(\mathbf{x}_i,\mathbf{x}_j)$ ein Kernel.

d) **Produkt**: $k(\mathbf{x}_i,\mathbf{x}_j) := k_1(\mathbf{x}_i,\mathbf{x}_j) \cdot k_2(\mathbf{x}_i,\mathbf{x}_j)$ ist ein Kernel.

e) **Potenz**: Für ein $p \in \mathbb{N}_+ : k(\mathbf{x}_i,\mathbf{x}_j) := k_1(\mathbf{x}_i,\mathbf{x}_j)^p$ ist ein Kernel.