▼ *3주차 예습과제*

1.Named Entity Recognition (NER)

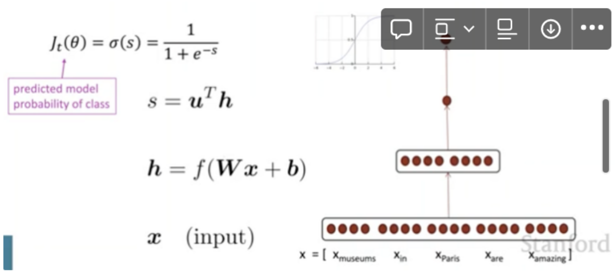
-텍스트 속 이름을 찾고 분류

-구조화 되지 않은 텍스트에서 언급된 단어를 이미 정의된 범주(사람, 나라 등)으로 분류

*Simple NER

-인접한 단어의 문맥에서 각 단어를 분류

-logistic classifier를 학습시킨다



2.Stochastic Gradient Descent

$$\theta^{new} = \theta^{old} - \alpha \nabla_\theta J(\theta)$$

α = step size or learning rate

기울기 함수를 계산하기 위해서

1) by hand

-행렬 계산: 다변수 미적분을 행렬을 사용하여 계산

-gradient를 벡터화시켜 유용한 계산 가능.

2) algorithmically : the backpropagation algorithm

3.Gradient

출력과 입력이 1개씩인 함수가 주어졌을 때, 그 함수에서 기울기는 미분을 통해 파생된다.

출력이 m개 입력이 n개인 함수가 주어졌을 때, 그 함수의 미분값은 편미분들의 m x n 행렬이다.

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \qquad \boxed{\left(\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\right)_{ij} = \frac{\partial f_i}{\partial x_j}}$$

4.Chain Rule

- 단일 변수일 경우: 도함수 곱

$$z = 3y$$
$$y = x^2$$
$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx} = (3)(2x) = 6x$$

- 다 변수일 경우: Jacobians 곱

$$\boldsymbol{h} = f(\boldsymbol{z})$$
$$\boldsymbol{z} = \boldsymbol{W}\boldsymbol{x} + \boldsymbol{b}$$
$$\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{z}}\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}} = \dots$$

Find $\frac{\partial s}{\partial b}$

1. 방정식을 간단히 분해

$s = u^T h \longrightarrow s = u^T h$

$h = f(Wx+b) \longrightarrow h = f(z)$

$\qquad\qquad\qquad\qquad z = Wx+b$

x(input) $\qquad\qquad x$(input)

2. Chain Rule 적용.

$s = u^T h$

$h = f(z)$

$z = Wx + b$

x(input)

$$\frac{\partial s}{\partial b} = \frac{\partial s}{\partial h} \cdot \frac{\partial h}{\partial z} \cdot \frac{\partial z}{\partial b}$$

3. Jacobians 사용

$\frac{\partial}{\partial u}(u^T h) = h^T$

$\frac{\partial}{\partial z}(f(z)) = \text{diag}(f'(z))$

$\frac{\partial}{\partial b}(Wx+b) = I$

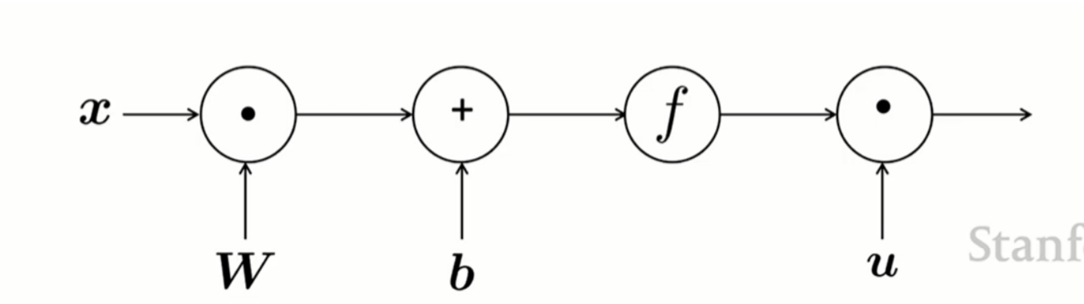
$$\frac{\partial s}{\partial b} = u^T \text{diag}(f'(z)) I$$
$$= u^T \circ f'(z)$$

$\therefore$ $\frac{\partial s}{\partial W}$ 를 얻기 위해서는 $\frac{\partial s}{\partial W} = \frac{\partial s}{\partial h} \cdot \frac{\partial h}{\partial z} \cdot \frac{\partial z}{\partial W} = \delta \frac{\partial z}{\partial W}$

$\frac{\partial s}{\partial b} = \delta \frac{\partial z}{\partial b} = \delta$

$\delta = \frac{\partial s}{\partial h} \cdot \frac{\partial h}{\partial z} = u^T \circ f'(z)$

5.Backpropagation

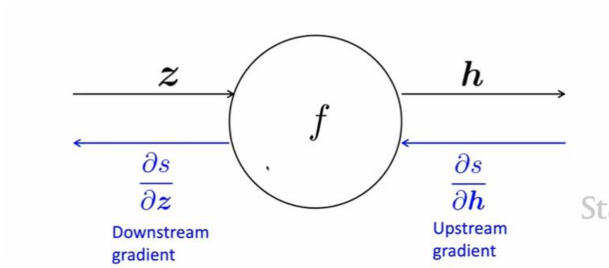


-원하는 값을 구하기 위해 각 단계의 편미분을 구하고 현재 단계의 편미분과 다음 단계의 편미분을 곱하는 것의 반복

-값을 구하기 위해서 forward 방향과는 반대 방향인 backward 과정으로 진행된다.

-위 그림은 forward 방향으로 진행되는 과정을 나타낸 것이고 backpropagation은 위 그림에 나타난 방향과 반대 방향으로 편미분을 진행시킨다.

1)single node일 경우



2)multiple node일 경우