과제 2

바자르와니주치 - 202046869

요 약

1. CLASSICAL RNN

1.1 문제 1

1.1.1 간략한 서술

$$\delta_z = \frac{\partial L}{\partial h}\frac{\partial h}{\partial z} \tag{1}$$

$$\delta_x = \frac{\partial L}{\partial h}\frac{\partial h}{\partial z}\frac{\partial z}{\partial x} \tag{2}$$

1.1.2 유도 과정

$$\delta_z = \frac{\partial L}{\partial z} => \frac{\partial L}{\partial h}\frac{\partial f(z)}{\partial z} \tag{3}$$

$$\delta_x = \frac{\partial L}{\partial x} => \frac{\partial L}{\partial h}\frac{\partial f(z)}{\partial x} => \frac{\partial L}{\partial h}\frac{\partial f(z)}{\partial z}\frac{\partial Ux}{\partial x} \tag{4}$$

1.2 문제 2

1.2.1 간략한 서술

$$\delta_U = \frac{\partial L}{\partial h}\frac{\partial h}{\partial z}\frac{\partial z}{\partial U} \tag{5}$$

1.2.2 유도 과정

$$\delta_x = \frac{\partial L}{\partial U} => \frac{\partial L}{\partial h}\frac{\partial f(z)}{\partial U} => \frac{\partial L}{\partial h}\frac{\partial f(z)}{\partial z}\frac{\partial Ux}{\partial U} \tag{6}$$

1.3 문제 3

1.3.1 유도 과정

$$\delta_{\tilde{o}} = \frac{\partial L}{\partial \tilde{o}} => \frac{\partial L}{\partial o}\frac{\partial o}{\partial \tilde{o}} \tag{7}$$

$$\frac{\partial L}{\partial o} = \frac{\partial \sum_i^T -y_i \cdot log(o)}{\partial o} = \frac{\partial \sum_i^T -y_i \cdot log(o)}{\partial o} = \frac{\sum_i^T -y_i}{o} \tag{8}$$

$$\frac{\partial o}{\partial \tilde{o}} = \frac{\partial softmax(\tilde{o})}{\partial \tilde{o}} = o(1-\tilde{o}) \tag{9}$$

$$\frac{\sum_i^T -y_i}{o} o(1-\tilde{o}) = \sum_i^T -y_i(1-\tilde{o}) = \sum_i^T y_i\tilde{o} - y = 1\tilde{o} - y = \tilde{o} - y \tag{10}$$

1.4 문제 4

1.4.1 유도 과정

$$\frac{\partial L}{\partial V} = \frac{\partial \sum_i^T -y_i \cdot log(o)}{\partial o}\frac{\partial softmax(Vh^t)}{\partial V} \tag{11}$$

$$\frac{\partial L}{\partial W} = \frac{\partial \sum_i^T -y_i \cdot log(o)}{\partial o}\frac{\partial softmax(Vh^t)}{\partial h^t}\frac{\partial f(z^t)}{\partial z^t}\frac{\partial Ux^t + Wh^{t-1}}{\partial W} \tag{12}$$

$$\frac{\partial L}{\partial U} = \frac{\partial \sum_i^T -y_i \cdot log(o)}{\partial o}\frac{\partial softmax(Vh^t)}{\partial h^t}\frac{\partial f(z^t)}{\partial z^t}\frac{\partial Ux^t + Wh^{t-1}}{\partial U} \tag{13}$$

$$\frac{\partial L}{\partial E} = \frac{\partial \sum_i^T -y_i \cdot log(o)}{\partial o}\frac{\partial softmax(Vh^t)}{\partial h^t}\frac{\partial f(z^t)}{\partial z^t}\frac{\partial Ux^t + Wh^{t-1}}{\partial x^t}\frac{\partial E \cdot onehot(w)}{\partial E} \tag{14}$$

1.5 문제 6

Classical RNN은 하나의 W를 모든 steps에서 공유한다. 그렇기에 매 batch에서 W가 중첩적으로 학습되면서 1보다 적어질경우 vanishing gradient가 값이 너무 높아지면 exploding gradients 문제가 생긴다.

2. LSTM

2.1 문제 1

$$\frac{\partial L}{\partial V} = \frac{\partial \sum_i^T -y_i \cdot log(o)}{\partial o}\frac{\partial softmax(Vh^t)}{\partial V} \tag{15}$$

$$\frac{\partial L}{\partial W^c} = \frac{\partial \sum_i^T -y_i \cdot log(o)}{\partial o} \frac{\partial softmax(Vh^t)}{\partial h^t} \frac{\partial r^t \odot c^t}{\partial c^t} \frac{\partial f^t \odot c^{t-1} + i^t \odot \tilde{c}^t}{\partial \tilde{c}^t} \frac{\partial f(k)}{\partial k} \frac{\partial [W^c U^c][h^{t-1} x^t]^T}{\partial W^c} \tag{16}$$

$$\frac{\partial L}{\partial U^c} = \frac{\partial \sum_i^T -y_i \cdot log(o)}{\partial o} \frac{\partial softmax(Vh^t)}{\partial h^t} \frac{\partial r^t \odot c^t}{\partial c^t} \frac{\partial f^t \odot c^{t-1} + i^t \odot \tilde{c}^t}{\partial \tilde{c}^t} \frac{\partial f(k)}{\partial k} \frac{\partial [W^c U^c][h^{t-1} x^t]^T}{\partial U^c} \tag{17}$$

$$\frac{\partial L}{\partial W^r} = \frac{\partial \sum_i^T -y_i \cdot log(o)}{\partial o} \frac{\partial softmax(Vh^t)}{\partial h^t} \frac{\partial r^t \odot c^t}{\partial r^t} = \frac{\partial [W^r U^r][h^{t-1} x^r]^T}{\partial W^r} \tag{18}$$

$$\frac{\partial L}{\partial U^r} = \frac{\partial \sum_i^T -y_i \cdot log(o)}{\partial o} \frac{\partial softmax(Vh^t)}{\partial h^t} \frac{\partial r^t \odot c^t}{\partial r^t} = \frac{\partial [W^r U^r][h^{t-1} x^r]^T}{\partial U^r} \tag{19}$$

$$\frac{\partial L}{\partial W^f} = \frac{\partial \sum_i^T -y_i \cdot log(o)}{\partial o} \frac{\partial softmax(Vh^t)}{\partial h^t} \frac{\partial r^t \odot c^t}{\partial c^t} \frac{\partial f^t \odot c^{t-1} + i^t \odot \tilde{c}^t}{\partial f^t} \frac{\partial \sigma(k)}{\partial k} \frac{\partial [W^f U^f][h^{t-1} x^t]^T}{\partial W^f} \tag{20}$$

$$\frac{\partial L}{\partial U^f} = \frac{\partial \sum_i^T -y_i \cdot log(o)}{\partial o} \frac{\partial softmax(Vh^t)}{\partial h^t} \frac{\partial r^t \odot c^t}{\partial c^t} \frac{\partial f^t \odot c^{t-1} + i^t \odot \tilde{c}^t}{\partial f^t} \frac{\partial \sigma(k)}{\partial k} \frac{\partial [W^f U^f][h^{t-1} x^t]^T}{\partial U^f} \tag{21}$$

$$\frac{\partial L}{\partial W^i} = \frac{\partial \sum_i^T -y_i \cdot log(o)}{\partial o} \frac{\partial softmax(Vh^t)}{\partial h^t} \frac{\partial r^t \odot c^t}{\partial c^t} \frac{\partial f^t \odot c^{t-1} + i^t \odot \tilde{c}^t}{\partial i^t} \frac{\partial \sigma(k)}{\partial k} \frac{\partial [W^i U^i][h^{t-1} x^t]^T}{\partial W^i} \tag{22}$$

$$\frac{\partial L}{\partial U^i} = \frac{\partial \sum_i^T -y_i \cdot log(o)}{\partial o} \frac{\partial softmax(Vh^t)}{\partial h^t} \frac{\partial r^t \odot c^t}{\partial c^t} \frac{\partial f^t \odot c^{t-1} + i^t \odot \tilde{c}^t}{\partial i^t} \frac{\partial \sigma(k)}{\partial k} \frac{\partial [W^i U^i][h^{t-1} x^t]^T}{\partial U^i} \tag{23}$$

$$\frac{\partial L}{\partial E} = \frac{\partial \sum_i^T -y_i \cdot log(o)}{\partial o} \frac{\partial softmax(Vh^t)}{\partial h^t} \frac{\partial r^t \odot c^t}{\partial c^t} \frac{\partial f^t \odot c^{t-1} + i^t \odot \tilde{c}^t}{\partial i^t} \tag{24}$$

3. 구현

코드 실행 : 학습 python [코드이름] -train -gpu
테스트 python [코드이름] -test -gpu
공통 : epochs = 7, hidden dimension = 200, lr = 0.1, batch = 16

표 1: 모델별 성능

MODEL	$tagged - test$
RNN	0.0007665
LSTM	0.004280
LSTM+Skipgram	0.09
LSTM+pretrain	0.00447630

표 2: 1 epoch에서의 gpu vs cpu

MODEL	$min - per - epochs$
cpu	측정불가(너무 오래걸려서 측정이 힘듬)
gpu	$26 \sim 30$

4. 향후 개선점

학습이 된다는 걸 확인했기에 오래 돌려보기, 추가로 구현코드의 최적화가 필요