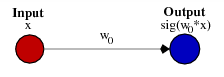
Bias term의 역할: <https://stackoverflow.com/questions/2480650/what-is-the-role-of-the-bias-in-neural-networks>

Bias term은 항상 helpful? Activaton function을 왼쪽에서 오른쪽으로 옮겨줌

Sigmoid(weight \* x) 를 생각 => 에서, x가 2라면 output을 0으로 만들기 위해 어떻게 해야할까? => bias term이 필요(shift의 역할)

Softmax는 혼자서 (logistic regression ~ linear regression with cross entropy)는 강력하지 않다. 확률로 바꿔주는 작업을 할텐데, 이 작업은 오직 linear decision boundary를 제공한다. 따라서 non-linear, 복잡한 모델 = neural net과 함께 동작할수록 강력하다.

Linear regression에 non-linear 기능을 더해 one vs all 형태로 classification을 class 만큼 하는 것 = neural net은 한번에!(more non-linear)

왜 non-linearity가 필요한가? Variance가 높아질지라도 더 다양한 분포의 데이터를 예측할 수 있음

Name Entity Recognition(NER) : 특정한, 특이한 entity를 찾는 것( ex) European Commission, German, Fischler ..) / QA의 답이 주로 named entities며 관련이 깊다. / slot-filling classifications에 똑같은 기술이 쓰인다.(빈칸 채우기) => 종종 KB에도 사용됨.

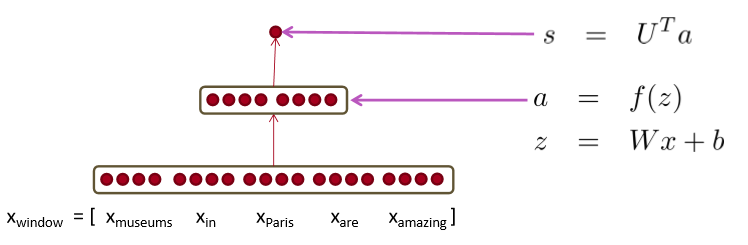
NER의 어려움: 어디까지가 N인지 구별하기 힘듬 First National Bank vs National Bank / Future School이 N인지 그저 그냥 평범한 future school 인지 알 수 없음(대소문자로는 판단하면 안됨, 변수 많아서) / Zig Ziglar가 새로운 N인데 사람인지 지역인지 알 수 가 없다. / 문맥에 따라 Charles Schwab이 사람인지 기관인지 애매하다.

B-ORG: ORG 개체명 시작 글자 / I-ORG: ORG 개체명 내부 글자 / O: 개체명이 아닌 글자

단어 한 개만으로 모델에 넣어 classfy 하지는 않는다. 단어는 여러 의미로 쓰일 수 있기 때문이다. = ambiguous를 조심 / window classification 이란 기법이 존재

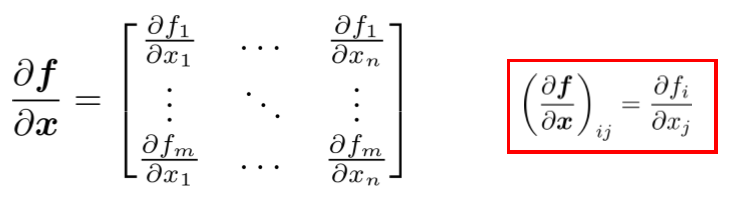
Window classification: neighboring words(in window)를 이용해 중심 word를 classify / 여러 방법들 있음.

1. 주변 단어들과 중심 단어를 벡터화 해서 평균 냄 => position 정보를 잃게 된다.
2. 또는 그냥 쭉 늘여 놓자! =>  
   



이런 형태로, 최종 스칼라의 스코어를 얻게 된다.

Gradients => Jacobian Matrix 까지 확장 가능



F(x) = [f1(x1,x2…, xn), …., fm(x1,x2,…,xn)] -> f는 여러 함수들을 포함하고 각 함수를 변수들에 대해 미분을 해준다.