

LSTMs와 그 이해.

LSTMs의 기본 형태인 RNN(Recurrent Neural Network)은 다른 구조와는 달리 한 순간의 데이터만을 사용하는것을 넘어 그 전 순간들의 데이터 역시 사용할 수 있는 특징을 보인다. 그러나 RNN은 근본적인 문제를 안고 있는데, 뉴럴 네트워크의 연산이 모두 곱연산으로 진행된다 보니 몇스텝 이상을 거슬러 올라가 데이터를 참조하려고 해도 그 값이 거의 0에 가깝게 사라져서 참조가 불가능한것이 그것이다(그림 1 참조)

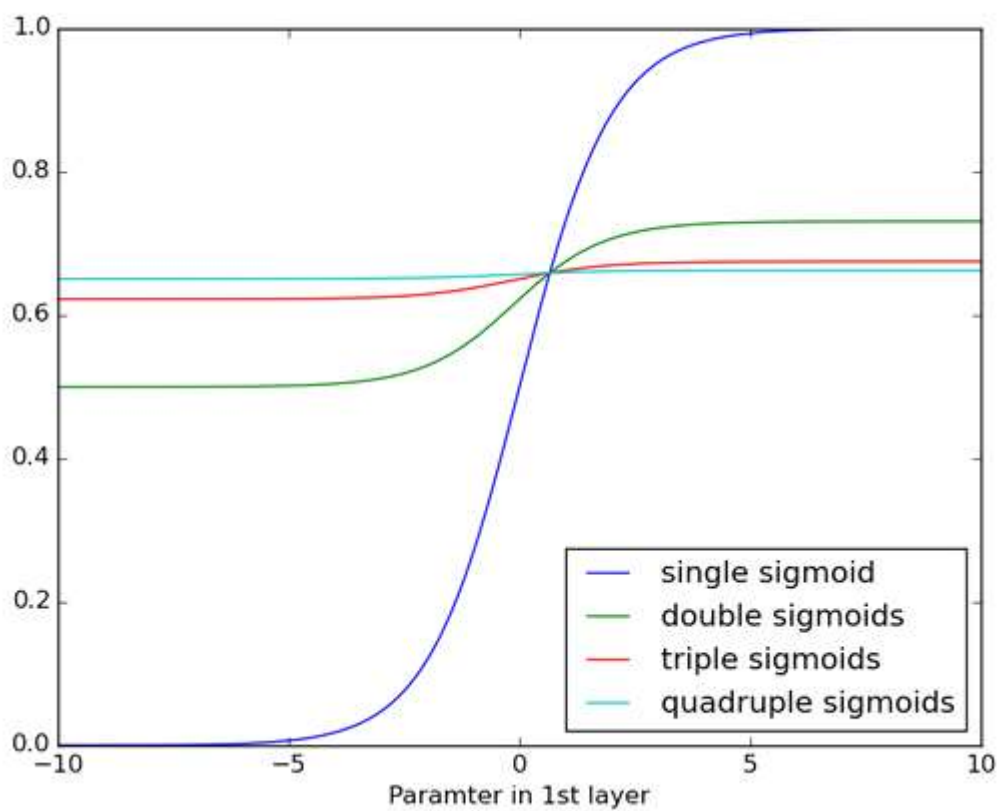
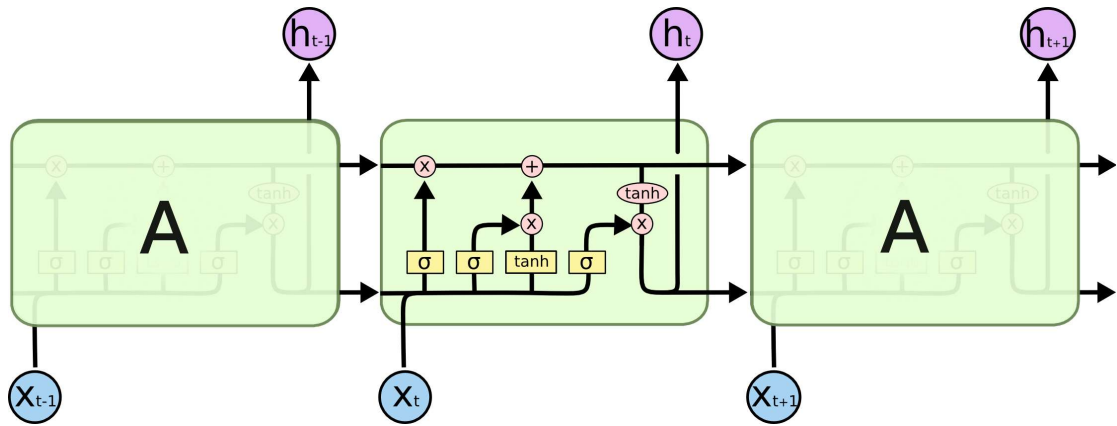


그림 1

하지만 LSTMs는 이러한 문제를 다른 접근법으로 해결했는데, 어쩌다 보니 '기억'을 활용할 수 있게 된 RNN과는 달리 체계적으로 '기억'을 관리하는 메커니즘을 만든것이다.



위 그림은 LSTMs의 기본 구조다. 이 그림에서 노란 상자는 뉴럴 네트워크 레이어를 뜻하고, 검은 선은 벡터의 이동, 붉은 원은 점대점 계산을 뜻한다. 이때 최상단을 지나가는 선이 바로 LSTMs의 '기억'을 나타내는 선이다. 이때 아래쪽에서 들어오는 x_t 와 그 전 사이클에서 들어오는 '연관성' 정보가 합쳐진 벡터가 시그모이드 레이어를 통과한 값이 위의 '기억'정보와 곱 연산이 일어나며 얼마나 '망각'할지 결정한다. 이는 데이터가 멍치로 이뤄져 있을때(한개 문장을 완성하고 다음 문장으로 넘어가는 등) 그 멍치의 경계를 인식하게 되는 구조가 된다. 그 후 두번째 시그모이드 레이어는 입력값 x_t 에서 새로이 '기억'에 추가할 정보를 추출하고, tanh를 통해 그 정보를 벡터화 시킨 후, '기억'에 합연산으로 추가한다. 그 뒤 기억 정보는 다음 사이클로 넘어가고, 세번째 시그모이드 레이어가 x_t 에 대해 연관성 연산을 수행한 뒤 tanh를 가한 기억정보에 곱연산 되어 연관성 정보로서 다음 사이클로 넘어간다.