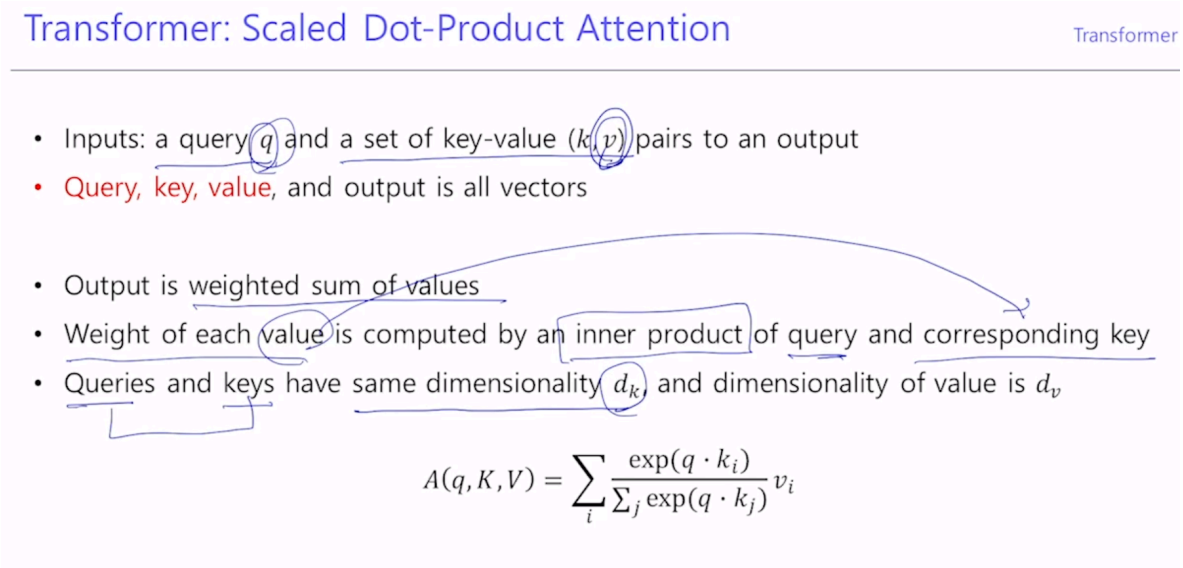
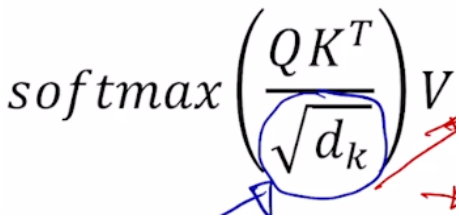
Week4-4

1. 학습정리
    1. transformer
        1. 어텐션 만으로 RNN 구조 대체
        2. 쿼리벡터와 키벡터를 내적해서 어느 벡터가 유사도가 높은지 계산
        3. 밸류벡터는 위에서 계산된 유사도를 적용할 벡터
        4. 밸류에 유사도를 가중평균 하면 인코딩 벡터가 나옴
        5. 이 인코딩 벡터는 모든 단어들을 고려한 인코딩 벡터
        6. 셀프어텐션 모델
            1. 기존에는 멀리있는 정보는 RNN 모듈을 타임스텝 차이만큼 반복적으로 통과해야함
            2. 셀프어텐션은 롱텀디펜던시 문제를 근본적으로 해결한 시퀀스 인코딩 기법



        7. 수식
        8. 밸류벡터는 키, 밸류벡터와 길이가 달라도 됨
        9. 내적에 참여하는 쿼리와 키 벡터의 분산에 따라 내적값이 크게 좌지될 수 있어서 내적값에 분산을 일정하게 유지시켜 줄 필요가 있음



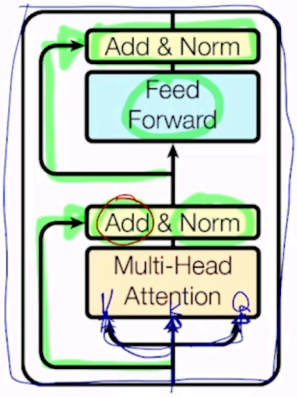
    2. 멀티헤드어텐션
        1. 구조
            1. 여러 버전의 q,k,v 벡터가 존재
            2. 서로 다른 버전의 인코딩 벡터가 나오고 이 것들은 컨캣함으로써 최종 인코딩 벡터를 얻음
        2. 필요성

        1. 여러 측면에서의 정보를 뽑기 위함
        2. 예 : 주체의 행동 중심, 장소 중심의 변화를 각자 캐치
    3. 연산
        1. 메모리 요구량이 큼
3. 포지셔널인코딩
    1. 셀프어텐션 문제점
        1. 셀프어텐션 기법은 단어들의 위치정보를 반영할 수 없음
    2. 방법
        1. 각 순서를 특정지을 수 있는 상수 벡터를 각 순서의 입력 워드 벡터에 더해줌
4. 러닝레이트 스케쥴링
    1. 학습중에 러닝레이트가 적절하게 변경되도록 하는 것

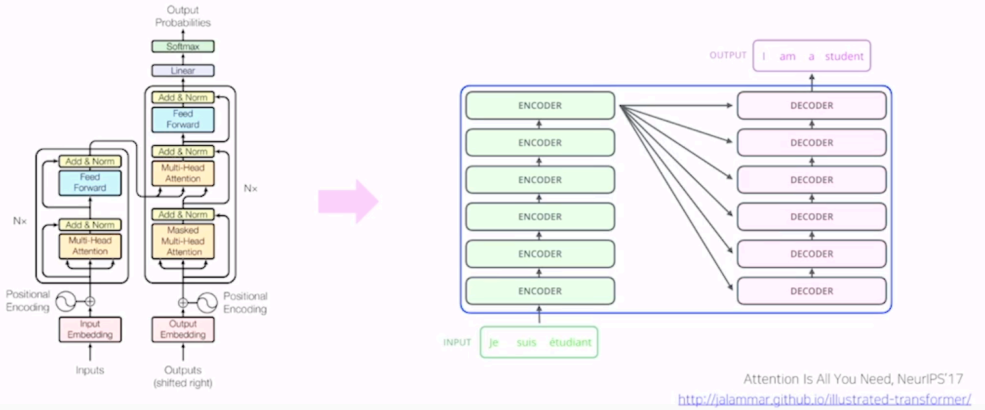
Transformer: Block-Based Model

Transformer

- Each block has two sub-layers
  - Multi-head attention
  - Two-layer feed-forward NN (with ReLU)
- Each of these two steps also has
  - Residual connection and layer normalization:
  - $LayerNorm(x + sublayer(x))$



- Attention is all you need, NeurIPS'17
  - No more RNN or CNN modules



    2. 마스크드 셀프 어텐션
        1. 디코딩 과정에서 접근하지 말아야 할 단어들의 접근정보를 0으로 처리
        2. kq 내적 해서 나온 행렬의 대각 윗부분을 0으로 만듦

1. 피어세션
    1. 인터뷰 연습
        1. https://docs.google.com/document/d/1F3ZWNVTLRPncF30iW10cup5R40ypXBQHlFi61CJk6Q0/edit