

딥러닝 1단계 1~2주차

이승목

2020.04.14.

범위

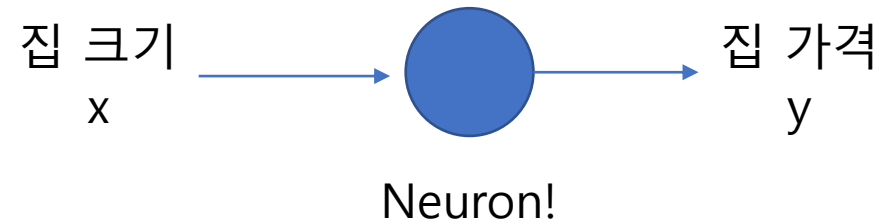
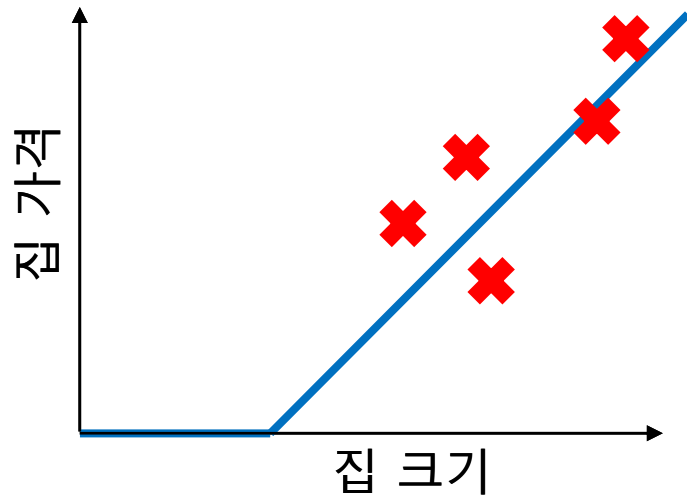
- Andrew Ng의 Deep Learning 1번 강의
 - <https://www.edwith.org/deeplearningai1>
- 1주차 '딥러닝 소개'
 - 환영합니다!
 - 신경망은 무엇인가?
 - 신경망을 이용한 지도학습
 - 왜 딥러닝이 뜨고 있을까요?
- 2주차 '신경망과 로지스틱 회귀'
 - 이진 분류
 - 로지스틱 회귀
 - 로지스틱 회귀의 비용함수
 - 경사하강법
 - 계산 그래프
 - 계산 그래프로 미분하기
 - 로지스틱 회귀의 경사하강법
 - M개 샘플의 경사하강법

Andrew Ng의 딥러닝 강의 목차

- 1단계 신경망과 딥러닝
- 2단계 심층 신경망 성능 향상시키기
 - 하이퍼 파라미터 조절, 정규화, 최적화
- 3단계 머신러닝 프로젝트 구조화하기
 - Dataset 구성법, end-to-end 딥러닝
- 4단계 합성곱 신경망 네트워크 (CNN)
- 5단계 자연어 처리
 - 시퀀스 모델. RNN, LSTM
 - Edwith 미제공. <https://www.coursera.org/learn/nlp-sequence-models#syllabus>

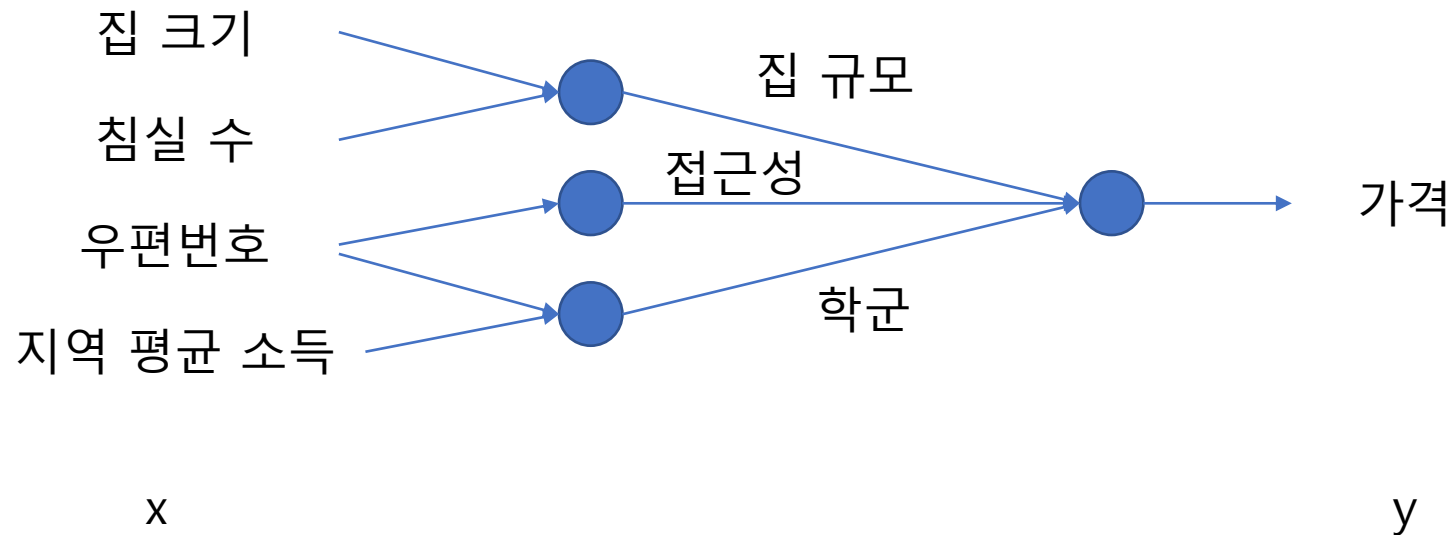
아주 간단한 신경망 예시

- 집 크기로부터 집 가격을 예측하고자 할 때 신경망을 사용할 수 있다.



간단한 신경망 예시

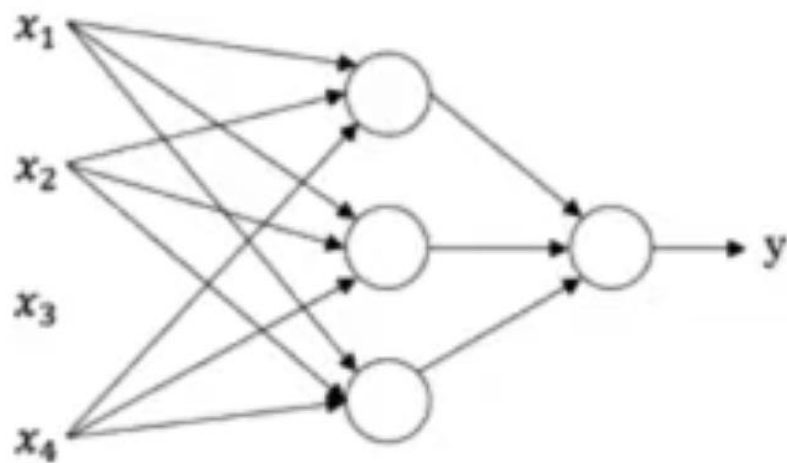
- 집과 관련한 여러 변수로부터 집 가격 예측하기
 - 아래와 같은 정보를 알아서 캐치하기를 기대



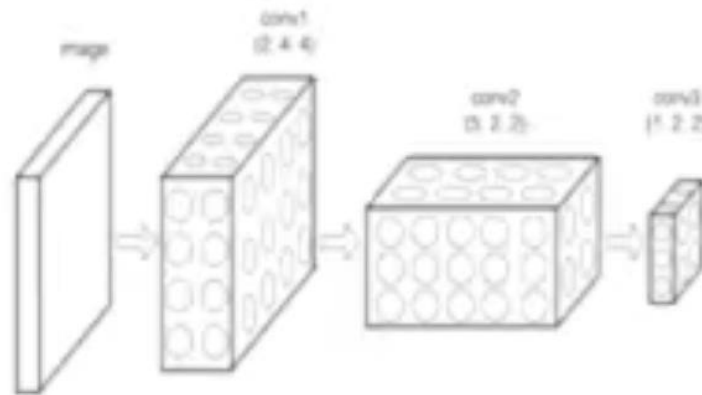
지도학습의 예시

입력 (x)	출력 (y)	적용	기법
집의 특성	가격	부동산	DNN
광고, 사용자 정보	광고 클릭 여부 (0/1)	온라인 광고	DNN
사진	사물 (1~1000)	사물인식	CNN
음성	대본	음성인식	RNN
영어	중국어	기계번역	RNN
사진 및 부가정보	다른 자동차의 위치	자율주행	복합

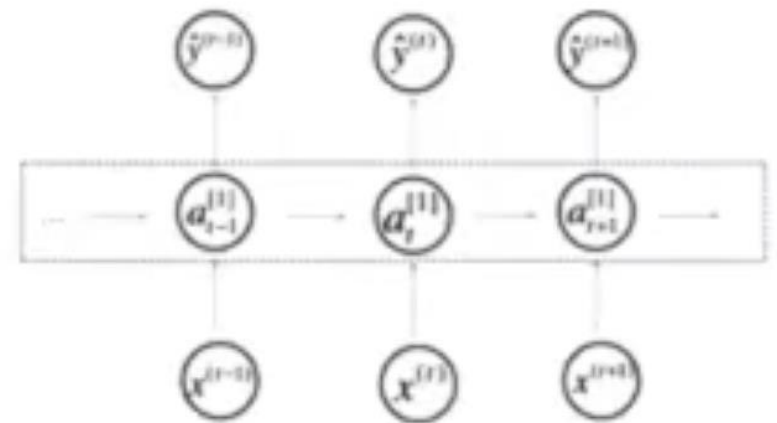
신경망 예시



Standard NN



Convolutional NN



Recurrent NN

구조화된 데이터

Structured Data

Size	#bedrooms	...	Price (1000\$s)
2104	3		400
1600	3		330
2400	3		369
!	!		!
3000	4		540

User Age	Ad Id	...	Click
41	93242		1
80	93287		0
18	87312		1

Unstructured Data



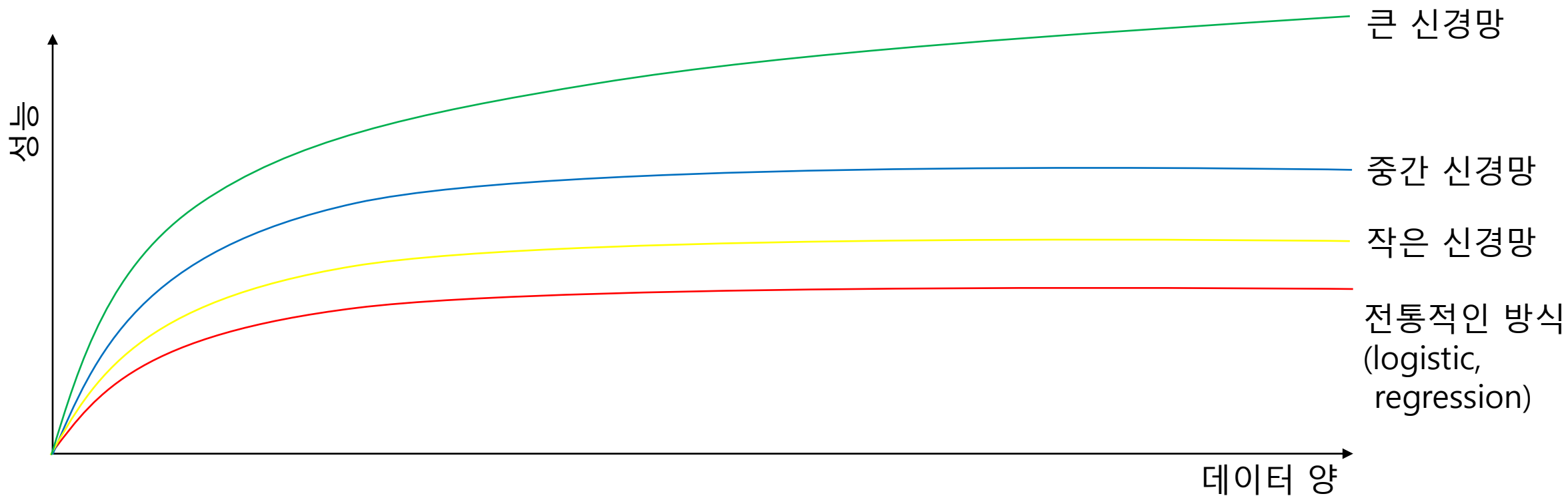
Audio



Image

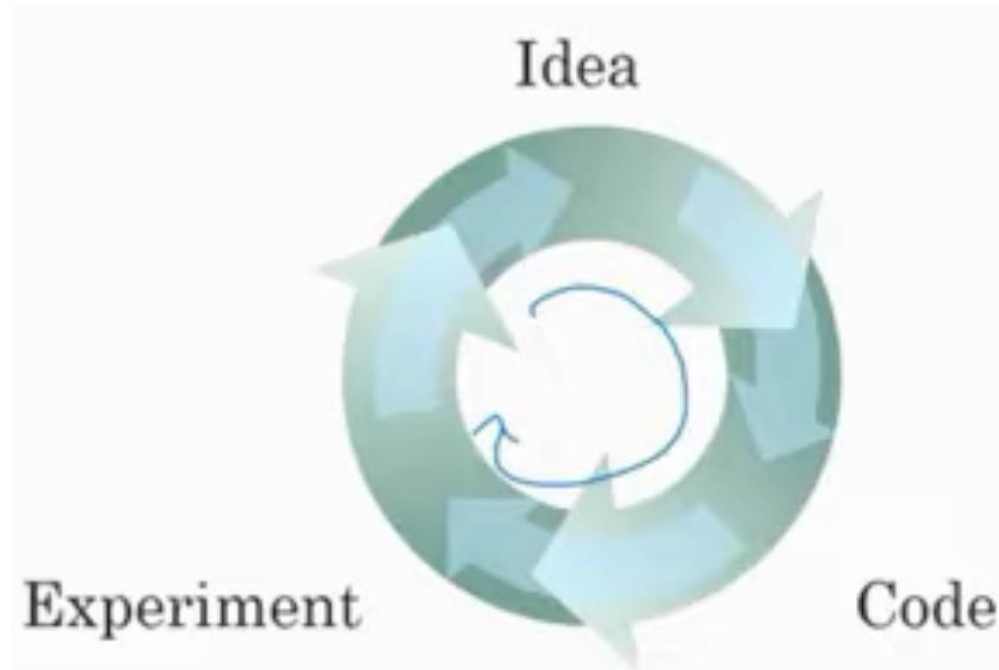
Four scores and seven
years ago...

딥러닝이 뜨고 있는 이유



딥러닝이 뜨고 있는 이유

- 데이터 양, 계산 속도 향상이 딥러닝을 더 강력하게 만들고 있다.
- 계산 속도가 빨라지면서 하이퍼 파라미터 튜닝도 용이해졌다.



이진 분류 문제

- 주어진 사진이 고양이인지 (1) 아닌지 (0) 맞추는 문제



1 (cat) vs 0 (non cat)

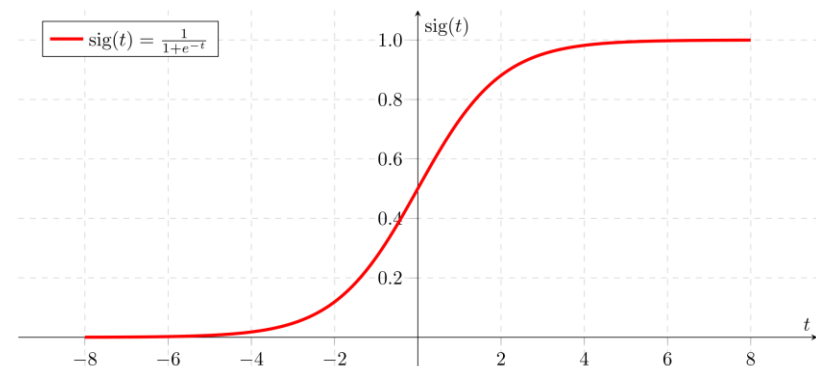
y

- 입력은 64 (가로) X 64 (세로) X 3 (color) = 12288 개의 값
- 출력은 0 또는 1

표기법

- 데이터: (x, y) . $x \in \mathbb{R}^{n_x}, y \in \{0, 1\}$
- m : 훈련 데이터 수. $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})\}$
 - 경우에 따라 m_{train}, m_{test} 등으로 표기하기도
- 데이터 행렬 $X = \begin{pmatrix} | & & | \\ x^{(1)} & \dots & x^{(m)} \\ | & & | \end{pmatrix}$. $X \in \mathbb{R}^{n_x \times m}$
- $Y = (y^{(1)}, \dots, y^{(m)})$. $Y \in \mathbb{R}^{1 \times m}$

Logistic Regression



- Given x , want $\hat{y} = P(y = 1|x)$
- 매개변수 $w \in \mathbb{R}^{n_x}$, $b \in \mathbb{R}$
- Logistic regression: $\hat{y} = \sigma(w^T x + b) \equiv \sigma(z)$
 - 시그모이드 함수: $\sigma(z) = \frac{1}{1+e^{-z}}$
 - (참고) $\theta = \begin{bmatrix} \theta_0 = b \\ \theta_1 = w_1 \\ \vdots \\ \theta_{n_x} = w_{n_x} \end{bmatrix}$, $x_0 = 1$, $\hat{y} = \sigma(\theta^T x)$ 로 표기하는 문헌도 있음.

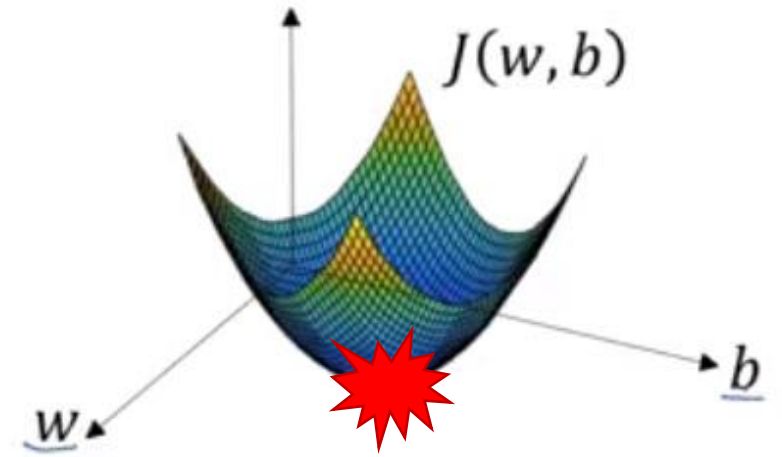
Logistic Regression 비용함수

- Loss (error) function $\mathcal{L}(\hat{y}, y)$ 를 먼저 정의하자.
- $\mathcal{L}(\hat{y}, y) = \frac{1}{2} (\hat{y} - y)^2$ 을 사용하면?
 - w, b 에 대해 볼록하지 못함 -> local minimum에 갇힐 수 있음.
- $\mathcal{L}(\hat{y}, y) = -y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})$
 - Crossentropy 함수. Logistic regression에서 자주 사용.
- 비용함수는 아래와 같이 정의

$$J(w, b) = -\frac{1}{m} \sum_{i=1}^m (y^{(i)} \log \hat{y}^{(i)} + (1 - y^{(i)}) \log (1 - \hat{y}^{(i)}))$$

경사하강법

- w 와 b 를 학습하는 알고리즘. $J(w, b)$ 를 최소로 하는 w 와 b 를 구한다.



경사하강법

- Repeat

- $w := w - \alpha \frac{\partial J}{\partial w} = w - \alpha dw$

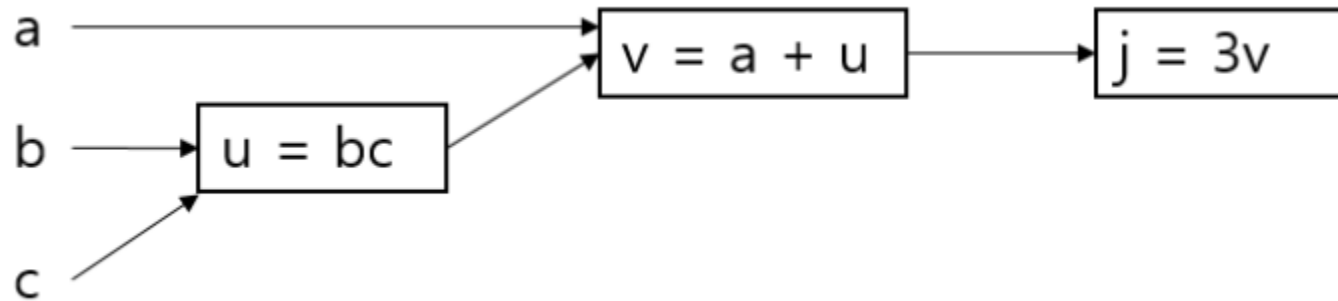
- $b := b - \alpha \frac{\partial J}{\partial b} = b - \alpha db$

- α 는 학습률 learning rate

- dw, db 는 코드 내에서 미분값을 표기할 때 자주 사용되는 명명법

계산 그래프와 미분의 연쇄법칙

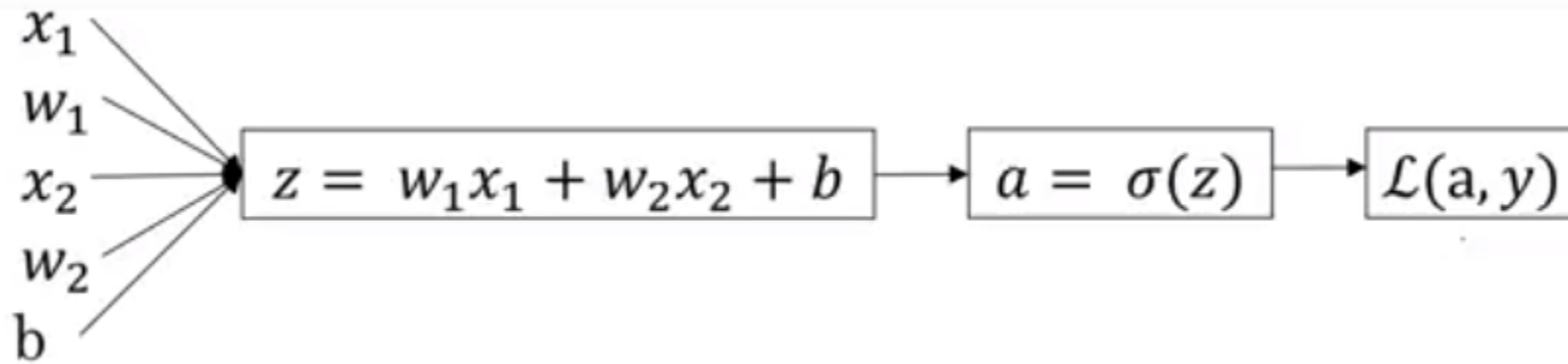
- $J(a, b, c) = 3(a + bc)$ 를 계산그래프로 나타내면 아래 그림과 같다.



- $\frac{dJ}{db} = \frac{dJ}{dv} \times \frac{dv}{du} \times \frac{du}{db}$ 가 된다.
- 이처럼 복잡한 네트워크라도 한 단계씩 거슬러 올라가면서 미분하면 $\frac{dJ}{dw} = dw$ 를 구할 수 있다.

Logistic Regression에서의 경사하강법

- Logistic regression에서의 계산그래프



- $da = -\frac{y}{a} + \frac{1-y}{1-a}$, $dz = a - y$, $dw_1 = x_1 dz$, $dw_2 = x_2 dz$, $db = dz$

m 개 샘플에서의 경사하강법

- $J = 0; dw_1 = 0; dw_2 = 0; db = 0$
- For $i = 1$ to m :
 - $z^{(i)} = w^T x^{(i)} + b$
 - $a^{(i)} = \sigma(z^{(i)})$
 - $J += -[y^{(i)} \log a^{(i)} + (1 - y^{(i)}) \log(1 - a^{(i)})]$
 - $dz^{(i)} = a^{(i)} - y^{(i)}$
 - $dw_1 += x_1^{(i)} dz^{(i)}$
 - $dw_2 += x_2^{(i)} dz^{(i)}$
 - $db += dz^{(i)}$
- $J /= m; dw_1 /= m; dw_2 /= m; db /= m$

다음 시간에는 vectorization을 이용해서
For문 없이, 더 빠르게 동작하는 코드를
소개할 예정.