Übung Künstliche Neuronale Netzwerke
Wintersemester 2015/2016

① Abgabe am 09. 12. 2015 bis Ende der Übung handschriftlich auf Papier (1x pro Person)
② Abgabe bis 16. 12. 2015, 23:55 Uhr im Moodle-Kurs (1x pro Gruppe)

3. Übungsblatt vom 02. 12. 2015

3.1 Lineare Diskriminanten

Gegeben sei die folgende lineare Diskriminante:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} -2 & -5 & 3 \\ 2 & 0 & 1 \\ 2 & -2 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix}$$

a) Zeichnen Sie das zugehörige Netzwerkdiagramm. Beschriften Sie die Kanten mit den Werten der entsprechenden synaptischen Gewichten. 2P ①

b) Zeichnen Sie die Entscheidungsregionen R_1, R_2 und R_3 in ein Koordinatensystem. Dabei ist R_1 die Region, in der y_1 maximal ist usw. 5P ①

Berechnen Sie dazu zunächst die Geraden, auf denen zwei der y_i gleich sind und zeichnen Sie diese ein. Dadurch wird die Ebene in 6 Teile $T_1 \dots T_6$ eingeteilt. Entscheiden Sie für jedes T_j, welches y_i maximal ist. Durch Zusammenfassen der entsprechenden T_j erhalten Sie R_1, R_2 und R_3.

3.2 Gradientenabstieg

Betrachten Sie folgende Fehlerfunktion:

$$E(\underline{w}) = 1 - e^{-\frac{w_1^2+w_2^2+w_3^2}{\beta}} - \alpha(\cos(w_1)\cos(w_2)\cos(w_3) - 1)$$

a) Berechnen Sie den Gradienten $\nabla E(\underline{w})$. Für welche Beziehung von α und β gibt es ein lokales Minimum bei $\underline{w}_{krit} = \underline{0}$? 5P ①

Hinweis: Ein hinreichendes Kriterium für ein lokales Minimum ist, wenn der Gradient $\nabla E(\underline{w}_{krit}) = \underline{0}$ und die Hesse-Matrix der zweiten Ableitungen $H\big(E(\underline{w}_{krit})\big)$ positiv definit ist, d.h. nur positive Eigenwerte besitzt.

$$H\big(E(\underline{w})\big) = \begin{pmatrix} \frac{\partial^2 E}{\partial w_1^2} & \frac{\partial^2 E}{\partial w_1 \partial w_2} & \frac{\partial^2 E}{\partial w_1 \partial w_3} \\ \frac{\partial^2 E}{\partial w_2 \partial w_1} & \frac{\partial^2 E}{\partial w_2^2} & \frac{\partial^2 E}{\partial w_2 \partial w_3} \\ \frac{\partial^2 E}{\partial w_3 \partial w_1} & \frac{\partial^2 E}{\partial w_3 \partial w_2} & \frac{\partial^2 E}{\partial w_3^2} \end{pmatrix}$$

b) Schreiben Sie jeweils eine Funktion, die den Fehler E und den Gradienten ∇E für einen gegebenen Gewichtsvektor $\underline{w}$ und ein gegebenes α berechnet. 2P ②

Der Gradientenabstieg erfolgt durch die Änderung des Gewichtsvektors $\underline{w}$. Die Änderung $\Delta\underline{w}$ wird berechnet mit:

$$\Delta\underline{w} = -\eta \cdot \nabla E(\underline{w})$$

c) Schreiben Sie eine Funktion, die durch einen Gradientenabstieg für eine gegebene Initialgewichtung $\underline{w}_{init}$ und gegebene Lernrate η ein Minimum der Fehlerfunktion $E(\underline{w})$ sucht und zurückgibt. Definieren Sie ein geeignetes Abbruchkriterium für den Gradientenabstieg. 4P ②

d) Wählen Sie $\alpha = 0.5$ und $\beta = 500.0$. Starten Sie den Gradientenabstieg mehrfach mit zufälligen Werten für $\underline{w}_{init}$ aus dem Bereich $[-10, 10]^3$ und betrachten Sie die resultierenden Gewichtsvektoren $\underline{w}$ und deren Fehler $E(\underline{w})$. Variieren Sie die Lernrate η im Bereich $[0.01, 10.0]$. Für welche Werte von η divergiert das Verfahren? Was beobachten Sie in Bezug auf die Laufzeit? 3P ②

e) Was geschieht für $\alpha = -0.05$, $\beta = 500.0$ und $\eta = 0.01$ bei $\underline{w}_{init} = \underline{0}$ und was bei $\underline{w}_{init} = (1, 1, 0)^{\mathrm{T}}$? Wird ein Minimum erreicht, und wenn nicht, warum? 1P ②