

RNN (Recurrent Neural Network)

• 재귀 / 순환 신경망

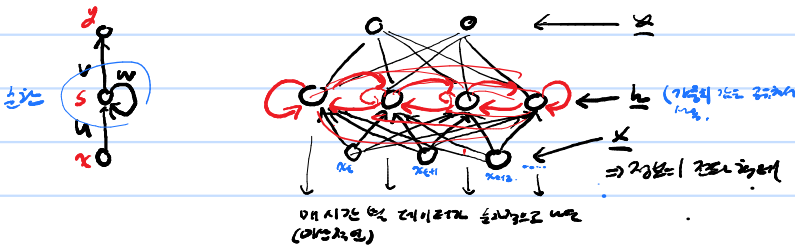
↳ 2차원/3차원 데이터는 주로

- 음성, 자연어 문장, 동영상. 푸기 번호 등 (순서가 있음, 시간이 흐름) 이 컴퓨터에 이해하는 데어려움

- 구성요소가 순차적으로 발생하거나, 지속적으로 순차에 있는 종류의 Data

↳ sequence

- 현재 시점의 값과 이전/미래 시점의 값이 서로 연관되어 있음 * CNN: 이미지 기 2차원 데이터는 정렬하여 연관성 있는 각해로운 분류
- 과거의 정보를 반영하기 위해, 순환/회환 같은 개념으로 사용함.



RNN 파라미터

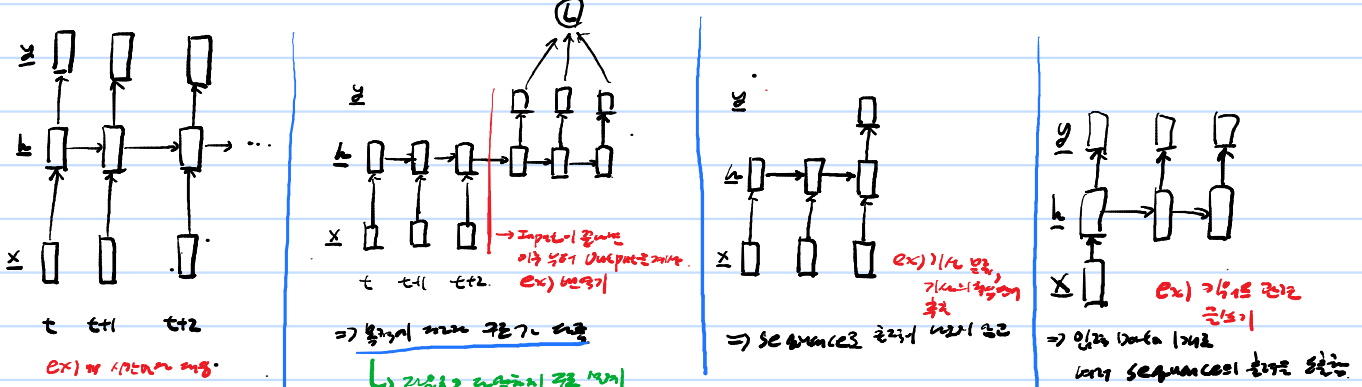
- 입력(세리터) 크기가 32원, h가 22원, h가 44의 노드를 갖는 경우,

- 입력-은닉층 간 가중치 $U: 4 \times 3$
- 은닉-은닉층 간 가중치 $V: 2 \times 4$
- 은닉-은출 간 가중치 $W: 4 \times 4$

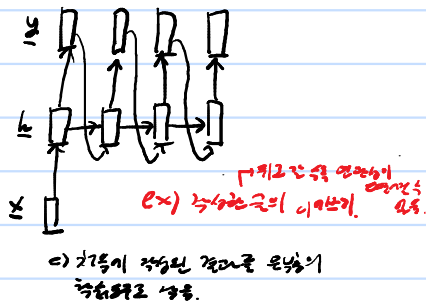
- RNN에서 현재시점(t)의 은닉층은 h는 과거의 모든 입력과 은닉층값이 영향을 받음

$$\begin{aligned} s_t &= Ux_t + Wh_{t-1} + b_s & \bullet b_s: \text{bias} \\ h_t &= f(s_t) \Rightarrow \text{RNN에 활용되도록 변환} & \bullet f: \text{시도하여 영향력 있는} \\ z_t &= V \cdot h_t + b_z \\ y_t &= g(z_t) \end{aligned}$$

RNN 출력 행동 형태



⇒ 입력 레이어가 1개이고 출력 레이어가 1개는 순환 방식으로 동작함.

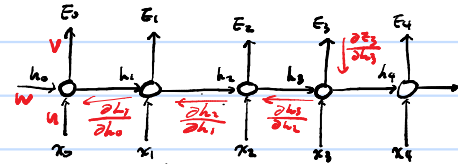


BPTT (Back Propagation Through Time)

- RNN에서 오차역전파와 순도리듬을 적용하기 위해서는 시도를 하고 문제해결

가중치 계승적에 한함

- 각 시점에서의 그래디언트를 각각 위해 편향점을 해당 매개변수의 그래디언트로 반영함



$$\begin{aligned} t=3, \quad \frac{\partial h_3}{\partial w} &= \frac{\partial h_3}{\partial z_3} \cdot \frac{\partial z_3}{\partial h_3} \cdot \frac{\partial h_3}{\partial w} \\ &\Rightarrow \frac{\partial h_3}{\partial z_3} \cdot \frac{\partial z_3}{\partial h_3} \left(\frac{\partial h_3}{\partial h_3} \cdot \frac{\partial h_3}{\partial w} + \frac{\partial h_3}{\partial h_2} \cdot \frac{\partial h_2}{\partial w} + \frac{\partial h_3}{\partial h_1} \cdot \frac{\partial h_1}{\partial w} + \frac{\partial h_3}{\partial h_0} \cdot \frac{\partial h_0}{\partial w} \right) \end{aligned}$$

$$\frac{\partial h_3}{\partial h_2} \cdot \frac{\partial h_2}{\partial w} = \frac{\partial h_3}{\partial h_2} \left(\frac{\partial h_2}{\partial h_2} \cdot \frac{\partial h_2}{\partial w} + \frac{\partial h_2}{\partial h_1} \cdot \frac{\partial h_1}{\partial w} + \frac{\partial h_2}{\partial h_0} \cdot \frac{\partial h_0}{\partial w} \right)$$

↳ 앞의 시간에서부터 곱해줄 층이 많음

⇒ 2차 시점으로 계속해가며 파라 변화하므로 곱해줄 층이 계속 늘림

$$\frac{\partial h_3}{\partial w} = \sum_{k=0}^3 \frac{\partial h_3}{\partial h_k} \cdot \frac{\partial h_k}{\partial h_{k-1}} \left(\frac{\partial h_k}{\partial h_{k-1}} \cdot \frac{\partial h_{k-1}}{\partial w} \right) \cdot \frac{\partial h_k}{\partial w}$$

$$\Rightarrow h_j = f(z_j) = f(Ux_j + Wh \cdot h_{j-1} + b_s) \text{ 이므로,}$$

$$\frac{\partial h_j}{\partial h_{j-1}} = W^T \cdot \text{diag}(f'(z_j))$$

→ 시간을 계속해가며 파라 가중치 행렬 W^T 가 계속재곱됨

→ 가중치 문제가 발생함. 가중치재곱으로 인해 너무 큰 값이나 너무 작은 값이 되어 가중치 업데이트에 문제가 생겨서 학습이 제대로 안될 수 있음

가중치 소멸 (Vanishing Gradient)

- 크기가 1보다 작은 가중치의 경우, 가중치재곱에 따라 가중치가 0이 됨.
- 가중치의 폭장/성장은 반대방향으로 발생함, 반대방향이 아니므로 가중치재곱이 발생함. 가중치재곱이 발생함. 가중치재곱이 발생함.

가중치 폭장 (Exploding Gradient)

- 크기가 1보다 큰 가중치이 경우, 가중치재곱에 따라 가중치가 계속적으로 폭장함
- ⇒ 시간이경과로 변화 된다면 레이어의 경우 가중치 문제를 유발하게 됨.

↳ W^T 의 가중치재곱이 수렴해 문제가 발생함.

가중치 폭장 문제의 대응방법

- RMSprop: 그래디언트 업데이트 시 평균값으로 나눠서 업데이트
- Truncated BPTT: BPTT를 모든 시점의 레이어에 대해 하는 대신, 최대 몇 층까지만 적용. 가중치재곱이 발생함.
- Gradient clipping: 그래디언트의 크기를 임계 (Threshold) 이하로 제한