

# Chapter 10. 순차열 모형화를 위한 순환 신경망과 재귀 신경망

2023.04.11

김다영

# 순환 신경망이란?

- recurrent neural networks, RNN
- 순차적인 자료를 처리하는 신경망의 한 종류
- 순차열, 즉 순서가 있는 일련의 값을 처리하는 데 특화됨
- 합성곱 신경망을 너비와 높이가 더 큰 이미지들로 손쉽게 확장할 수 있는 것과 마찬가지로, 순환 신경망 역시 더 긴 순차열로 손쉽게 확장 가능

# 10.1 계산 그래프 펼치기

- 계산 그래프란?

계산들의 구조를 공식화하는 한 방법

-> 입력을 출력으로 사상하는 계산, 매개변수들을 손실값으로 사상하는 계산을 계산 그래프 형태로 표현 가능

- 펼치기(Unpolding)?

재귀적인 또는 순환적인 계산을 펼쳐서 반복구조를 가진 계산 그래프 형태를 얻는 것

-> 심층 신경망 구조 전체에서 매개변수들이 공유되는 효과가 생김

# 10.1 계산 그래프 펼치기

ex)

고전적인 형태의 동역학계 (classical form of a dynamical system)

시간 색인  $t$ 는 1에서  $r$ 까지의 정수

$$\mathbf{s}^{(t)} = f(\mathbf{s}^{(t-1)}; \boldsymbol{\theta})$$

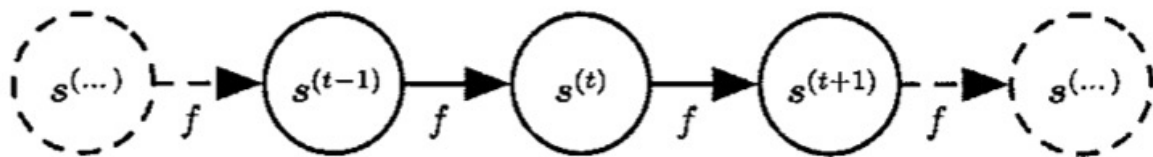


계의 상태(state)

시간  $t$ 에서의 상태  $s$ 의 정의가 시간  $t-1$ 에서의 동일한 정의를 참조 -> 점화식

## 10.1 계산 그래프 펼치기

- 식 10.1이 서술하는 고전적인 동역학계를 펼친 계산 그래프



- $r = 3$ 의 시간 단계에 대해 식을 펼친 것

$$\begin{aligned} \mathbf{s}^{(3)} &= f(\mathbf{s}^{(2)}; \boldsymbol{\theta}) \\ &= f(f(\mathbf{s}^{(1)}; \boldsymbol{\theta}); \boldsymbol{\theta}). \end{aligned}$$

## 10.1 계산 그래프 펼치기

- 외부 신호  $x^{(t)}$ 가 이끄는 동역학계

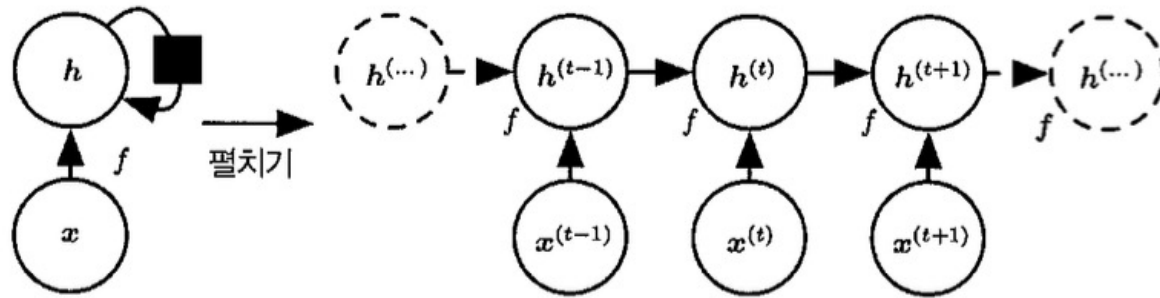
$$\mathbf{s}^{(t)} = f(\mathbf{s}^{(t-1)}, \mathbf{x}^{(t)}; \boldsymbol{\theta}).$$

- 순환 신경망의 은닉 단위를 정의하는 공식

-> 신경망의 은닉 단위들의 상태임을 나타내기 위해 변수  $h$ 를 씀

$$\mathbf{h}^{(t)} = f(\mathbf{h}^{(t-1)}, \mathbf{x}^{(t)}; \boldsymbol{\theta}).$$

## 10.1 계산 그래프 펼치기



1. 모형의 물리적 구현에 존재할만한 각 구성요소를 개별 노드로 두는 것  
-> 신경망은 실시간으로 작동하는 하나의 회로를 정의

2. RNN을 펼쳐진 계산 그래프로 표현하는 것  
-> 여러 계산 요소의 개별 시간 단계에서의 상태를 나타내는 다수의 변수로 표현

# 10.1 계산 그래프 펼치기

t단계로 펼쳐진 점화식을  $g^{(t)}$ 함수로 표현 가능

$$\begin{aligned} \mathbf{h}^{(t)} &= g^{(t)}(\mathbf{x}^{(t)}, \mathbf{x}^{(t-1)}, \mathbf{x}^{(t-2)}, \dots, \mathbf{x}^{(2)}, \mathbf{x}^{(1)}) \\ &= f(\mathbf{h}^{(t-1)}, \mathbf{x}^{(t)}; \theta). \end{aligned}$$

## 펼치기 과정의 장점

1. 학습된 모형은 길이가 가변적인 상태 역사(과거 상태의 순차열)가 아니라 한 상태에서 다른 상태로의 전이를 명시, 입력 크기는 순차열의 길이와 무관하게 항상 동일
2. 모든 시간 단계에서 동일한 전이 함수  $f$ 를 동일한 매개변수들과 함께 적용하는 것이 가능



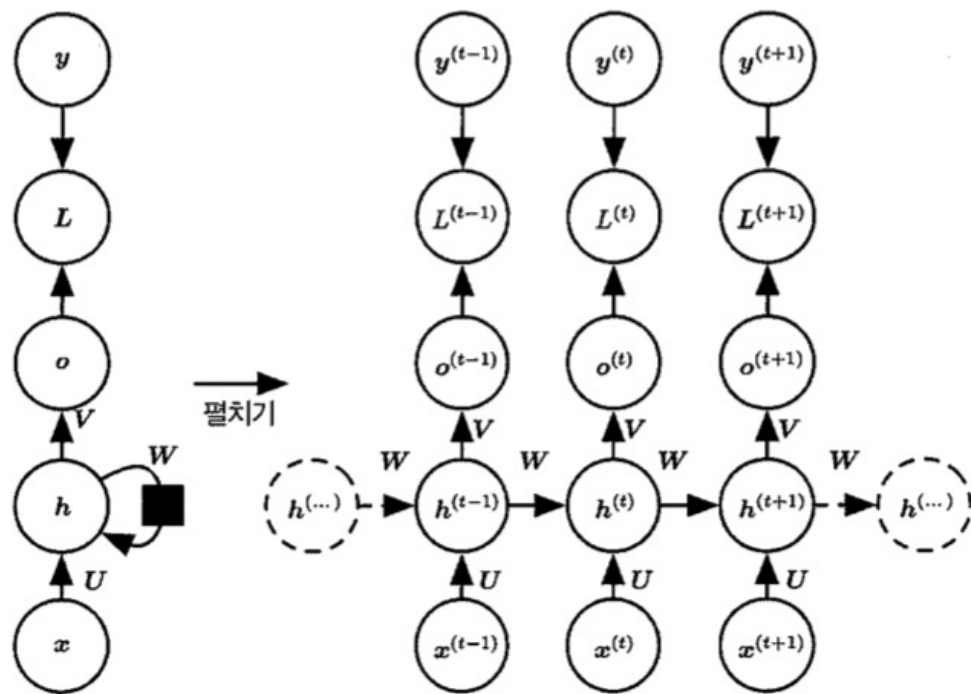
## 10.2 순환 신경망

### 순환 신경망의 설계에서 자주 보이는 중요한 패턴

1. 각 시간 단계에서 하나의 출력을 산출, 은닉 단위들 사이에 순환 연결들이 존재하는 순환 신경망
2. 각 시간 단계에서 하나의 출력을 산출, 한 시간 단계의 출력과 그 다음 시간 단계의 은닉 단위들 사이에서만 순환 연결들이 존재하는 순환 신경망
3. 은닉 단위들 사이에 순환 연결들이 존재, 순차열 전체를 읽어서 하나의 출력을 산출하는 순환 신경망

## 10.2 순환 신경망

- 순환 신경망의 대표적인 예시는 1번



$x$  값들로 이루어진 입력 순차열을  $o$  값들로 이루어진 해당 출력 순차열로 사상하는 순환 신경망의 훈련 손실을 계산하는 계산 그래프

손실함수  $L$ 은 각  $o$ 가 훈련 목표  $y$ 와 얼마나 떨어져 있는지를 나타냄

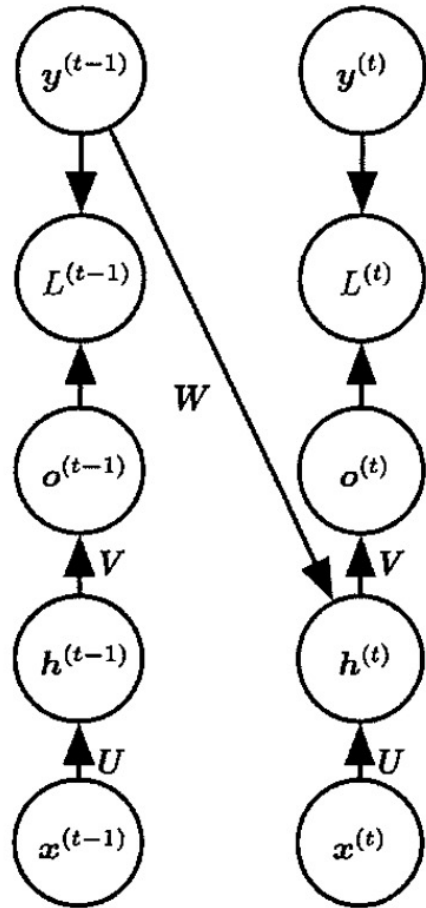
## 10.2.1 교사 강제와 출력 순환 연결이 있는 신경망

- 2번 신경망은 은닉 대 은닉 순환 연결이 없어서 표현력이 약함
- 신경망이 미래를 예측하는 데 사용할 과거의 모든 정보를 출력 단위들이 반영해야함
- 출력 단위들은 훈련 집합의 목표값들과 부합하도록 명시적으로 훈련되므로, 입력의 과거 역사에 관한 필수 정보를 출력 단위들이 포착할 가능성은 작음
- 손실함수가 시간  $t$ 에서의 예측과 시간  $t$ 에서의 훈련 목표값의 비교에 기초하는 경우, 순환 연결이 없으면 모든 시간 단계를 분리  
-> 각 단계  $t$ 에서의 기울기를 개별적으로 계산하는 것이 가능

## 10.2.1 교사 강제와 출력 순환 연결이 있는 신경망

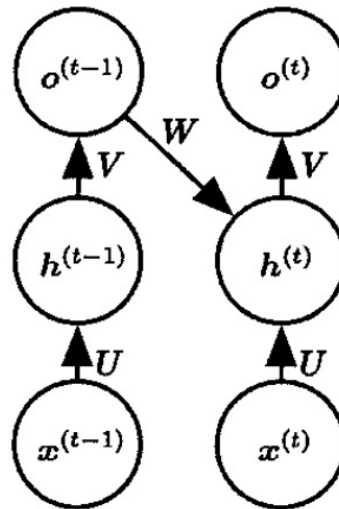
- 출력에서 모형의 내부로 돌아가는 순환 연결이 있는 모형에는 교사 강제(teacher forcing)라는 훈련 기법을 적용할 수 있음
- 교사 강제는 최대가능도 판정기준에서 비롯된 하나의 절차, 모형을 훈련하는 과정에서 바탕 출력 참값  $y^{(t)}$ 를 시간  $t+1$ 에서의 입력으로 사용

## 10.2.1 교사 강제와 출력 순환 연결이 있는 신경망



훈련 시간

훈련 도중에는 훈련 집합에 있는 진 출력(출력의 참값)  $y^{(t)}$ 를  $h^{(t+1)}$ 의 입력으로 공급



시험 시간

모형을 과제에 적용할 때에는 진 출력을 모르는 것이 일반적  
-> 진 출력  $y^{(t)}$ 를 모형의 출력  $o^{(t)}$ 로 근사  
-> 다시 모형의 입력으로 공급

## 10.2.2 순환 신경망의 기울기 계산

- 6절에서 나온 일반화된 역전파 알고리즘을 펼쳐진 계산 그래프에 적용

$$\frac{\partial L}{\partial L^{(t)}} = 1.$$

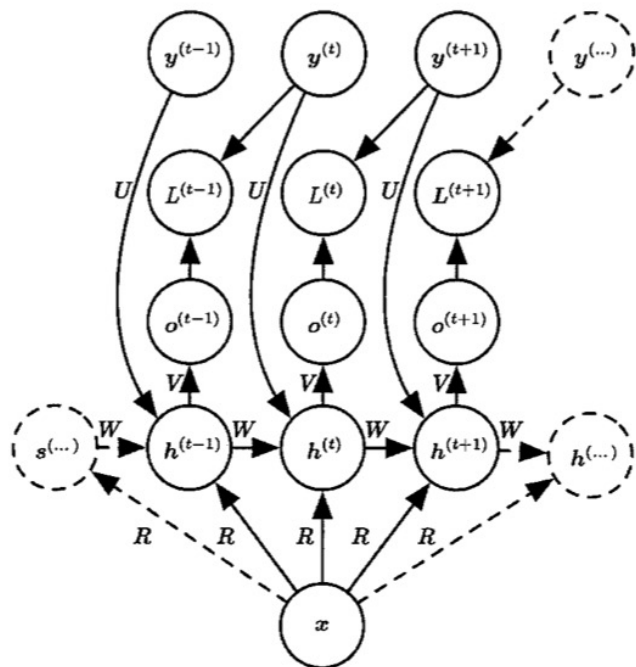
$$\nabla_{\mathbf{h}^{(t)}} L = \mathbf{V}^\top \nabla_{\mathbf{o}^{(t)}} L.$$

$$\begin{aligned}\nabla_{\mathbf{h}^{(t)}} L &= \left( \frac{\partial \mathbf{h}^{(t+1)}}{\partial \mathbf{h}^{(t)}} \right)^\top (\nabla_{\mathbf{h}^{(t+1)}} L) + \left( \frac{\partial \mathbf{o}^{(t)}}{\partial \mathbf{h}^{(t)}} \right)^\top (\nabla_{\mathbf{o}^{(t)}} L) \\ &= \mathbf{W}^\top (\nabla_{\mathbf{h}^{(t+1)}} L) \text{diag}(1 - (\mathbf{h}^{(t+1)})^2) + \mathbf{V}^\top (\nabla_{\mathbf{o}^{(t)}} L).\end{aligned}$$

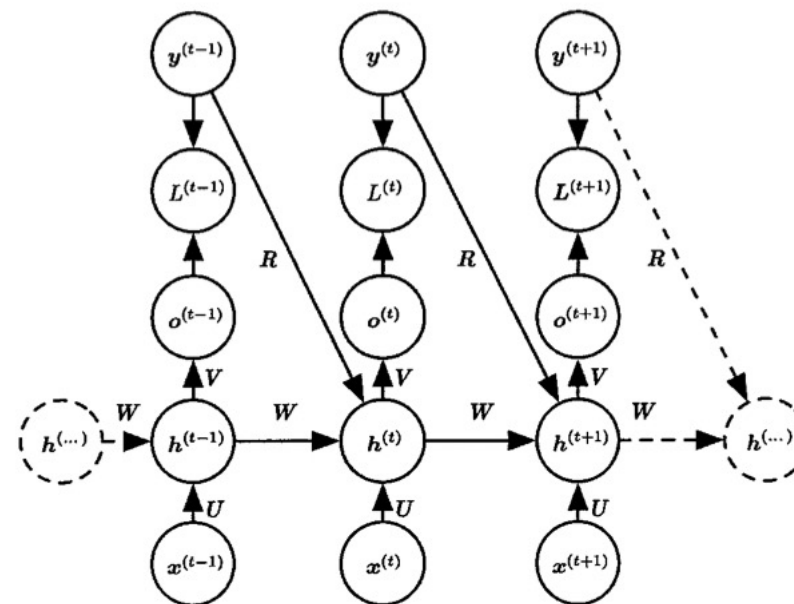
## 10.2.4 문맥에 따라 조건화되는 순차열을 모형화하는 RNN

- 앞에서 설명한 RNN은  $t = 1, \dots, r$ 에 대한 벡터  $x^{(t)}$ 들의 순차열을 입력 받음
- 벡터  $x$  하나를 입력받는 RNN도 가능함
  - >  $x$ 가 고정 크기 벡터일 때는  $y$  순차열을 산출하는 RNN의 여분의 입력으로 두면 가능함
- RNN에 여분의 입력을 제공하는 데 쓰이는 방법
  1. 각 시간 단계에 여분의 입력을 추가
  2. 여분의 입력을 초기 상태  $h^{(0)}$ 으로 둠
  3. 두 방법 모두 적용

## 10.2.4 문맥에 따라 조건화되는 순차열을 모형화하는 RNN



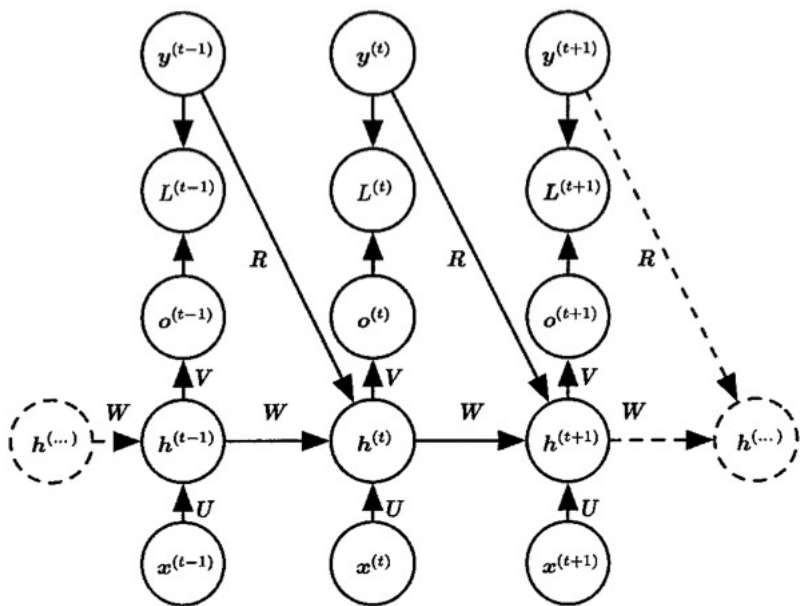
고정 길이 벡터  $x$ 를 순차열  $y$ 에 관한 분포로  
사상하는 RNN



x 값들의 가변 길이 순차열을 같은 길이의 y 값들의 순차열로 사상하는 조건부 순환 신경망



## 10.2.4 문맥에 따라 조건화되는 순차열을 모형화하는 RNN



시간  $t$ 에서의 출력에서 시간  $t+1$ 에서의 은닉 단위로의 연결을 신경망에 추가하면 그러한 조건부 독립성 가정을 제거 가능

-> 그러한 연결들이 추가된 모형은  $y$  순차열에 관한 임의의 확률분포를 표현 가능

-> 두 순차열의 길이가 같아야 함

$$\prod_t P(\mathbf{y}^{(t)} | \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(t)}).$$

->  $y$  순차열에 관한 확률분포를 표현 가능

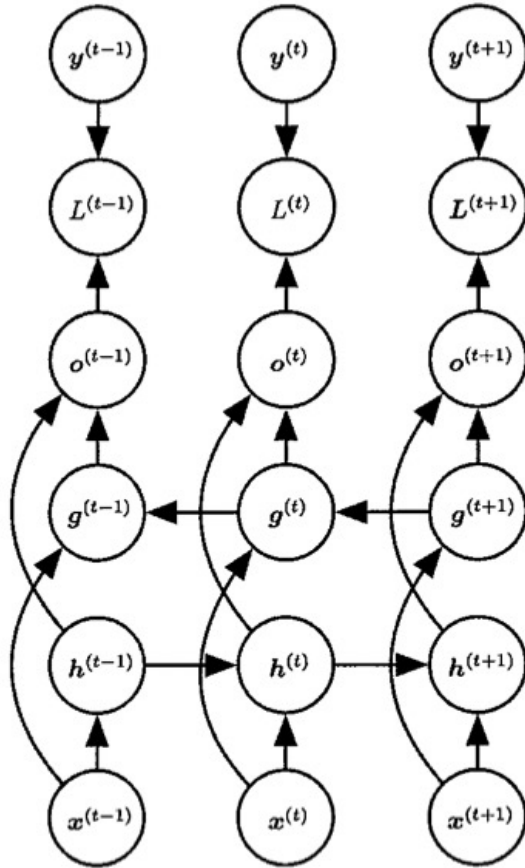
## 10.3 양방향 순환 신경망

- 여러 응용에서 신경망이 출력할 예측값  $y^{(t)}$ 가 입력 순차열 전체에 의존할 수 있음

ex) 음성 인식 -> 하나의 유효한 음소로 정확히 해석하려면 그다음 몇몇 음소도 고려해야함

- 양방향 순환 신경망  
-> 시간순 RNN과 역시간순 RNN을 결합

## 10.3 양방향 순환 신경망



$h^{(t)}$  는 시간순 RNN 의 상태,  
 $g^{(t)}$  는 역시간순 RNN의 상태

-> 출력 단위  $o^{(t)}$ 들은 과거와 미래 모두에 의존, 시간  $t$  근처의 입력값들에 가장 민감하게 반응하는 표현을 계산 가능

-> 고정 크기 미리 보기 버퍼를 가진 보통의 RNN으로 표현을 계산하려면  $t$ 를 중심으로 한 고정 크기 구간을 정해야 하지만, 이 구조에서는 그럴 필요 x

## 10.4 부호기-복호기 순차열 대 순차열 아키텍처

- 입력 순차열과는 길이가 다를 수도 있는 출력 순차열을 산출하도록 RNN을 훈련하는 방법

-> 음성 인식, 기계 번역, 질문 응답 등 응용

두 가지 아키텍처

1. 부호기-복호기 아키텍처
2. 순차열 대 순차열

## 10.4 부호기-복호기 순차열 대 순차열 아키텍처

- 부호기 복호기 순차열

1. 판독기 또는 입력 RNN이라고도 부르는 부호기 RNN은 입력 순차열을 처리
2. 기록기 또는 출력 RNN이라고도 부르는 복호기 RNN은 고정 길이 벡터를 조건으로 하여 출력 순차열을 산출

-> 순차열 길이  $n_x, n_y$ 가 서로 다를 수 있음

## 10.4 부호기-복호기 순차열 대 순차열 아키텍처

- 순차열 대 순차열 아키텍처

훈련의 목적이 훈련 집합의 순차열  $x$ 와  $y$ 의 모든 쌍에 대한

$$\log P(\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(n_y)} | \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n_x)})$$

평균을 최대화

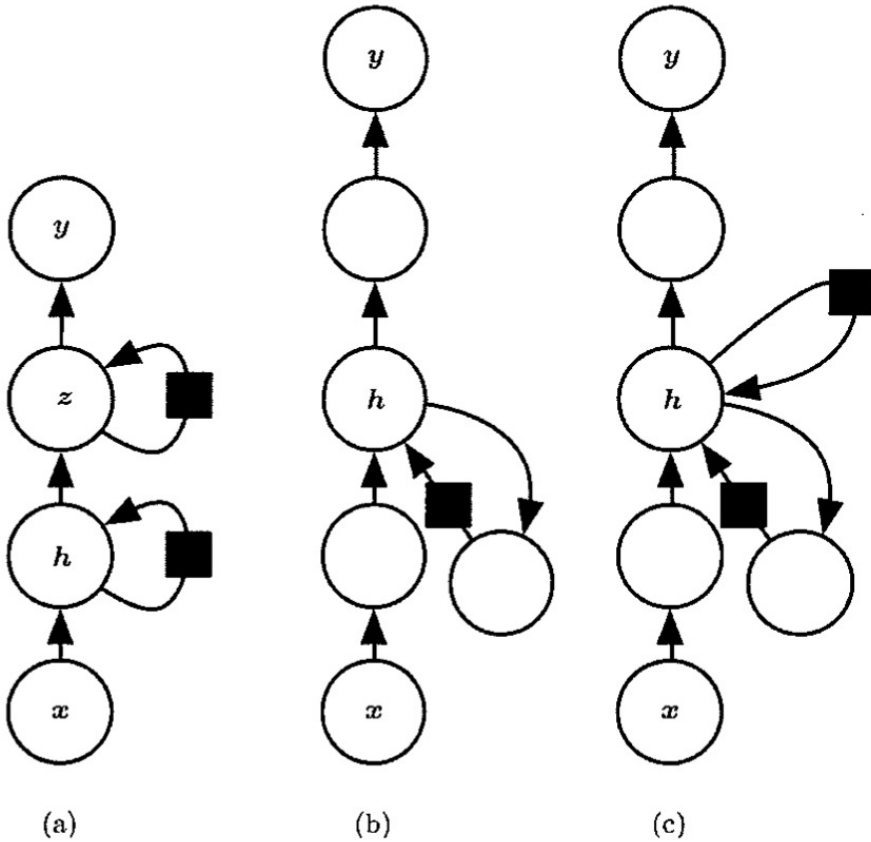
-> 부호기의 은닉층과 복호기의 은닉층의 크기가 같아야 한다는 제약이 없음

## 10.5 심층 순환 신경망

대부분의 RNN이 수행하는 계산은 매개변수들을 크게 3가지로 나누고 각각을 적절한 방식으로 변환하는 것에 해당한다.

1. 입력에서 은닉 상태로의 변환
2. 이전 은닉 상태에서 다음 은닉 상태로의 변환
3. 은닉 상태에서 출력으로의 변환

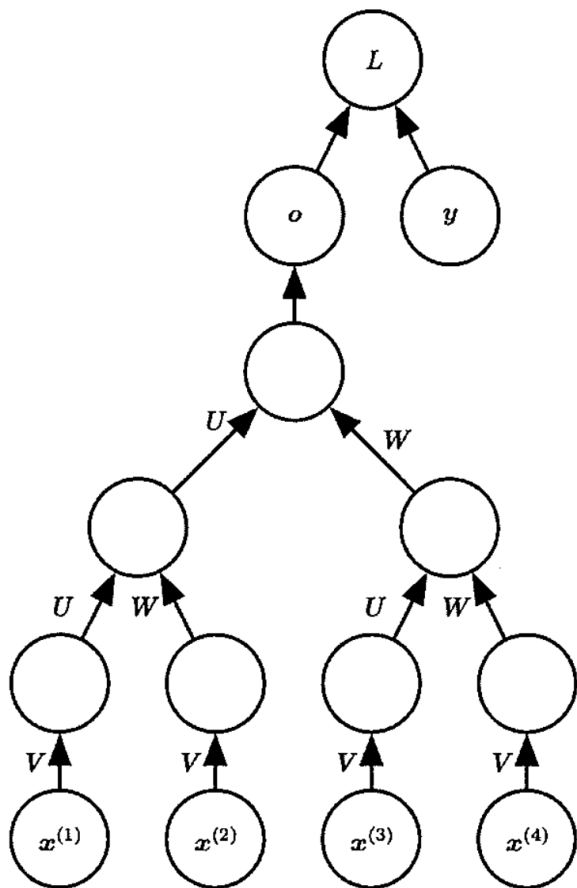
## 10.5 심층 순환 신경망



- (a) 하나의 은닉 순환 상태를 계통구조를 따르는 여러 그룹으로 분해
- (b) 입력-은닉, 은닉-은닉, 은닉-출력 블록에 더 깊은 계산을 도입
- (c) 경로 연장 문제는 건너뛰기 연결을 도입해서 완화



## 10.5 심층 순환 신경망



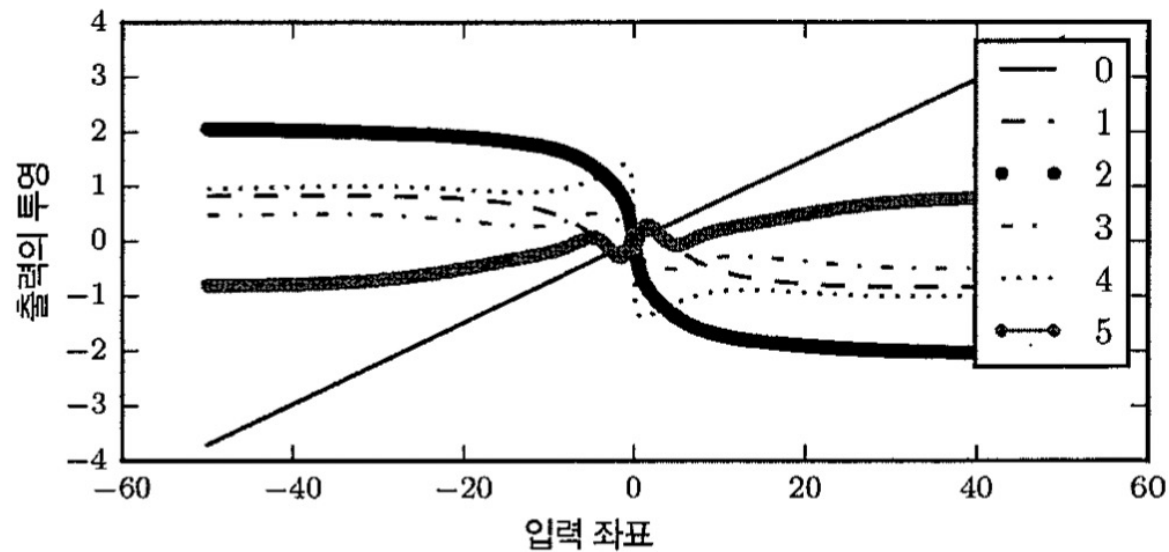
순차열 전체와 연관된 어떤 목표값  $y$ 가  
주어진 지도 학습 상황을 나타낸 것

## 10.6 재귀 신경망

- 순환 신경망의 또 다른 일반화
  - 다른 종류의 계산 그래프를 사용
  - 재귀 신경망의 계산 그래프는 순환 신경망에 쓰이는 사슬 형태의 계산 그래프를 트리 형태로 일반화 한 것
  - 자료구조를 신경망의 입력으로 삼아서 처리하는 응용 분야와 자연어 처리 그리고 컴퓨터 비전에 성공적으로 적용됨
  - 순차열의 길이가  $r$ 이라고 할 때 순환 신경망의 깊이는  $r$ 이지만 재귀 신경망의 깊이는  $O(\log r)$
- > 장기 의존관계를 다루는 데 도움됨

## 10.7 장기 의존성의 어려움

- 여러 단계에 걸쳐 전파되는 기울기들이 소멸하거나 폭발하는 경향이 나타남



## 10.8 반향 상태 신경망

- $h^{(t-1)}$ 에서  $h^{(t)}$ 로의 순환 가중치 사상과  $x^{(t)}$ 에서  $h^{(t)}$ 로의 입력 가중치 사상은 순환 신경망에서 가장 어려운 매개변수들
- 출력 가중치들만 배우게 하는 것으로 해소
- 반향 상태 신경망(ESN)과 액체 상태 기계가 제시됨  
-> 두가지를 통칭해서 저장소 컴퓨팅이라고 부름
- 반향 상태 신경망은 연속값 은닉 단위들을 사용

## 10.8 반향 상태 신경망

- 순환 신경망 상태가 풍부한 역사를 표현할 수 있도록 입력 가중치들과 순환 가중치들을 설정하는 방법은?

-> 순환 신경망을 하나의 동역학계로 간주해서, 입력 가중치들과 순환 가중치들을 동역학계가 안정성의 가장자리에 놓이게 하는 값들로 설정하는 것

## 10.9 누출 단위 및 여러 다중 시간 축척 전략

- 장기 의존성을 처리하는 방법

- > 모형이 여러 시간 축척들에서 작동하도록 설계

- 모형의 일부는 조밀한 시간 축척 척도에서 작동

- > 먼 미래의 정보가 현재로 좀 더 효율적으로 전달되게 만드는 것

- > 건너뛰기 연결을 모형에 추가

- > 서로 다른 시간 상수를 가진 누출 단위들로 여러 신호를 통합

- > 조밀한 연결들의 일부를 제거함

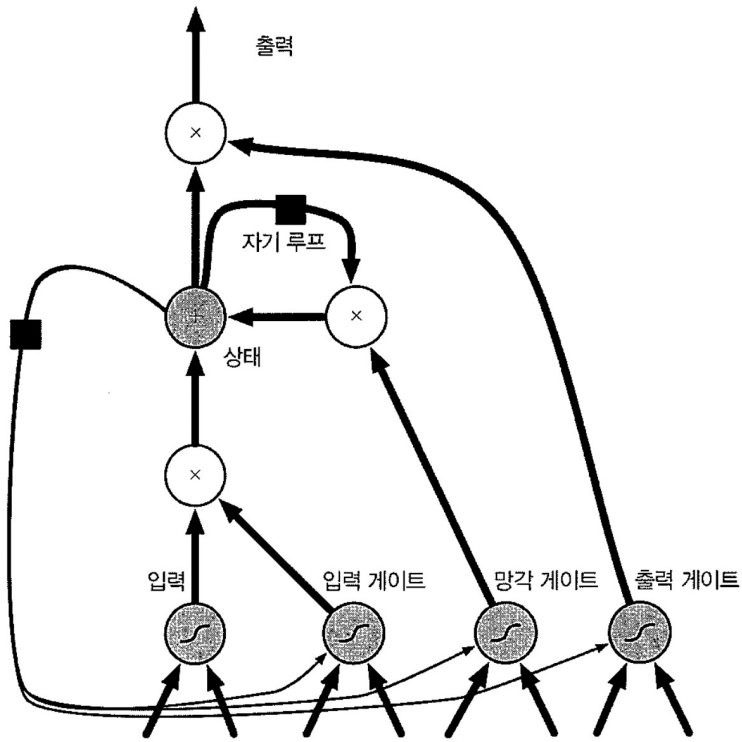
## 10.10 장단기 기억과 기타 게이트 제어 RNN들

- 장단기 기억(LSTM)

-> 초창기 장단기 기억 모형은 자기 루프를 도입함으로써 기울기가 오랜 기간 흐를 수 있는 경로가 만들어지게 한다는 착안에 기초

-> 가중치를 고정하는 것이 아닌 문맥에 대해 조건화한다는 중요한 수정이 가해짐

## 10.10 장단기 기억과 기타 게이트 제어 RNN들



장단기 기억의 구조도

-> 그 아래의 수식은얇은 순환 신경망 아키텍처를 위한 순전파를 정의



## 10.10 장단기 기억과 기타 게이트 제어 RNN들

- 순환 신경망과 장단기 기억(LSTM)의 주된 차이점

-> 순환 신경망은 하나의 게이트 단위가 상태 단위 갱신 여부와 망각 인자를 동시에 제어

$$h_i^{(t)} = u_i^{(t-1)} h_i^{(t-1)} + (1 - u_i^{(t-1)}) \sigma \left( b_i + \sum_j U_{i,j} x_j^{(t-1)} + \sum_j W_{i,j} r_j^{(t-1)} h_j^{(t-1)} \right).$$

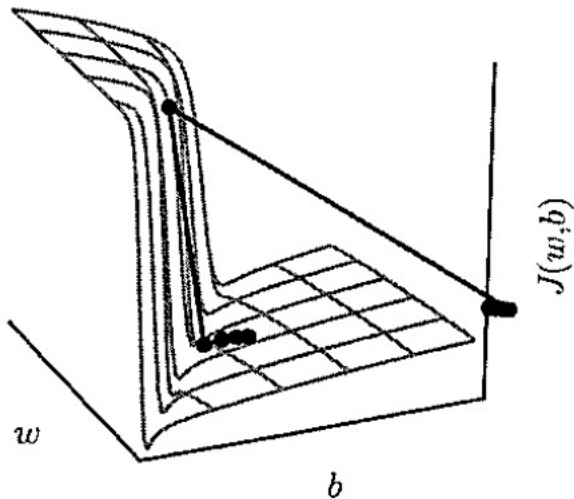
u는 갱신(update) r은 재설정(reset) 게이트에 해당

## 10.11 장기 의존성을 위한 최적화

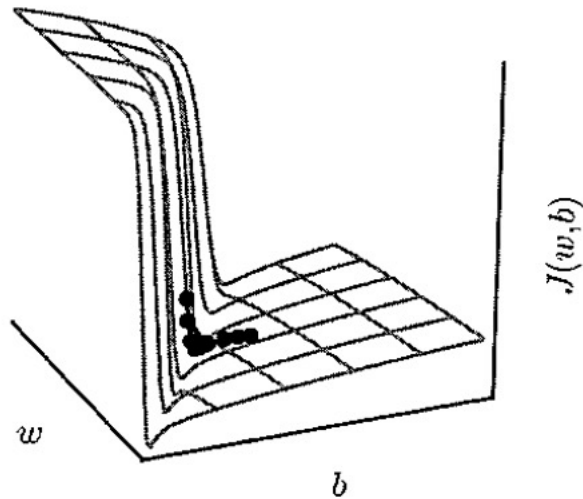
- 기울기 절단

고도의 비선형적인 함수들은 그 미분의 크기가 아주 커지거나 아주 작아지는 경향이 있음

절단 미적용



절단 적용



기울기 절단은 기울기의 폭발을 다루는 데  
도움됨  
-> 기울기의 소멸을 다루는 데에는 도움x

## 10.12 명시적 기억

- 명시적 기억에 관한 어려움을 해소하기 위해 기억망(memory network)을 소개

-> 초기에는 기억세포의 활용 방법을 지도하는 신호를 외부에서 공급해야 함

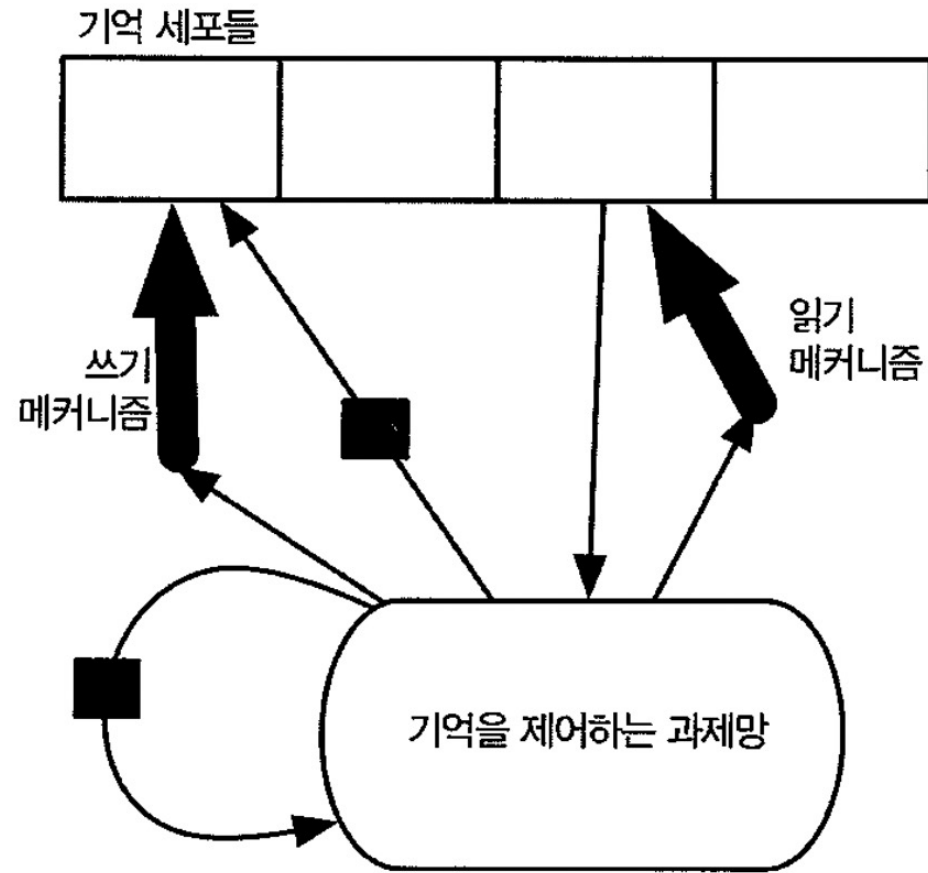
- 신경 튜링 기계(neural Turing machine)NTM

-> 어떤 행동을 취해야 하는지에 대한 외부 지도 신호 없이도 기억 세포들에서 임의의 내용을 읽고 쓰는 방법을 학습 가능

## 10.12 명시적 기억

- 신경 튜링 기계의 기억 세포는 여러 값으로 이루어진 벡터를 담는 것이 일반적
- 기억 세포의 크기를 늘리는 이유는 두가지
  - 1) 신경 튜링 기계에서는 기억 세포의 접근 비용이 더 큼
  - 2) 벡터를 사용하면 한 기억세포의 읽기나 쓰기에 쓰이는 가중치가 그 세포의 한 함수인 내용 기반 주소 접근이 가능

## 10.12 명시적 기억



명시적 기억을 가진 신경망의 구조도

감사합니다.