hw 8

Chen Yuanteng

3039725444

1. Backprop through a Simple RNN

(a). $p = u_1, \quad q = W \cdot u_1, \quad r = u_2 + q = u_2 + W \cdot u_1$

$\quad\quad S = W \cdot r = W \cdot u_2 + W^2 \cdot u_1$

$$\begin{cases} t = S + u_3 = u_3 + W \cdot u_2 + W^2 \cdot u_1 \\ y = W \cdot t = W \cdot u_3 + W^2 \cdot u_2 + W^3 u_1 \end{cases}$$

(b). $\dfrac{dy}{dW} = u_3 + 2 \cdot W \cdot u_2 + 3 W^2 \cdot u_1$

(c) $\dfrac{\partial y}{\partial t} = W; \dfrac{\partial y}{\partial s} = W; \dfrac{\partial y}{\partial r} = W^2; \dfrac{\partial y}{\partial q} = \dfrac{\partial y}{\partial r} = W^2$

$\quad\dfrac{\partial y}{\partial p} = \dfrac{\partial y}{\partial q} \dfrac{\partial q}{\partial u_1} \cdot \dfrac{\partial u_1}{\partial p} = W^2 \cdot W \cdot 1 = W^3$

(d).

$\quad\dfrac{dy}{dW} = t \dfrac{\partial y}{\partial y} + r \cdot \dfrac{\partial y}{\partial s} + p \dfrac{\partial y}{\partial q}$

$\quad = u_3 + W \cdot u_2 + W^2 u_1 + (u_2 + W \cdot u_1) \cdot W + u_1 \cdot W^2$

$\quad = 3 W^2 u_1 + 2 W u_2 + u_3$

5. self-supervised Linear Autoencoders

(a). (iv) 2 layers (1 for encoder and
$\quad\quad\quad\quad\quad\quad\quad$ 1 for decoder)

(iii) use. nn. MSELoss as we hope for each vector can be close to its reconstruction.

(iii)
Weight-Decay + SGD-optimizer

(b). not sure

7. Homework Process and Study Group
(a). GPT. CSDN
(b) None
(c) two days