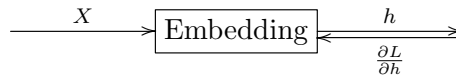


4강 연습문제

1. Embedding층의 가중치 행렬 W_{in} , 입력되는 단어 id의 배치 묶음 X , 은닉층에서 흘러들어오는 미분 $\frac{\partial L}{\partial h}$ 이 각각

$$W_{in} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix}, \quad X = [2, 0, 3, 0], \quad \frac{\partial L}{\partial h} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

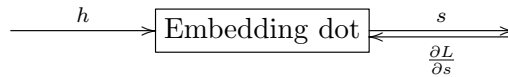
이다.



- (i) Embedding층 방법으로 출력값 h 와 가중치 행렬에 대한 미분 $\frac{\partial L}{\partial W_{in}}$ 을 각각 계산하시오.
(ii) 행렬곱층 계산 그래프로 검증하시오.
(iii) `layers.py`안의 Embedding 클래스로 인스턴스를 만들어서 검증하시오.
2. Embedding dot층의 가중치 행렬 W_{out} , 입력되는 은닉 벡터의 배치 묶음 h , 위에서 흘러들어오는 미분 $\frac{\partial L}{\partial s}$, 타겟 단어 id의 배치 묶음 t 가 각각

$$W_{out} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{pmatrix}, \quad h = \begin{pmatrix} 0 & -1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 1 \end{pmatrix}, \quad \frac{\partial L}{\partial s} = [1, 2, 3], \quad t = [2, 0, 0]$$

이다.



- (i) 출력되는 스코어 s 를 구하시오. (PPT 7 페이지)
(ii) W_{out} 을 transpose하고 Embedding층을 거쳐 W 를 구하시오. (PPT 8페이지)
(iii) 계산 그래프를 통해

$$s, \quad \frac{\partial L}{\partial h}, \quad \frac{\partial L}{\partial W}$$

을 구하시오. (PPT 8, 9페이지)

- (iv) $\frac{\partial L}{\partial W}$ 을 embedding층 역전파를 한후 transpose를 해서 $\frac{\partial L}{\partial W_{out}}$ 을 구하시오.
(v) `negative_sampling_layer.py`안의 EmbeddingDot 클래스로 인스턴스를 만들어서 검증하시오.

3. 행렬곱층과 Embedding층의 속도가 몇배 차이가 나는지 다음과 같은 코드를 작성하여 확인하시오.

- 말뭉치는 PTB 데이터셋 사용
- 행은 PTB 데이터셋의 어휘수, 열은 100개인 가중치 행렬을 표준정규분포를 따라 랜덤하게 생성
- MatMul 클래스와 Embedding 클래스로 인스턴스 생성
- time 라이브러리를 사용하여 앞 100,000개 단어의 묶음에 대하여 두 인스턴스의 순전과 속도 산출