

习题 6-1

1. 延时神经网络在时间维度上共享权值, 以降低参数数量. 因此对于序列输入来讲, 延时神经网络就相当于卷积神经网络.
2. 循环神经网络会保留/提取历史输入的所有信息 (有损), 同时能处理不定长的输入数据; 卷积神经网络中每个卷积核的输出只能保留/提取被卷积核处理的部分的输入信息, 且输入数据的长度是限定好的.

习题 6-2

we have:

$$\begin{aligned} z_t &= U h_{t-1} + W x_t + b \\ h_t &= f(z_t) \\ \delta_{t,k} &= \frac{\partial \mathcal{L}_t}{\partial z_k} \end{aligned}$$

for $\frac{\partial \mathcal{L}}{\partial W}$:

$$\begin{aligned} \frac{\partial \mathcal{L}_t}{\partial W} &= \frac{\partial z_t}{\partial W} \frac{\partial h_t}{\partial z_t} \frac{\partial \mathcal{L}_t}{\partial h_t} \\ &= \left(\frac{\partial h_{t-1}}{\partial W} \frac{\partial z_t}{\partial h_{t-1}} + x_t^T \right) \frac{\partial h_t}{\partial z_t} \frac{\partial \mathcal{L}_t}{\partial h_t} \\ &= \left(\frac{\partial z_{t-1}}{\partial W} \frac{\partial h_{t-1}}{\partial z_{t-1}} \frac{\partial z_t}{\partial h_{t-1}} + x_t^T \right) \frac{\partial h_t}{\partial z_t} \frac{\partial \mathcal{L}_t}{\partial z_t} \\ &= \frac{\partial z_{t-1}}{\partial W} \frac{\partial \mathcal{L}_t}{\partial z_{t-1}} + x_t^T \frac{\partial \mathcal{L}_t}{\partial z_t} \\ &= \frac{\partial z_{t-1}}{\partial W} \delta_{t,t-1} + x_t^T \delta_{t,t} \\ &= \sum_{k=1}^t x_k^T \delta_{t,k} \\ \frac{\partial \mathcal{L}}{\partial W} &= \sum_{t=1}^T \sum_{k=1}^t x_k^T \delta_{t,k} \end{aligned}$$

for $\frac{\partial \mathcal{L}}{\partial b}$:

$$\begin{aligned}
\frac{\partial \mathcal{L}_t}{\partial b} &= \frac{\partial z_t}{\partial b} \frac{\partial h_t}{\partial z_t} \frac{\partial \mathcal{L}_t}{\partial h_t} \\
&= \left(\frac{\partial z_{t-1}}{\partial b} \frac{\partial h_{t-1}}{\partial z_{t-1}} \frac{\partial z_t}{\partial h_{t-1}} + I \right) \frac{\partial h_t}{\partial z_t} \frac{\partial \mathcal{L}_t}{\partial h_t} \\
&= \frac{\partial z_{t-1}}{\partial b} \delta_{t,t-1} + \delta_{t,t} \\
&= \sum_{k=1}^t \delta_{t,k} \\
\frac{\partial \mathcal{L}}{\partial b} &= \sum_{t=1}^T \sum_{k=1}^t \delta_{t,k}
\end{aligned}$$