Neuronale Netze Sommer 2019

Prof. Dr.-Ing. Sebastian Stober, Michelle I. Bieber, Jan-Ole Perschewski, Christian Wustrau

7. Übungsblatt

Aufgabe 28 RNN

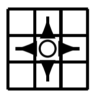
Betrachten Sie ein RNN mit einer versteckten Schicht, die wie folgt aktiviert wird :

$$h_{t+1} = Wx_t + Uh_t + b$$

$$W, U \in \mathbb{R}^{4\times 4}, b \in \mathbb{R}^{4\times 1}$$

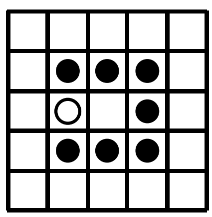
In den folgenden Skizzen betrachten wir den Agenten $\circ$, der sich entlang des Pfades von $\bullet$ bewegen soll. Die Eingabe x_t in das RNN ist die Umgebung des Agenten, wobei sich der Vektor wie folgt zusammensetzt:

$$x_t = \begin{bmatrix} \uparrow \\ \rightarrow \\ \leftarrow \\ \downarrow \end{bmatrix} \in \{0,1\}^4$$



Hierbei soll ein Eintrag 1 sein, wenn sich in der jeweiligen Richtung ein $\bullet$ befindet. Die Ausgabe h_t soll dabei analog zu x_t die nächste Richtung des Agenten angeben. In der folgenden Anwendung wird nach dieser Ausgabe der Agent in die Richtung bewegt und beeinflusst die Eingabe x_{t+1}

Bestimmen sie für das Labyrinth unten mit $h_0 = (1,0,0,0)^T$ die Gewichtsmatrizen W, U, b, sodass der Agent dem Pfad aus $\bullet$ folgt.



Könnte ein normales MLP dieses Problem lösen? Was können wir über die Berechnungsfähigkeit der RNNs ableiten?

Aufgabe 29 Architekturen

Schlagen Sie für die folgenden Anwendungsgebiete jeweils eine Architektur vor und erklären Sie warum die jeweilige Architektur geeignet ist.

- Sie haben eine Sammlung von den Werken von Shakespear und wollen Texte im gleichen Stil generieren. Dabei sollen Sie auf der Ebene von Buchstaben arbeiten.
- Sie haben eine Sammlung von Katzen- und Hundebilder und wollen bei neuen Bildern entscheiden, ob es sich um einen Hund oder eine Katze handelt.
- Sie sollen ein Model entwickeln, das markiert, ob ein Buchstabe in einem Text zur wörtlichen Rede gehört oder nicht.
- Sie sollen ein Model entwickeln, das Englisch in Russisch und Deutsch übersetzt.
- Sie sollen ein Model entwickeln, das in der Lage ist einen Code zu zu reduzieren und den ursprünglichen Code zu rekonstruieren.

Aufgabe 30 Gradienten in RNNs

Eine besondere Herausforderung bei RNNs ist das bestimmen eines geeigneten Gradientens. Dabei ist ein Aspekt das Lernen von Abängigkeiten zwischen Elementen am Anfang und am Ende der Sequenz z.B. Subjekt und Verb in einem Nebensatz.
Wir betrachten nun als Vereinfachung nur die Verbindungen zwischen versteckten Neuronen mit linearer Aktivierungsfunktion.

$$h_t = U^t \cdot h$$

Weiterhin sei $f(h_k)$ der Fehler, den die versteckte Schicht macht. Bestimmen Sie den Gradienten der Fehlerfunktion für U in Abhängigkeit von $k > 1$.

- Wie verhält sich der Gradient, wenn $U = U_1$ bzw. U_2 ist, für sehr große t?

$$U_1 = \begin{bmatrix} 1.2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1.1 \end{bmatrix}, U_2 = \begin{bmatrix} 0.8 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.9 \end{bmatrix}$$

- Warum sind diese Werte problematisch für Gradienten-basiertes Training?
- Was sind mögliche Strategien, die bei den Problemen helfen können ?

Aufgabe 31 Anfangswertproblem

Für ein vereinfachtes Modell eines Fahrzeugs, dass beschleunigt wird, kann man mit Hilfe des Impulserhaltungssatzes die Geschwindigkeit des Fahrzeugs bestimmen.
Das Modell ist definiert als folgende Differentialgleichung (Anfangswertproblem):
$\frac{dv}{dt} = \dot{v}(t) = \frac{1}{m}(F_A - F_R) = \frac{1}{m}(F_A(t) - d \cdot v(t))$
Dabei ist m die Masse unseres Fahrzeugs , F_A die Antriebskraft, die auf das Fahrzeug wirkt, F_R der Rollwiderstand,der der Antriebskraft entgegenwirkt, d der Rollwiderstandskoeffizient und $v(t)$ die zu bestimmende Geschwindigkeit des Fahrzeugs abhängig von der Zeit.

Dieses Problem lässt sich numerisch mit dem Euler-Cauchy-Verfahren durch den Polygonzug $v(t_i) = v(t_{i-1}) + \dot{v}(t_{i-1})\Delta t$ lösen.
Bei gegebenem Anfangswert $v(t_0)$ können wir jeden Wert $v(t)$ näherungsweise berechen, indem wir das Intervall $[v(t_0), v(t)]$ in n Teile gleicher Länge $\Delta t = \frac{t-t_0}{n}$ zerlegen und dann schrittweise $\vartheta(t_1)$ bis $\vartheta(t)$ berechnen.

a) Erstellen Sie mit Hilfe der Formel ein sequence to fixed-size vector Netz, welches schrittweise den Polygonzug $v(t_i)$ berechnet und dann $v(t)$ ausgibt.
Hinweis: Versuchen Sie die Formel für den Polygonzug und die Differrentialgleichung in eine Gleichung für die versteckte Schicht in der Form $h_t = W \cdot x + U \cdot h_{t-1}$ zu bringen.

b) Berechnen Sie für $v(t_0) = 0, m = 150, d = 10, F_A(t) = 100 \cdot t$ und $\Delta t = 0,5$ näherungsweise $v(t)$ mit $t = 8$. Vergleichen Sie ihre Ergebnisse mit der analytischen Lösung des Problems:
$v(t) = \frac{1}{d} \cdot F_A(t) \cdot (1 - e^{-\frac{d}{m}t})$
Was stellen Sie fest ? Wie kann man die Approximation verbesseren ?

c) Bei der Erstellung des rekurrenten Netzes haben Sie die Gewichte festgelegt und damit schrittweise den nächsten Näherungspunkt berechnet. Wie würde der Prozess der Erstellung eines Netzes aussehen, dass keine numerische Lösungsformel vorgeben hat ?
Welche Daten müssten dafür vorliegen ?