

Linear Regression & Neural Network

2017.07.22

최건호

01

Linear Regression

- 정의
- Loss
- Gradient Descent
- 문제점

02

Neural Network

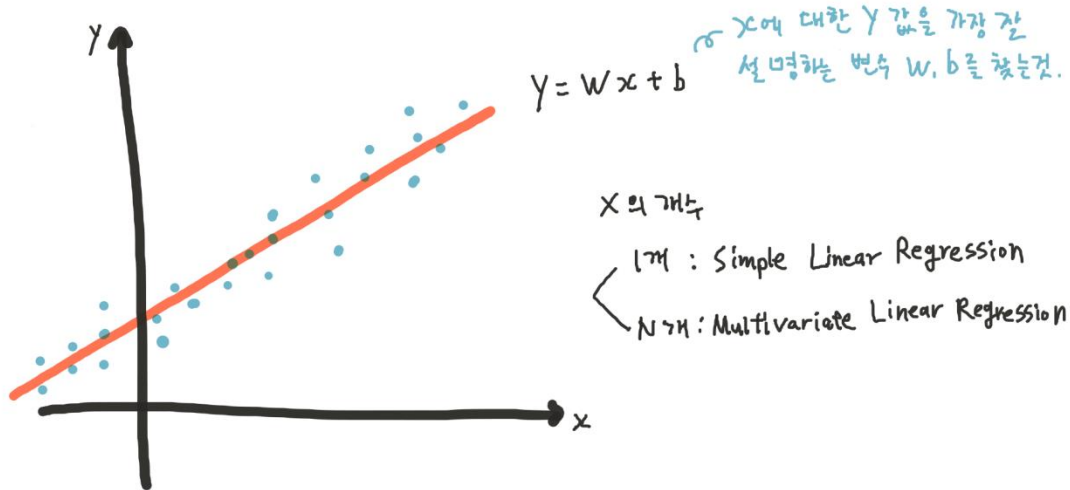
- 정의
- Deep?
- 활용

03

Propagation

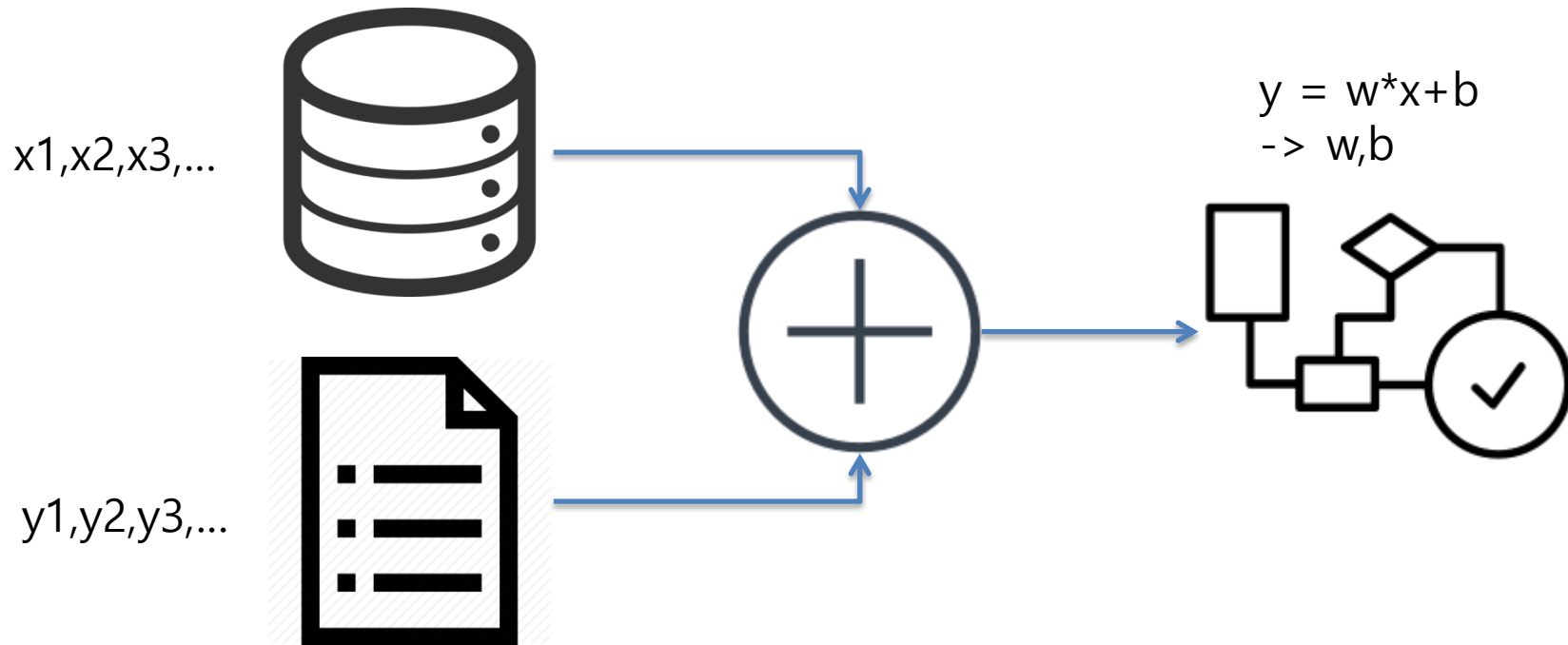
- 정의
 - Forward Prop.
 - Back Prop.
-

Linear Regression

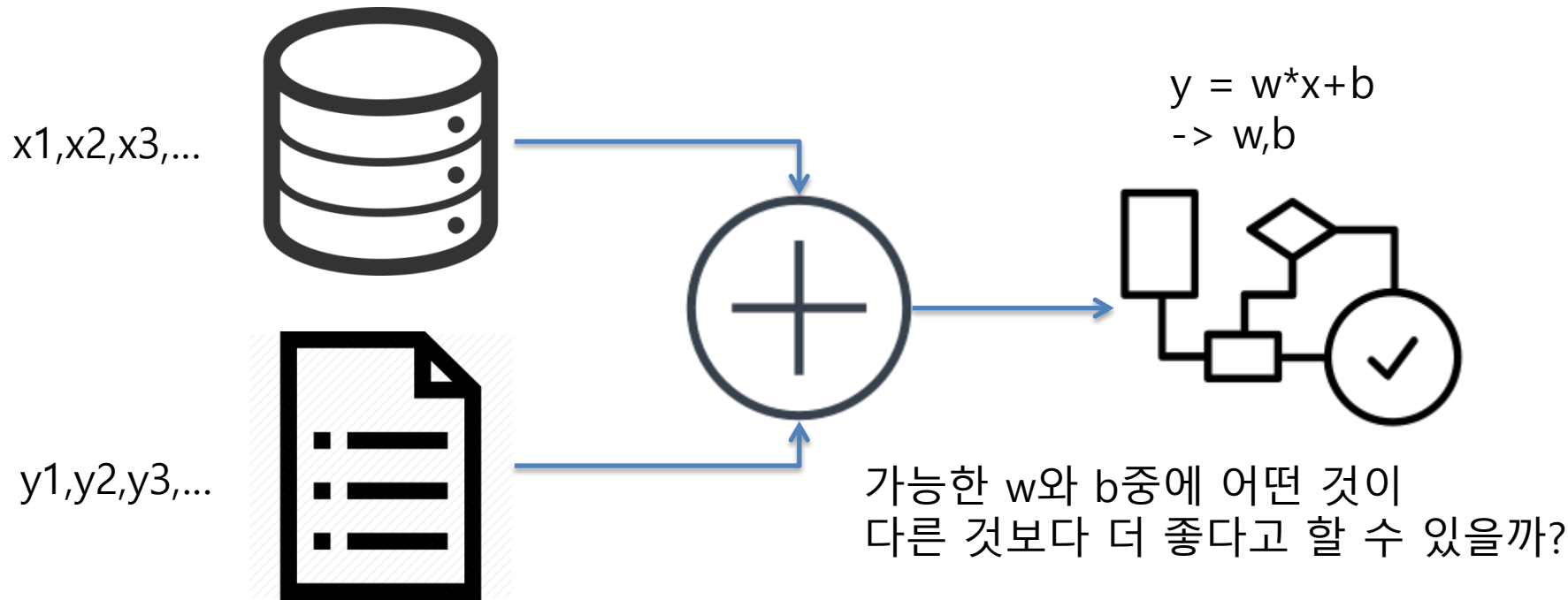


종속 변수 y 와 한 개 이상의 독립 변수 x 와의
선형 상관 관계를 모델링하는 회귀분석 기법
(출처: 위키피디아)

Linear Regression



Linear Regression

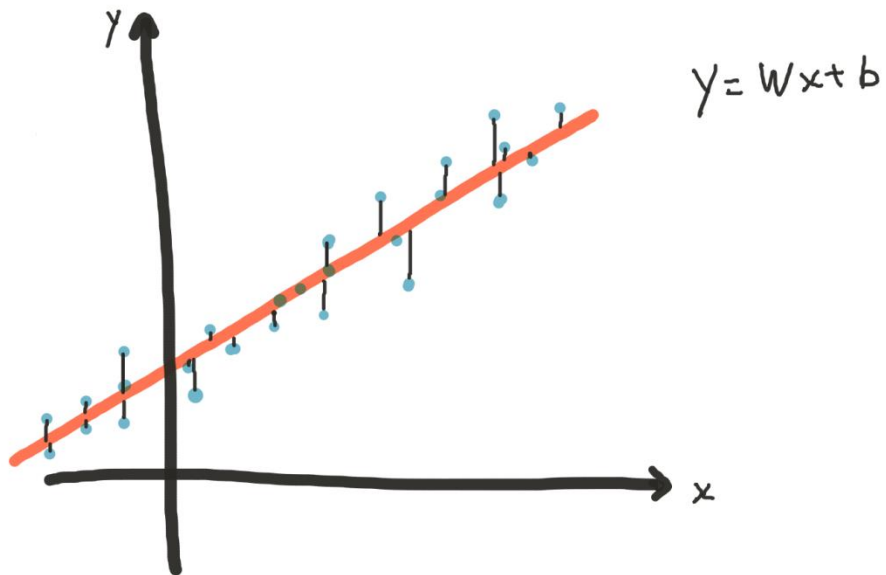


Linear Regression

잘 예측했는지 아닌지 측정할 척도(metric)가 필요함

Linear Regression

잘 예측했는지 아닌지 측정할 척도(metric)가 필요함



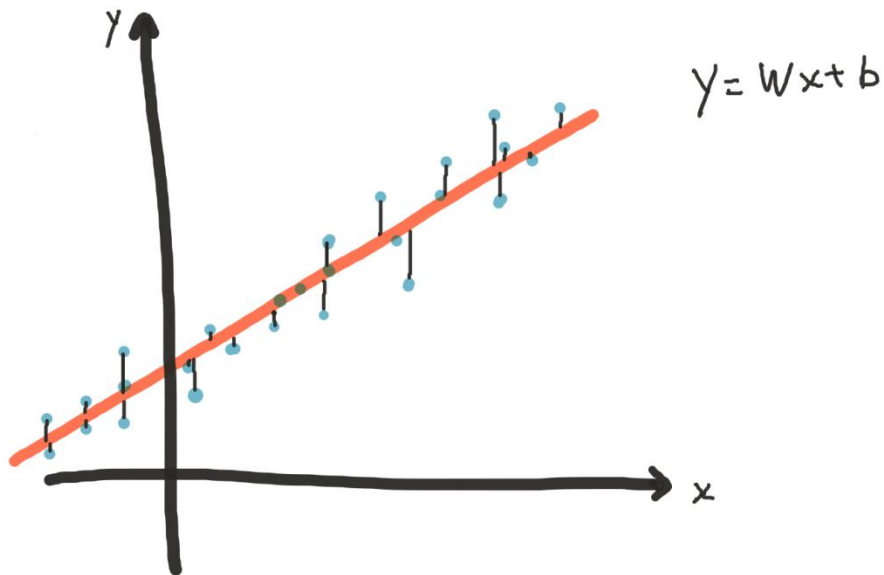
Linear Regression

잘 예측했는지 아닌지 측정할 척도(metric)가 필요함

Mean Squared Error(MSE)

$$\text{MSE} = (x_1 - x_2)^2$$

두 값의 거리의 제곱



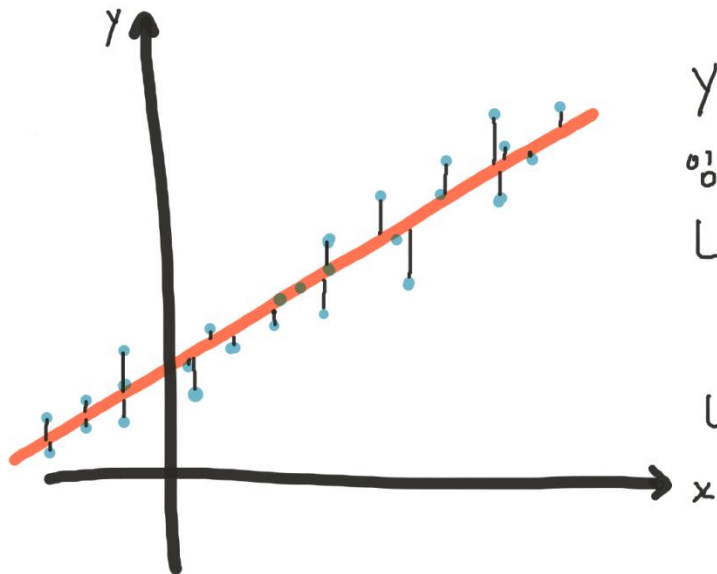
Linear Regression

잘 예측했는지 아닌지 측정할 척도(metric)가 필요함

Mean Squared Error(MSE)

$$\text{MSE} = (x_1 - x_2)^2$$

두 값의 거리의 제곱



$$y = wx + b$$

임의의 w^*, b^* 를 초기 값으로 한다면

$$\text{Loss} = (\underbrace{y^*}_{\text{예측}} - \underbrace{y}_{\text{실제}})^2 = (wx + b - y)^2$$

Loss 함수는 고정된 x, y 에서 w^*, b^* 에 의해 정해짐

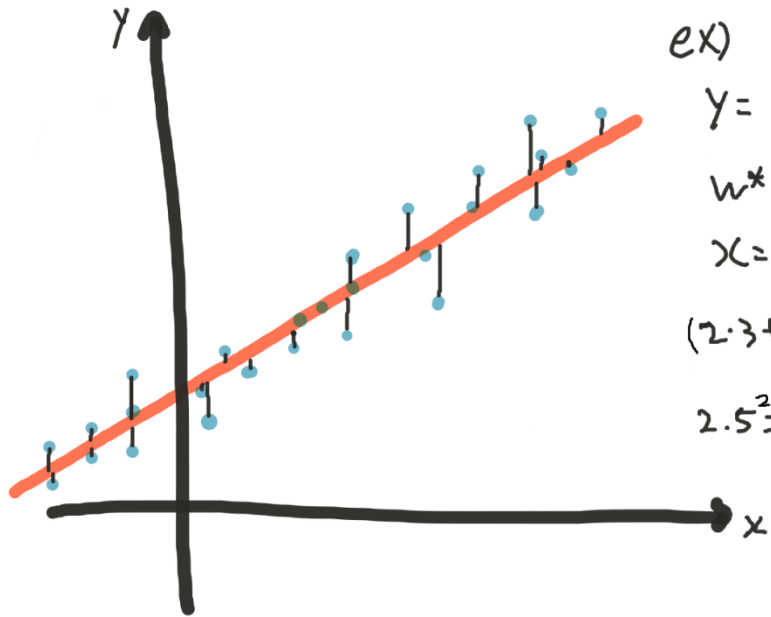
Linear Regression

잘 예측했는지 아닌지 측정할 척도(metric)가 필요함

Mean Squared Error(MSE)

$$\text{MSE} = (x_1 - x_2)^2$$

두 값의 거리의 제곱



ex)

$$y = 0.5x + 4 \text{ 이고}$$

$$w^* = 2, b^* = 2 \text{ 일때,}$$

$x = 3$ 에서 Loss 는

$$(2 \cdot 3 + 2 - (1.5 + 4))^2 = 2.5^2$$

$$2.5^2 = (\underbrace{w^* \cdot 3 + b^*}_{y^*} - y)^2$$

Linear Regression

Loss를 minimize하는 w, b 를 구하고 싶다.



Linear Regression

Loss를 minimize하는 w, b 를 구하고 싶다.

-> Random Search?

Linear Regression

Loss를 minimize하는 w, b 를 구하고 싶다.

-> Random Search? 어느 세월에..

Linear Regression

Loss를 minimize하는 w, b 를 구하고 싶다.

- > Random Search? 어느 세월에..
- > Loss 값을 통해서 구할 수 있을까?

Linear Regression

Loss를 minimize하는 w, b 를 구하고 싶다.

-> Random Search? 어느 세월에..

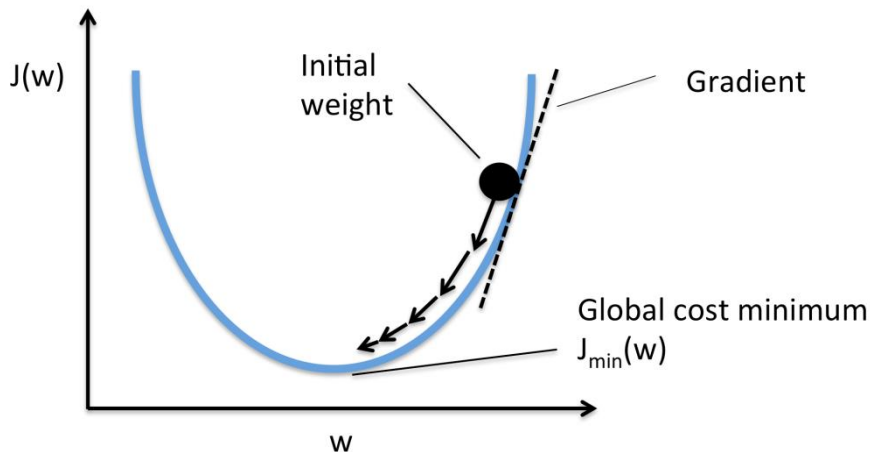
-> Loss 값을 통해서 구할 수 없을까? Gradient Descent!

Linear Regression

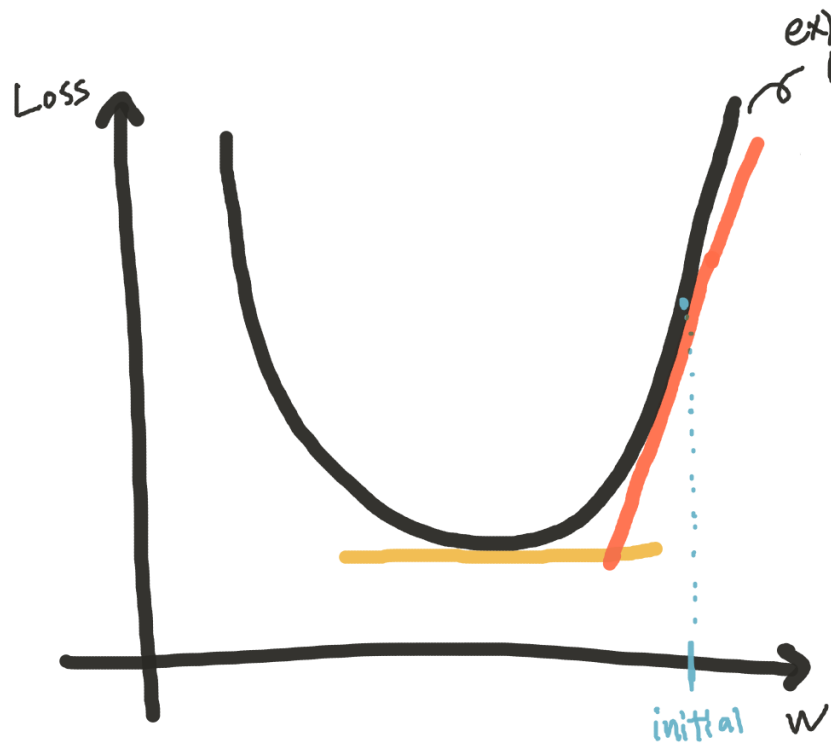
Loss를 minimize하는 w, b 를 구하고 싶다.

-> Random Search? 어느 세월에..

-> Loss 값을 통해서 구할 수 없을까? Gradient Descent!



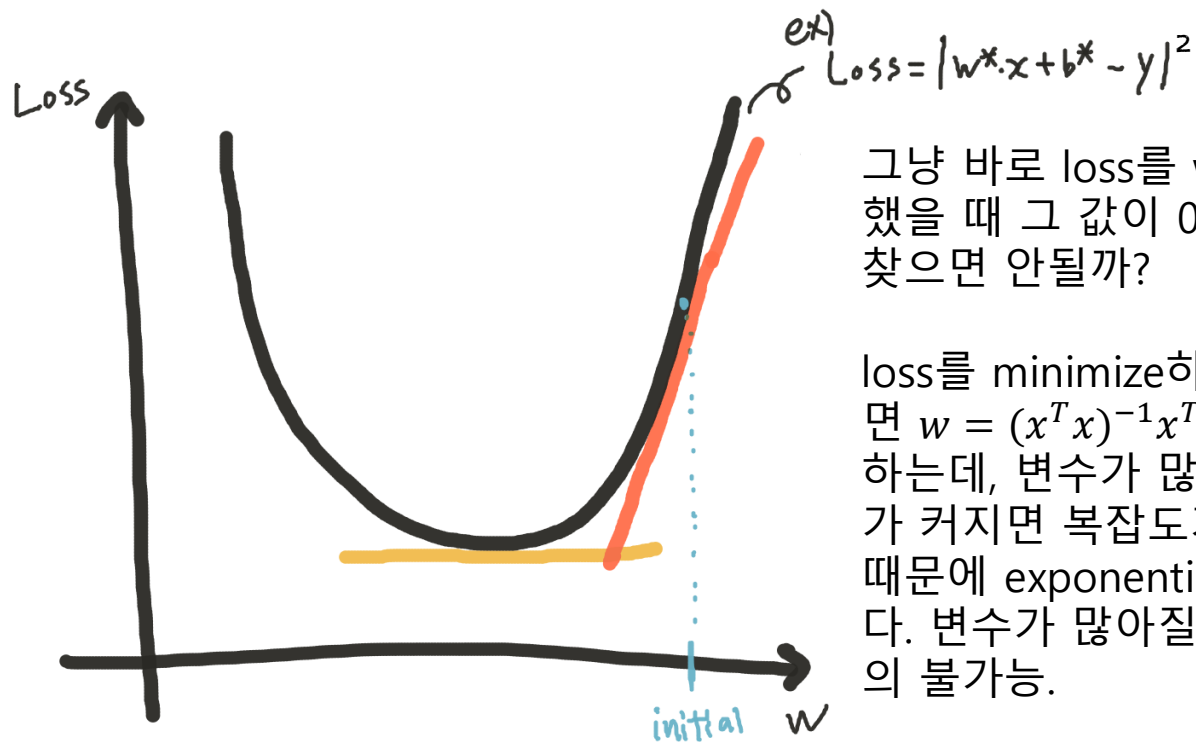
Linear Regression



ex) $Loss = |w^* \cdot x + b^* - y|^2$

그냥 바로 loss를 w 에 대해 미분
했을 때 그 값이 0이 되는 w 를
찾으면 안될까?

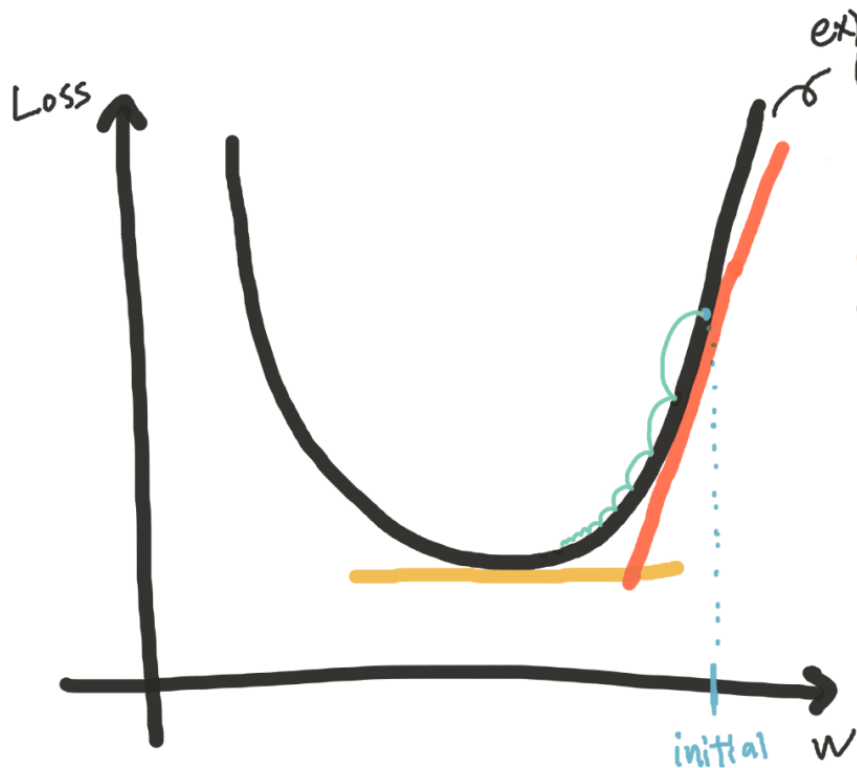
Linear Regression



그냥 바로 loss를 w 에 대해 미분했을 때 그 값이 0이 되는 w 를 찾으면 안될까?

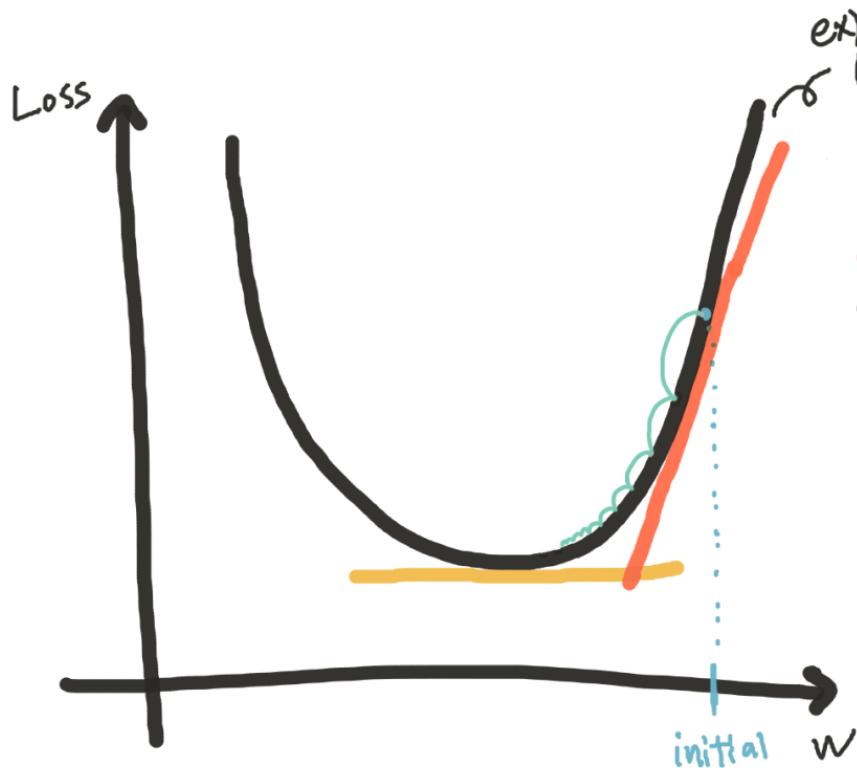
loss를 minimize하는 w 를 구하려면 $w = (x^T x)^{-1} x^T y$ 식을 풀어야 하는데, 변수가 많아지고 matrix가 커지면 복잡도가 $O(n^3)$ 이기 때문에 exponential 하게 증가한다. 변수가 많아질수록 계산이 거의 불가능.

Linear Regression



바로 minimum이 되는 지점을 구하기보다, 현재 loss에 대한 w 의 **gradient(경사도)**를 구하여 gradient x learning rate만큼 w 를 업데이트

Linear Regression

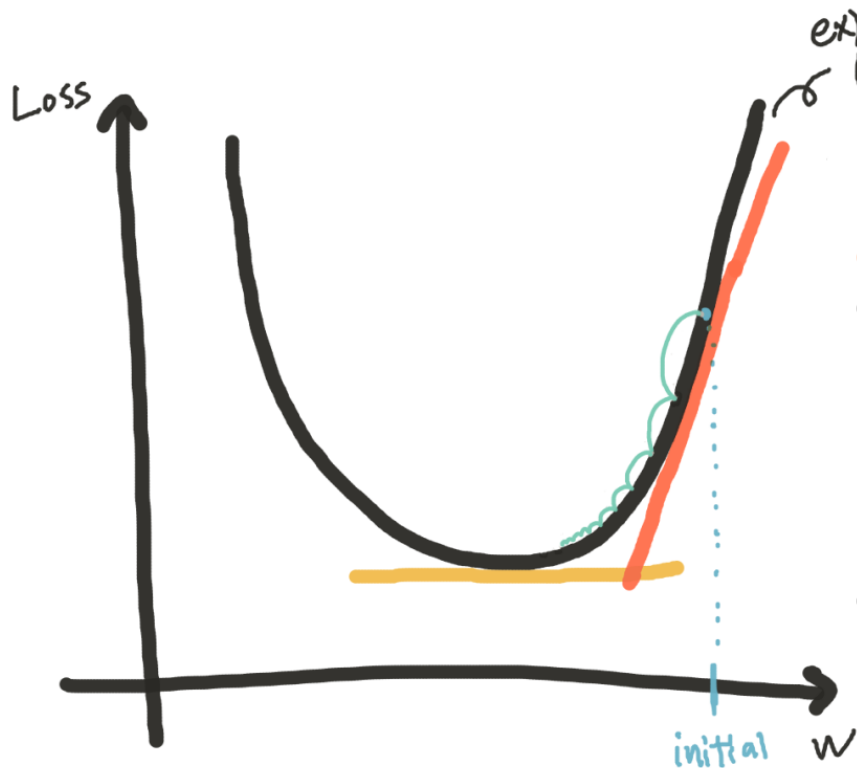


ex) $Loss = |w^* \cdot x + b^* - y|^2$

바로 minimum이 되는 지점을 구하기보다, 현재 loss에 대한 w 의 **gradient(경사도)**를 구하여 $gradient \times learning\ rate$ 만큼 w 를 업데이트

$$w_{t+1} = w_t - \underline{gradient} \times \underline{learning\ rate}$$

Linear Regression



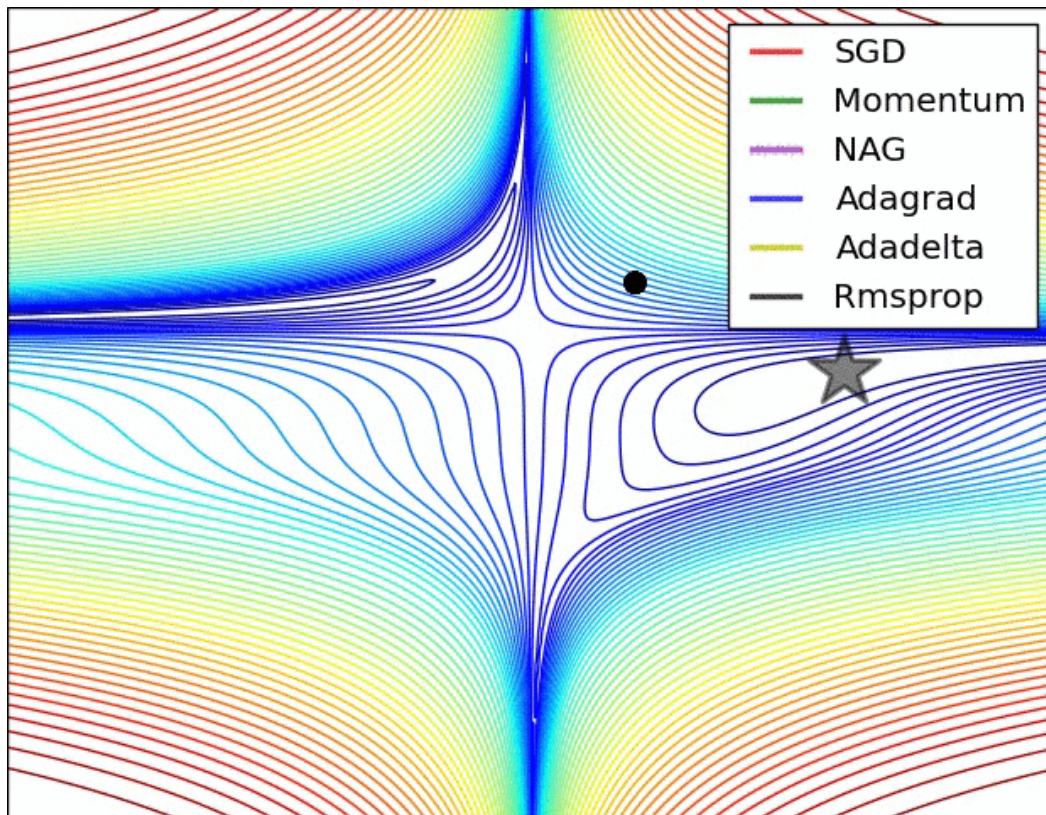
ex) $Loss = |w^* \cdot x + b^* - y|^2$

바로 minimum이 되는 지점을 구하기보다, 현재 loss에 대한 w 의 **gradient(경사도)**를 구하여 $gradient \times learning\ rate$ 만큼 w 를 업데이트

$$w_{k+1} = w_k - \underline{gradient} \times \underline{learning\ rate}$$

계속 업데이트 하다 보면 optimal한 w 에 근접하게 됨.

Linear Regression



Linear Regression

```
linear_regression.py x
1 import numpy as np
2 import torch
3 import torch.nn as nn
4 import torch.optim as optim
5 import torch.nn.init as init
6 from torch.autograd import Variable
7 from visdom import Visdom
8 viz = Visdom()
9
10 # data generation
11
12 num_data = 1000
13
14 noise = init.normal(torch.FloatTensor(num_data,1),std=0.2)
15 x = init.uniform(torch.Tensor(num_data,1),-10,10)
16 y = 2*x+3
17 y_noise = 2*(x+noise)+3
```

Linear Regression

필요한 라이브러리

```
linear_regression.py x
1 import numpy as np
2 import torch
3 import torch.nn as nn
4 import torch.optim as optim
5 import torch.nn.init as init
6 from torch.autograd import Variable
7 from visdom import Visdom
8 viz = Visdom()
9
10 # data generation
11
12 num_data = 1000
13
14 noise = init.normal(torch.FloatTensor(num_data,1),std=0.2)
15 x = init.uniform(torch.Tensor(num_data,1),-10,10)
16 y = 2*x+3
17 y_noise = 2*(x+noise)+3
```


Linear Regression

필요한 라이브러리

```
1 import numpy as np
2 import torch
3 import torch.nn as nn
4 import torch.optim as optim
5 import torch.nn.init as init
6 from torch.autograd import Variable
7 from visdom import Visdom
8 viz = Visdom()
9
```

데이터 생성

```
10 # data generation
11
12 num_data = 1000
13
14 noise = init.normal(torch.FloatTensor(num_data,1),std=0.2)
15 x = init.uniform(torch.Tensor(num_data,1),-10,10)
16 y = 2*x+3
17 y_noise = 2*(x+noise)+3
```

Linear Regression

```
41 # model & optimizer
42
43 model = nn.Linear(1,1)
44 output = model(Variable(x))
45
46 loss_func = nn.L2Loss()
47 optimizer = optim.SGD(model.parameters(), lr=1)
48
49 # train
50 loss_arr = []
51 label = Variable(y_noise)
52 for i in range(1000):
53     output = model(Variable(x))
54     optimizer.zero_grad()
55
56     loss = loss_func(output, label)
57     loss.backward()
58     optimizer.step()
59     #print(loss)
60     loss_arr.append(loss.data.numpy()[0])
61
62
63 param_list = list(model.parameters())
64 print(param_list[0].data, param_list[1].data)
```

Linear Regression

linear regression 모델 생성

```
41 # model & optimizer
42
43 model = nn.Linear(1,1)
44 output = model(Variable(x))
45
46 loss_func = nn.L2Loss()
47 optimizer = optim.SGD(model.parameters(), lr=1)
48
49 # train
50 loss_arr = []
51 label = Variable(y_noise)
52 for i in range(1000):
53     output = model(Variable(x))
54     optimizer.zero_grad()
55
56     loss = loss_func(output, label)
57     loss.backward()
58     optimizer.step()
59     #print(loss)
60     loss_arr.append(loss.data.numpy()[0])
61
62
63 param_list = list(model.parameters())
64 print(param_list[0].data, param_list[1].data)
```

Linear Regression

linear regression 모델 생성

loss function 및
gradient descent
optimizer 생성

```
41 # model & optimizer
42
43 model = nn.Linear(1,1)
44 output = model(Variable(x))
45
46 loss_func = nn.L2Loss()
47 optimizer = optim.SGD(model.parameters(), lr=1)
48
49 # train
50 loss_arr = []
51 label = Variable(y_noise)
52 for i in range(1000):
53     output = model(Variable(x))
54     optimizer.zero_grad()
55
56     loss = loss_func(output, label)
57     loss.backward()
58     optimizer.step()
59     #print(loss)
60     loss_arr.append(loss.data.numpy()[0])
61
62
63 param_list = list(model.parameters())
64 print(param_list[0].data, param_list[1].data)
```

Linear Regression

linear regression 모델 생성

loss function 및
gradient descent
optimizer 생성

<training 단계>

1. 모델로 결과값 추정
2. loss 및 gradient 계산
3. 모델 업데이트

```
41 # model & optimizer
42
43 model = nn.Linear(1,1)
44 output = model(Variable(x))
45
46 loss_func = nn.L2Loss()
47 optimizer = optim.SGD(model.parameters(), lr=1)
48
49 # train
50 loss_arr = []
51 label = Variable(y_noise)
52 for i in range(1000):
53     output = model(Variable(x))
54     optimizer.zero_grad()
55
56     loss = loss_func(output, label)
57     loss.backward()
58     optimizer.step()
59     #print(loss)
60     loss_arr.append(loss.data.numpy()[0])
61
62
63 param_list = list(model.parameters())
64 print(param_list[0].data, param_list[1].data)
```

Linear Regression

linear regression 모델 생성

loss function 및
gradient descent
optimizer 생성

<training 단계>

1. 모델로 결과값 추정
2. loss 및 gradient 계산
3. 모델 업데이트

training 이후 파라미터 값 확인

```
41 # model & optimizer
42
43 model = nn.Linear(1,1)
44 output = model(Variable(x))
45
46 loss_func = nn.L2Loss()
47 optimizer = optim.SGD(model.parameters(), lr=1)
48
49 # train
50 loss_arr = []
51 label = Variable(y_noise)
52 for i in range(1000):
53     output = model(Variable(x))
54     optimizer.zero_grad()
55
56     loss = loss_func(output, label)
57     loss.backward()
58     optimizer.step()
59     #print(loss)
60     loss_arr.append(loss.data.numpy()[0])
61
62
63 param_list = list(model.parameters())
64 print(param_list[0].data, param_list[1].data)
```

Linear Regression

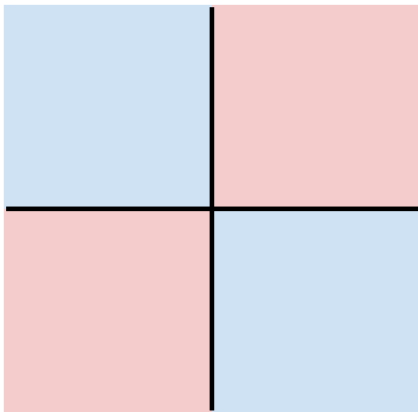
Hard cases for a linear classifier

Class 1:

number of pixels > 0 odd

Class 2:

number of pixels > 0 even

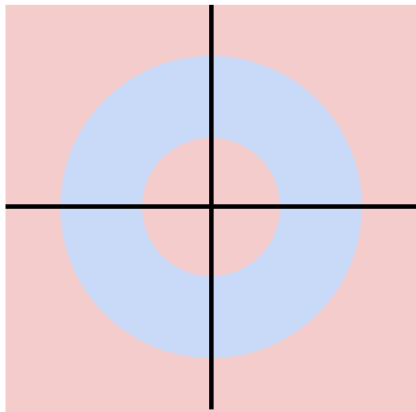


Class 1:

$1 \leq \text{L2 norm} \leq 2$

Class 2:

Everything else

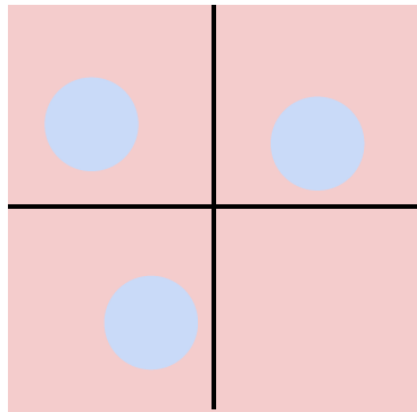


Class 1:

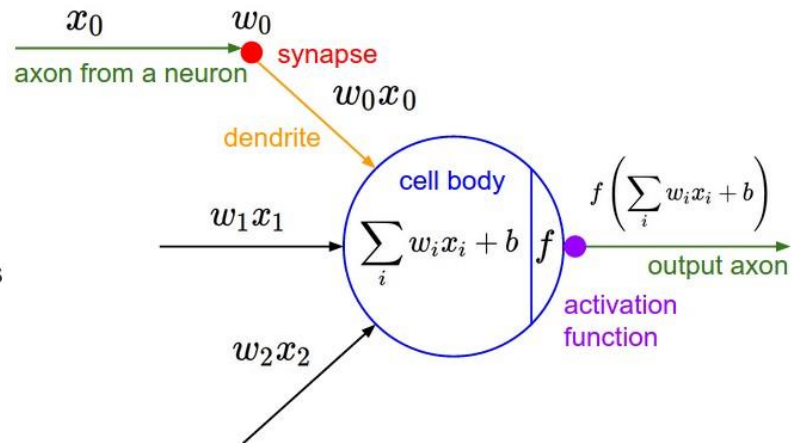
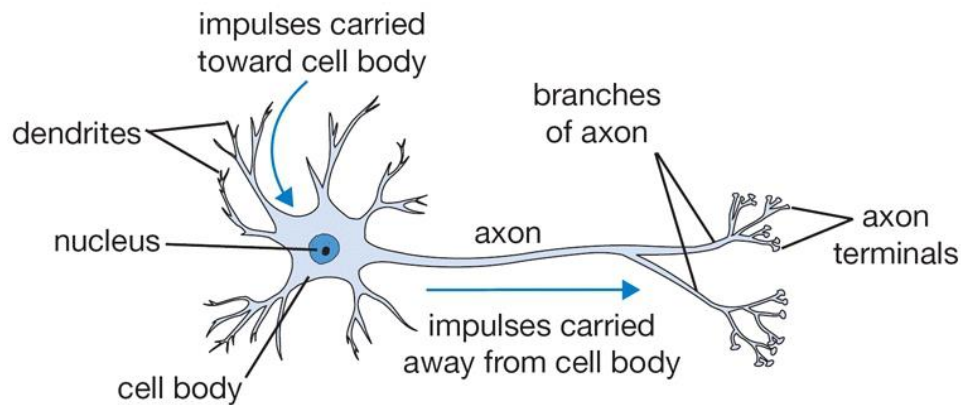
Three modes

Class 2:

Everything else



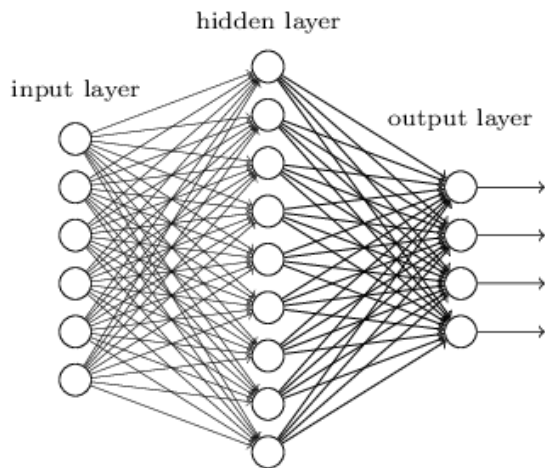
Neural Network



여러 자극이 들어오고 일정 기준을 넘으면 이를 다른 뉴런에 전달하는 구조

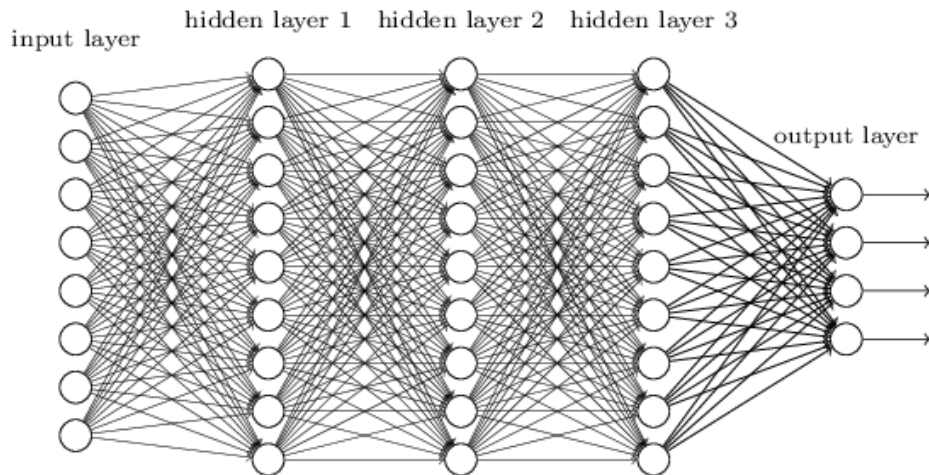
Neural Network

"Non-deep" feedforward
neural network



$$y = w2(act(w1 * input + b1)) + b2$$

Deep neural network



$$y = w4(act(w3(act(w2(act(w1 * input + b1)) + b2)) + b3)) + b4$$

Neural Network

$$Y = W \cdot x + b$$

$$= \begin{bmatrix} x_{00} & x_{01} & \dots & \dots & \dots \\ x_{10} & x_{11} & & & \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ x_{mn} & & & & \end{bmatrix} \begin{bmatrix} w_{00} & w_{01} & w_{02} & \dots & \dots \\ w_{10} & w_{11} & & & \\ w_{20} & & & & \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ w_{nl} & & & & \end{bmatrix} + \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ \vdots \\ \vdots \\ b_m \end{bmatrix}$$

$m \times n$ $n \times l$ $m \times 1$

Neural Network

$$Y = W \cdot X + b$$

Diagram illustrating the matrix representation of the equation $Y = W \cdot X + b$.

The input matrix X (labeled $m \times n$) is shown as a large bracketed matrix. Its first row is highlighted with an orange box and labeled x in green. The elements are $x_{00}, x_{01}, \dots, x_{0n}$. The first column is labeled $x_{10}, x_{11}, \dots, x_{1n}$.

The weight matrix W (labeled $n \times l$) is shown as a large bracketed matrix. Its first column is highlighted with an orange box and labeled w in green. The elements are $w_{00}, w_{01}, w_{02}, \dots, w_{0l}$. The first row is labeled $w_{10}, w_{11}, \dots, w_{1l}$.

The bias vector b (labeled $m \times 1$) is shown as a large bracketed vector. Its first element is highlighted with an orange box and labeled b in green. The elements are b_0, b_1, \dots, b_m .

The equation is represented as:

$$= \begin{bmatrix} x_{00} & x_{01} & \dots & x_{0n} \\ x_{10} & x_{11} & \dots & x_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m0} & x_{m1} & \dots & x_{mn} \end{bmatrix} \begin{bmatrix} w_{00} & w_{01} & w_{02} & \dots & w_{0l} \\ w_{10} & w_{11} & \dots & \dots & w_{1l} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ w_{n0} & w_{n1} & \dots & \dots & w_{nl} \end{bmatrix} + \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_m \end{bmatrix}$$

Neural Network

$$y = \text{act}(wx + b)$$

$$= \text{activation} \left(\begin{bmatrix} wx + b \end{bmatrix} \right)$$

$m \times 1$

Neural Network

만약 activation function이 없다면 아래의 식은 결국 linear function.

$$y = w_4(\text{act}(w_3(\text{act}(w_2(\text{act}(w_1 * \text{input} + b_1)) + b_2)) + b_3)) + b_4$$

Neural Network

만약 activation function이 없다면 아래의 식은 결국 linear function.

$$y = w_4(\text{act}(w_3(\text{act}(w_2(\text{act}(w_1 * \text{input} + b_1)) + b_2)) + b_3)) + b_4$$

activation function으로 non-linearity를 추가해야 함

Neural Network

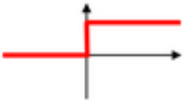
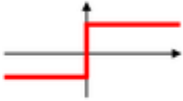

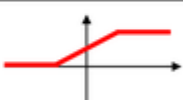


만약 activation function이 없다면 아래의 식은 결국 linear function.

$$y = w_4(\text{act}(w_3(\text{act}(w_2(\text{act}(w_1 * \text{input} + b_1)) + b_2)) + b_3)) + b_4$$

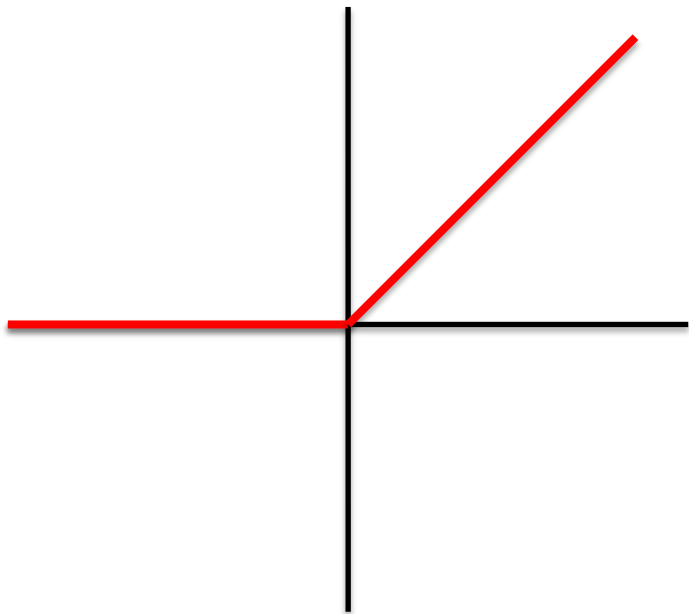
activation function으로 non-linearity를 추가해야 함

그렇다면 어떤 activation function을 써야 할까?

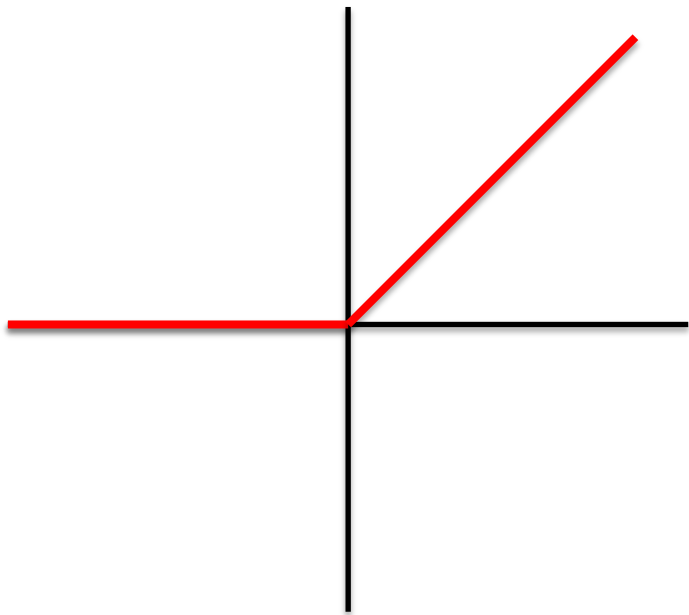
Neural Network

Activation function	Equation	Example	1D Graph
Unit step (Heaviside)	$\phi(z) = \begin{cases} 0, & z < 0, \\ 0.5, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron variant	
Sign (Signum)	$\phi(z) = \begin{cases} -1, & z < 0, \\ 0, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron variant	
Linear	$\phi(z) = z$	Adaline, linear regression	
Piece-wise linear	$\phi(z) = \begin{cases} 1, & z \geq \frac{1}{2}, \\ z + \frac{1}{2}, & -\frac{1}{2} < z < \frac{1}{2}, \\ 0, & z \leq -\frac{1}{2}, \end{cases}$	Support vector machine	
Logistic (sigmoid)	$\phi(z) = \frac{1}{1 + e^{-z}}$	Logistic regression, Multi-layer NN	
Hyperbolic tangent	$\phi(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$	Multi-layer NN	

Neural Network

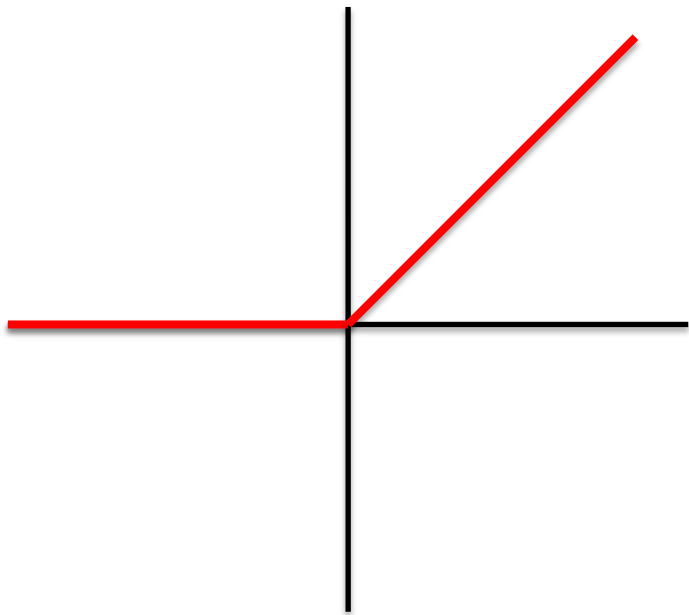


Neural Network



Rectified Linear Unit (ReLU)

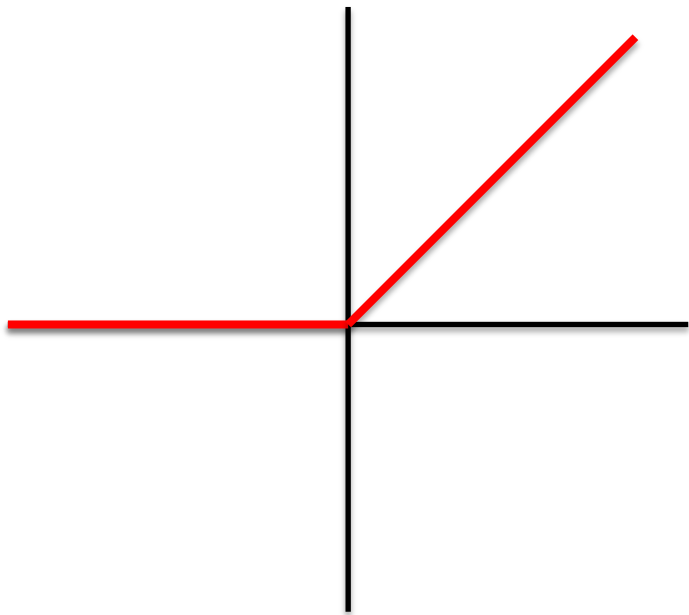
Neural Network



Rectified Linear Unit (ReLU)

$$f(x) = \max(0, x)$$

Neural Network

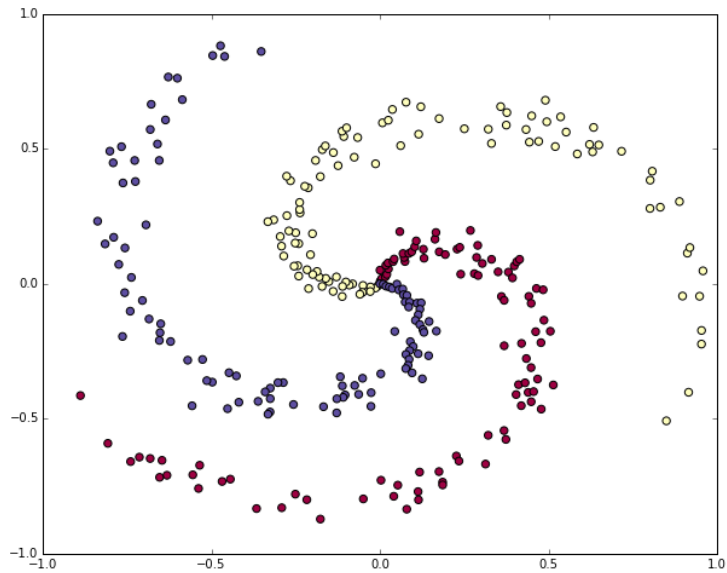


Rectified Linear Unit (ReLU)

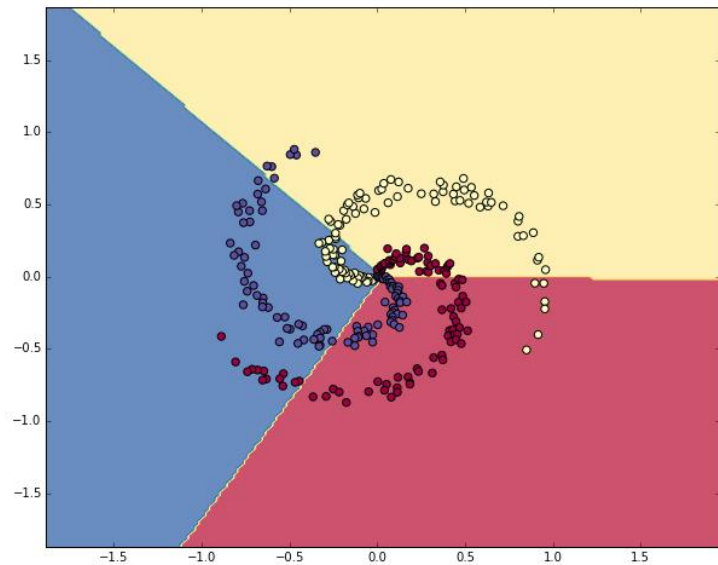
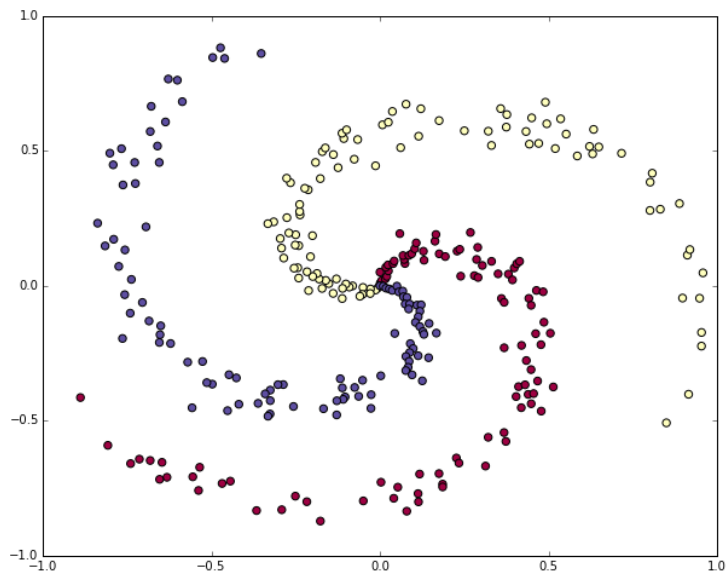
$$f(x) = \max(0, x)$$

기존의 sigmoid와 tanh로는
학습이 잘 안됐었는데 relu는
gradient의 전달이 좋아서
default로 사용되고 있음

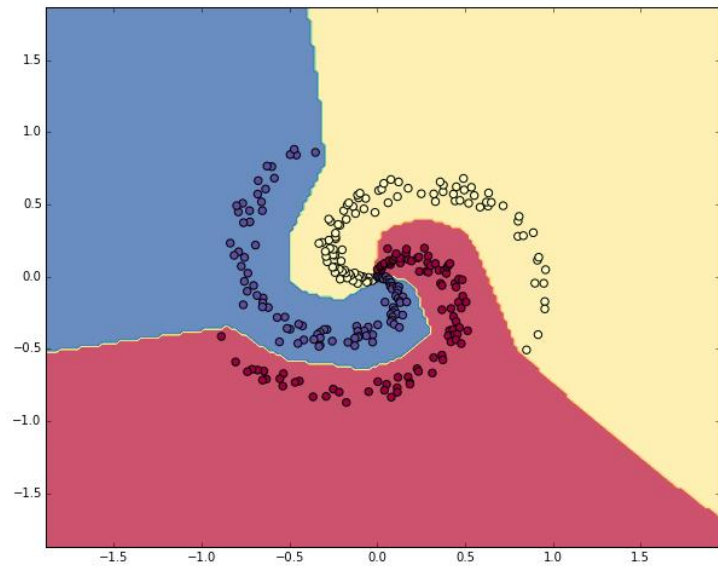
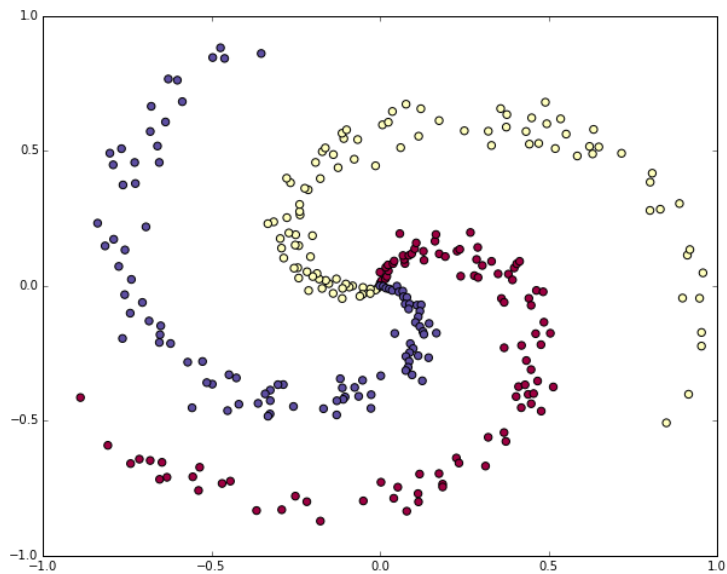
Neural Network



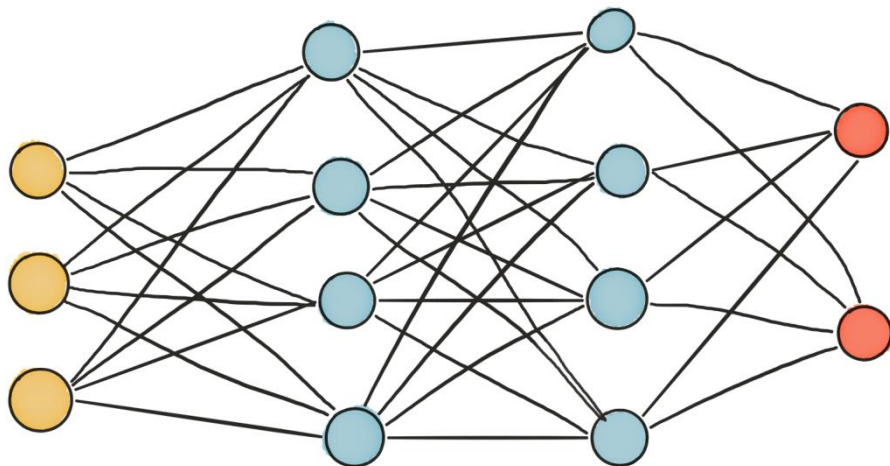
Neural Network



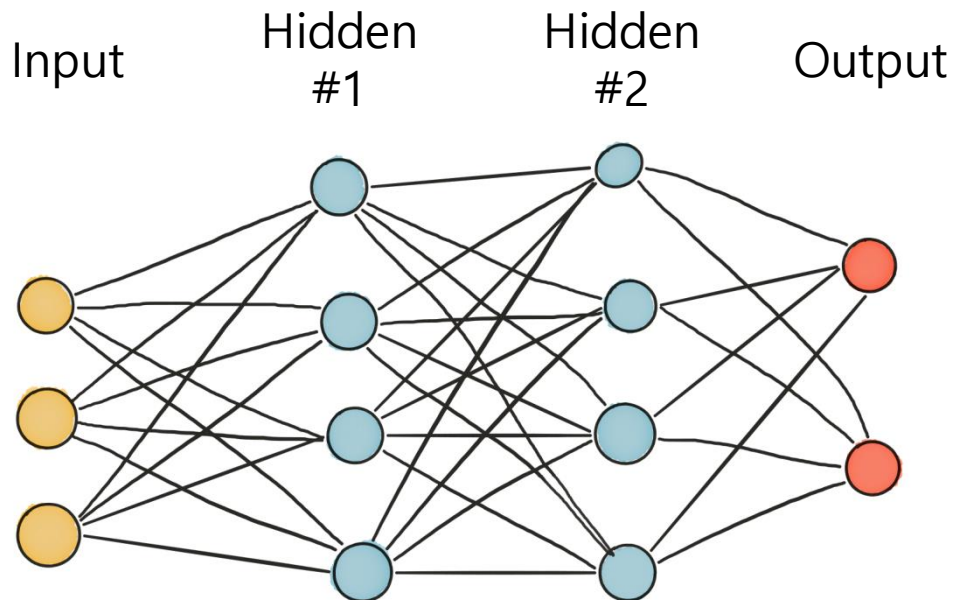
Neural Network



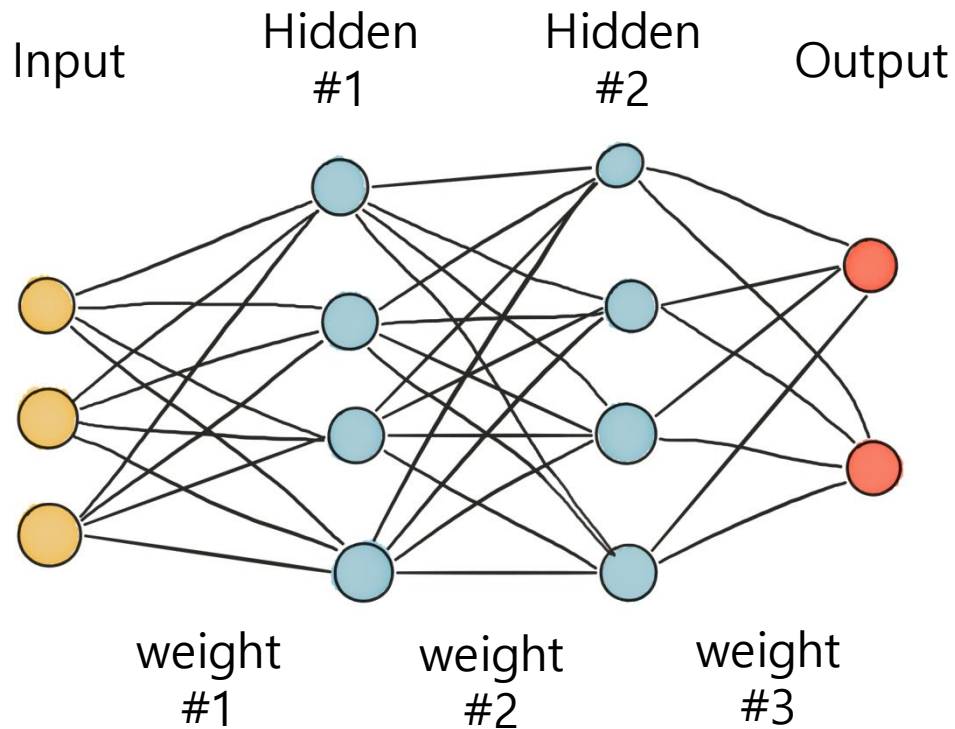
Forward & Back Prop.



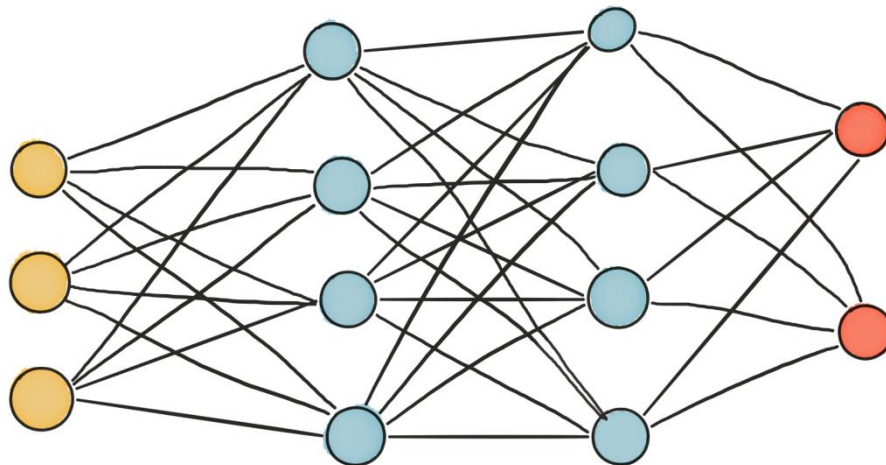
Forward & Back Prop.



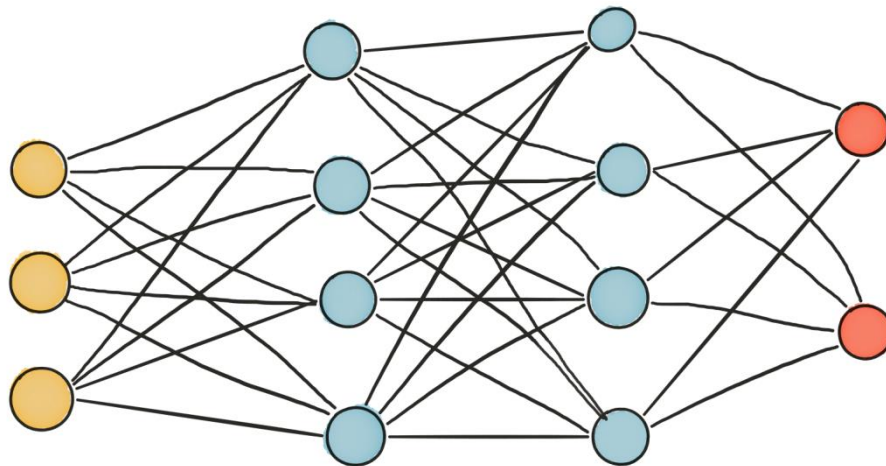
Forward & Back Prop.



Forward & Back Prop.

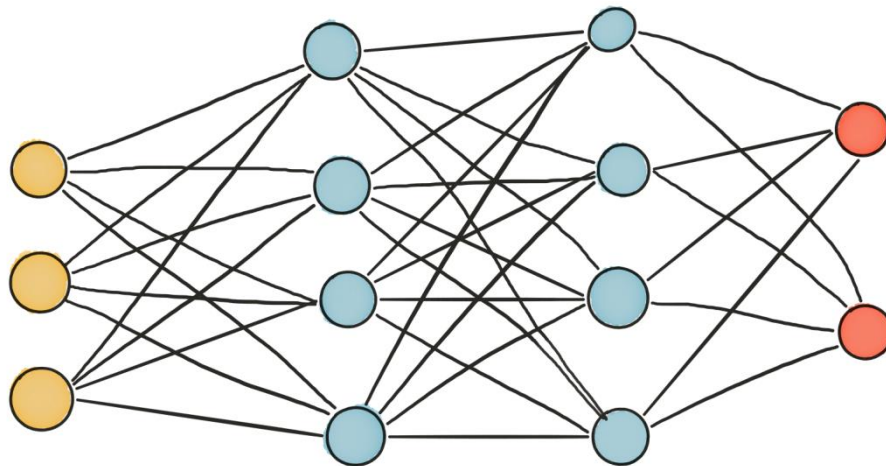


Forward & Back Prop.



$$\begin{bmatrix} w_{00} & w_{01} & w_{02} & w_{03} \\ w_{10} & w_{11} & w_{12} & w_{13} \\ w_{20} & w_{21} & w_{22} & w_{23} \end{bmatrix} \times \begin{bmatrix} w_{00} & w_{01} & w_{02} & w_{03} \\ w_{10} & w_{11} & w_{12} & w_{13} \\ w_{20} & w_{21} & w_{22} & w_{23} \\ w_{30} & w_{31} & w_{32} & w_{33} \end{bmatrix} \times \begin{bmatrix} w_{00} & w_{01} \\ w_{10} & w_{11} \\ w_{20} & w_{21} \\ w_{30} & w_{31} \end{bmatrix}$$

Forward & Back Prop.

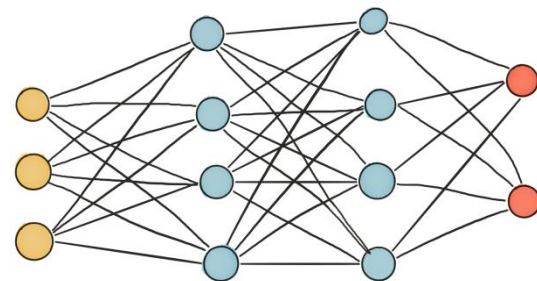


$$\begin{bmatrix} w_{00} & w_{01} & w_{02} & w_{03} \\ w_{10} & w_{11} & w_{12} & w_{13} \\ w_{20} & w_{21} & w_{22} & w_{23} \end{bmatrix} \times \begin{bmatrix} w_{00} & w_{01} & w_{02} & w_{03} \\ w_{10} & w_{11} & w_{12} & w_{13} \\ w_{20} & w_{21} & w_{22} & w_{23} \\ w_{30} & w_{31} & w_{32} & w_{33} \end{bmatrix} \times \begin{bmatrix} w_{00} & w_{01} \\ w_{10} & w_{11} \\ w_{20} & w_{21} \\ w_{30} & w_{31} \end{bmatrix}$$

$3 \times 4 \qquad \qquad \qquad 4 \times 4 \qquad \qquad \qquad 4 \times 2$

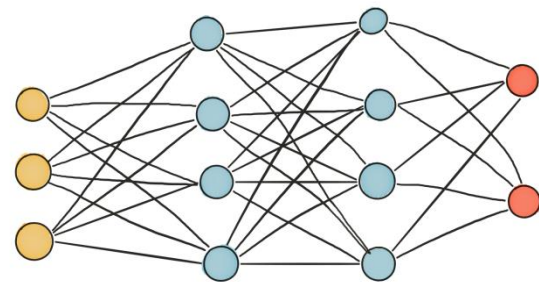
Forward & Back Prop.

$$y^* = w3 * sig(w2 * sig(w1 * x + b1) + b2) + b3$$



쉽게 이해되도록
loss = 예측값-실제로 설정

Forward & Back Prop.

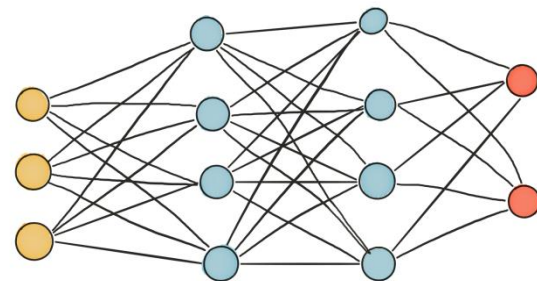


쉽게 이해되도록
loss = 예측값 - 실제로 설정

$$y^* = w3 * sig(w2 * sig(w1 * x + b1) + b2) + b3$$

$$\begin{aligned} loss &= y^* - y \\ &= w3 * sig(w2 * sig(w1 * x + b1) + b2) + b3 - y \end{aligned}$$

Forward & Back Prop.



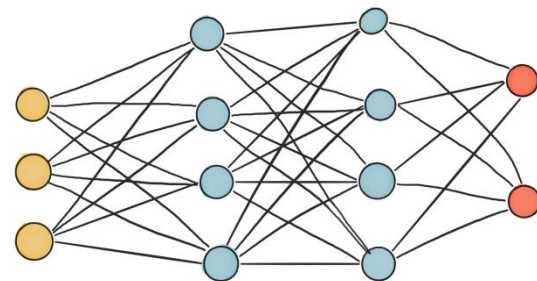
쉽게 이해되도록
loss = 예측값-실제로 설정

$$y^* = w3 * sig(w2 * sig(w1 * x + b1) + b2) + b3$$

$$\begin{aligned} loss &= y^* - y \\ &= w3 * sig(w2 * sig(w1 * x + b1) + b2) + b3 - y \end{aligned}$$

$$\frac{\partial loss}{\partial w3} = sig(w2 * sig(w1 * x + b1) + b2)$$

Forward & Back Prop.



쉽게 이해되도록
loss = 예측값-실제로 설정

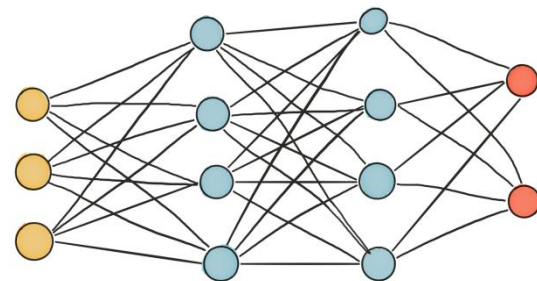
$$y^* = w3 * sig(w2 * sig(w1 * x + b1) + b2) + b3$$

$$\begin{aligned} loss &= y^* - y \\ &= w3 * sig(w2 * sig(w1 * x + b1) + b2) + b3 - y \end{aligned}$$

$$\frac{\partial loss}{\partial w3} = sig(w2 * sig(w1 * x + b1) + b2)$$

$$\frac{\partial loss}{\partial b3} = 1$$

Forward & Back Prop.



쉽게 이해되도록
loss = 예측값-실제로 설정

$$y^* = w3 * sig(w2 * sig(w1 * x + b1) + b2) + b3$$

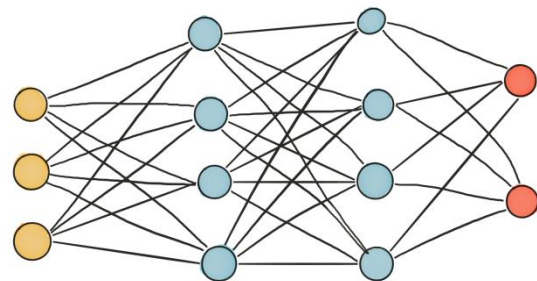
$$\begin{aligned} loss &= y^* - y \\ &= w3 * sig(w2 * sig(w1 * x + b1) + b2) + b3 - y \end{aligned}$$

$$\frac{\partial loss}{\partial w3} = sig(w2 * sig(w1 * x + b1) + b2)$$

$$\frac{\partial loss}{\partial b3} = 1$$

$$\frac{\partial loss}{\partial w2} = ??$$

Forward & Back Prop.



쉽게 이해되도록
loss = 예측값-실제로 설정

$$y^* = w3 * sig(w2 * sig(w1 * x + b1) + b2) + b3$$

$$\begin{aligned} loss &= y^* - y \\ &= w3 * sig(w2 * sig(w1 * x + b1) + b2) + b3 - y \end{aligned}$$

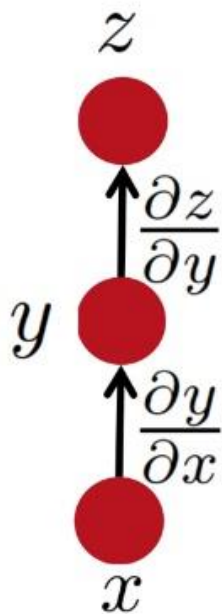
$$\frac{\partial loss}{\partial w3} = sig(w2 * sig(w1 * x + b1) + b2)$$

$$\frac{\partial loss}{\partial b3} = 1$$

$$\frac{\partial loss}{\partial w2} = chain\ rule!!$$

Forward & Back Prop.

Simple Chain Rule



$$\Delta z = \frac{\partial z}{\partial y} \Delta y$$

$$\Delta y = \frac{\partial y}{\partial x} \Delta x$$

$$\Delta z = \frac{\partial z}{\partial y} \frac{\partial y}{\partial x} \Delta x$$

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y} \frac{\partial y}{\partial x}$$

Forward & Back Prop.

Backpropagation: a simple example

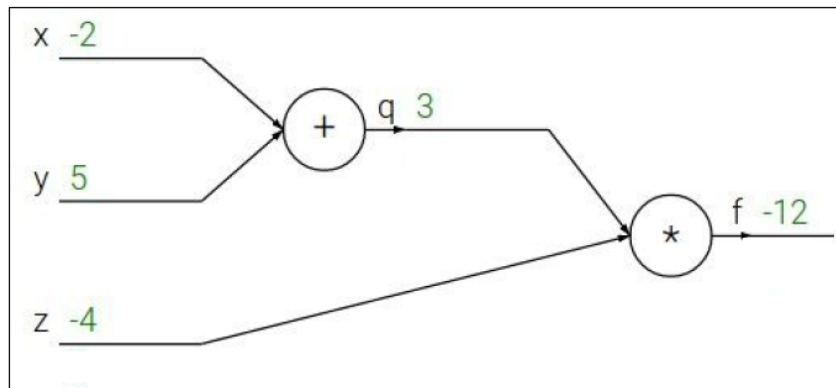
$$f(x, y, z) = (x + y)z$$

e.g. $x = -2, y = 5, z = -4$

$$q = x + y \quad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \quad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Forward & Back Prop.

Backpropagation: a simple example

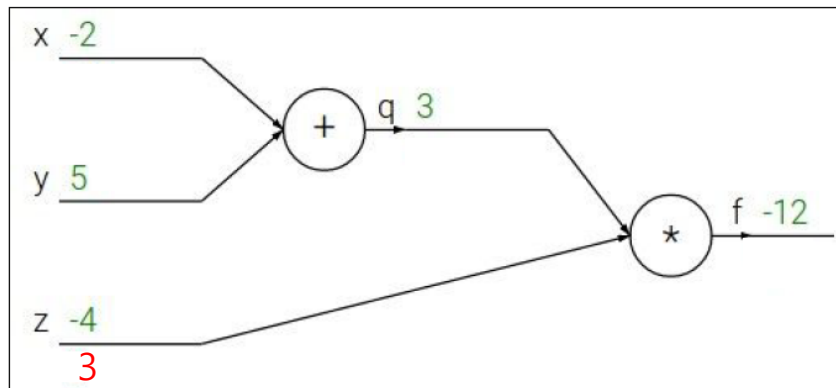
$$f(x, y, z) = (x + y)z$$

e.g. $x = -2, y = 5, z = -4$

$$q = x + y \quad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \quad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



$$\frac{\partial f}{\partial z} = q = x + y = -2 + 5 = 3$$

Forward & Back Prop.

Backpropagation: a simple example

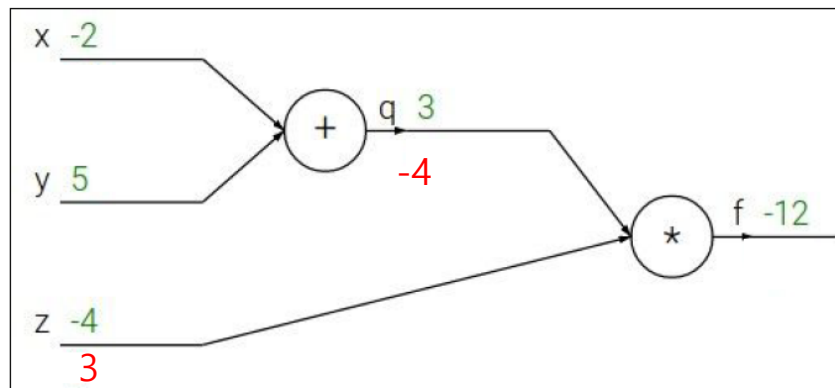
$$f(x, y, z) = (x + y)z$$

e.g. $x = -2, y = 5, z = -4$

$$q = x + y \quad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \quad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



$$\frac{\partial f}{\partial z} = q = x + y = -2 + 5 = 3$$

$$\frac{\partial f}{\partial q} = z = -4$$

Forward & Back Prop.

Backpropagation: a simple example

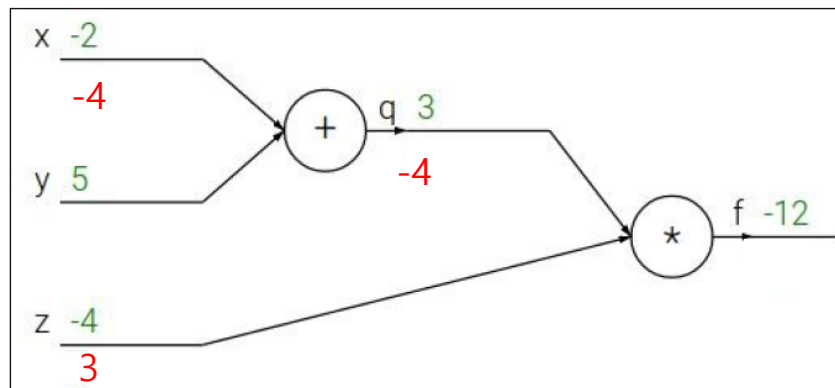
$$f(x, y, z) = (x + y)z$$

e.g. $x = -2, y = 5, z = -4$

$$q = x + y \quad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \quad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



$$\frac{\partial f}{\partial z} = q = x + y = -2 + 5 = 3$$

$$\frac{\partial f}{\partial q} = z = -4 \quad \frac{\partial f}{\partial x} = \frac{\partial f}{\partial q} \frac{\partial q}{\partial x} = -4 * 1 = -4$$

Forward & Back Prop.

Backpropagation: a simple example

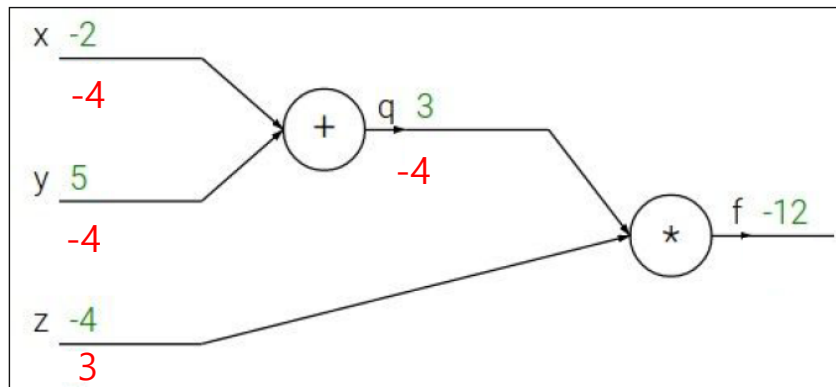
$$f(x, y, z) = (x + y)z$$

e.g. $x = -2, y = 5, z = -4$

$$q = x + y \quad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \quad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$

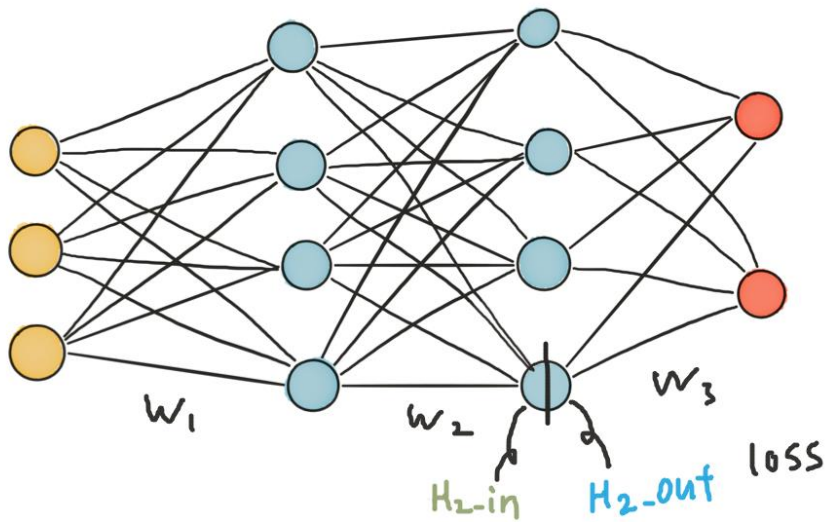


$$\frac{\partial f}{\partial z} = q = x + y = -2 + 5 = 3$$

$$\frac{\partial f}{\partial q} = z = -4 \quad \frac{\partial f}{\partial x} = \frac{\partial f}{\partial q} \frac{\partial q}{\partial x} = -4 * 1 = -4$$

$$\frac{\partial f}{\partial y} = \frac{\partial f}{\partial q} \frac{\partial q}{\partial y} = -4 * 1 = -4$$

Forward & Back Prop.



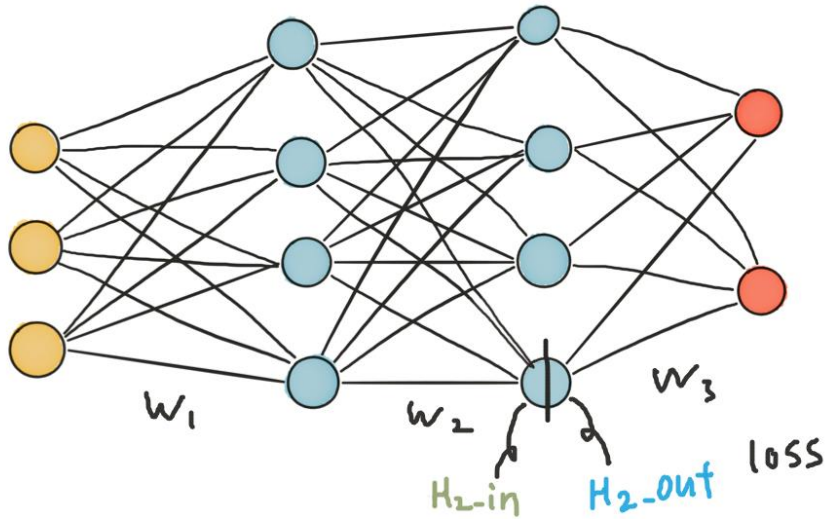
$$loss = w_3 \times \text{sig}(\underbrace{w_2 \times \text{sig}(w_1 x + b_1)}_{H_{2-in}} + b_2)$$

H_{2-out}

$$\frac{\partial loss}{\partial w_2} = \frac{\partial loss}{\partial H_{2-out}} \times \frac{\partial H_{2-out}}{\partial H_{2-in}} \times \frac{\partial H_{2-in}}{\partial w_2}$$

$\frac{\partial loss}{\partial w_2}$: $loss$ 의 w_2 에 대한 편도함수
 $\frac{\partial H_{2-out}}{\partial H_{2-in}}$: H_{2-out} 의 H_{2-in} 에 대한 편도함수
 $\frac{\partial H_{2-in}}{\partial w_2}$: H_{2-in} 의 w_2 에 대한 편도함수

Forward & Back Prop.



$$loss = w_3 \times \text{sig}(\underbrace{w_2 \times \text{sig}(w_1 x + b_1)}_{H_{2-in}} + b_2) + b_3 - y$$

H_{2-out}

$$\frac{\partial loss}{\partial w_2} = \frac{\partial loss}{\partial H_{2-out}} \times \frac{\partial H_{2-out}}{\partial H_{2-in}} \times \frac{\partial H_{2-in}}{\partial w_2}$$

$\frac{\partial loss}{\partial w_2}$: $loss$ 의 w_2 에 대한 편도함수
 $\frac{\partial H_{2-out}}{\partial H_{2-in}}$: H_{2-out} 의 H_{2-in} 에 대한 편도함수
 $\frac{\partial H_{2-in}}{\partial w_2}$: H_{2-in} 의 w_2 에 대한 편도함수

$$\frac{\partial loss}{\partial w_2} = w_3 * \text{sigmoid}'(h2_in) * \text{sigmoid}(w_1 * x + b)$$

Forward & Back Prop.

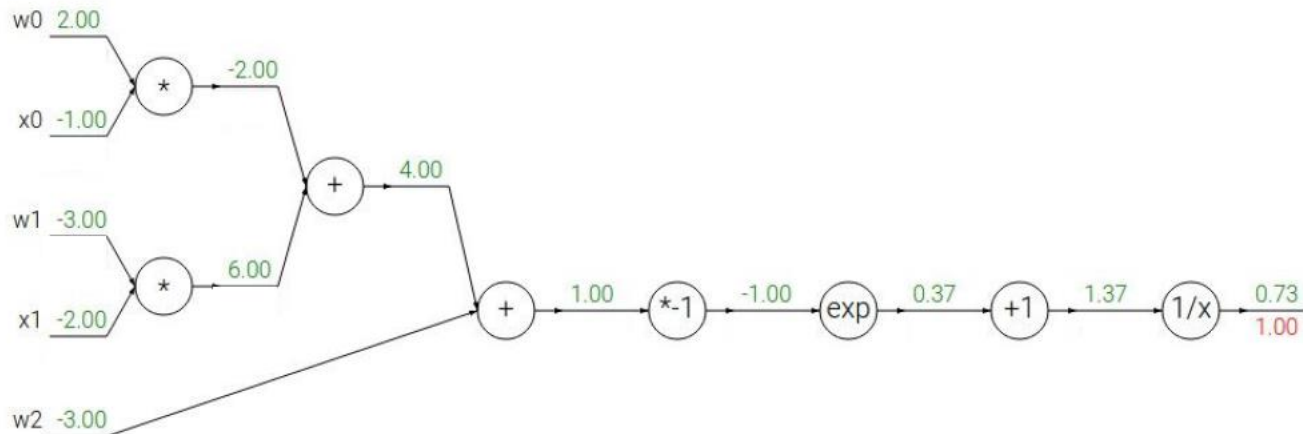
(참고) sigmoid 함수의 미분

$$\sigma(x)' = \frac{\delta\{1+e^{-x}\}^{-1}}{\delta x} = -(1+e^{-x})^{-2} \cdot e^{-x} = \frac{e^{-x}}{(1+e^{-x})^2}$$

$$\sigma(x)(1-\sigma(x)) = \frac{1}{1+e^{-x}} \left(1 - \frac{1}{1+e^{-x}}\right) = \frac{1}{1+e^{-x}} \left(\frac{e^{-x}}{1+e^{-x}}\right) = \frac{e^{-x}}{(1+e^{-x})^2}$$

Forward & Back Prop.

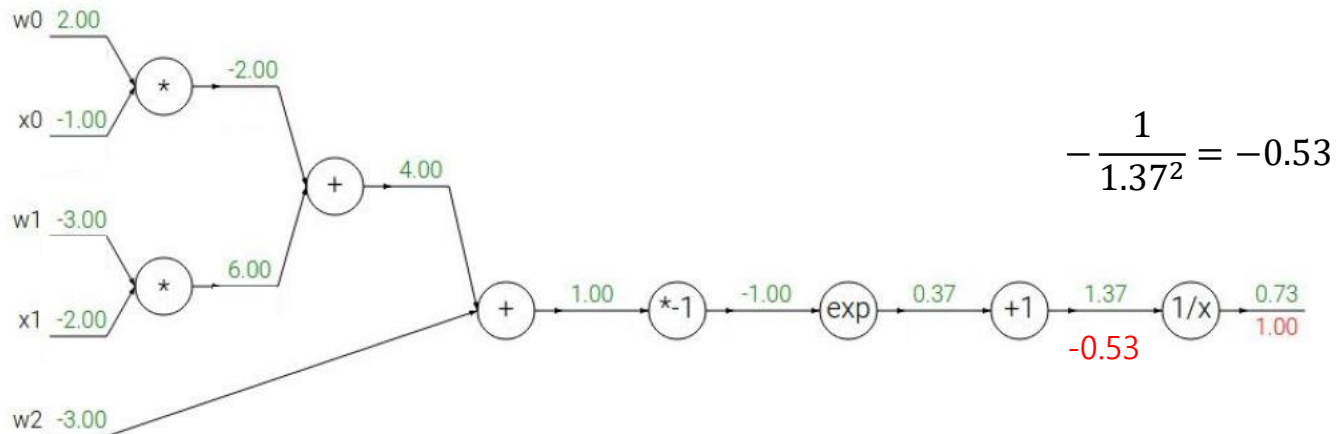
Another example: $f(w, x) = \frac{1}{1 + e^{-(w_0x_0 + w_1x_1 + w_2)}}$



$f(x) = e^x$	\rightarrow	$\frac{df}{dx} = e^x$		$f(x) = \frac{1}{x}$	\rightarrow	$\frac{df}{dx} = -1/x^2$
$f_a(x) = ax$	\rightarrow	$\frac{df}{dx} = a$		$f_c(x) = c + x$	\rightarrow	$\frac{df}{dx} = 1$

Forward & Back Prop.

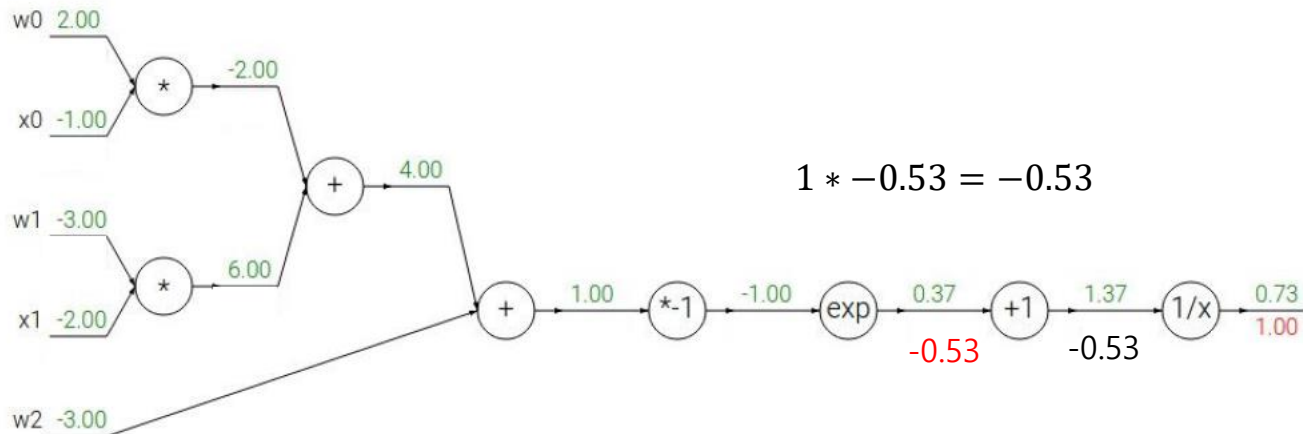
Another example: $f(w, x) = \frac{1}{1 + e^{-(w_0x_0 + w_1x_1 + w_2)}}$



$f(x) = e^x$	→	$\frac{df}{dx} = e^x$		$f(x) = \frac{1}{x}$	→	$\frac{df}{dx} = -1/x^2$
$f_a(x) = ax$	→	$\frac{df}{dx} = a$		$f_c(x) = c + x$	→	$\frac{df}{dx} = 1$

Forward & Back Prop.

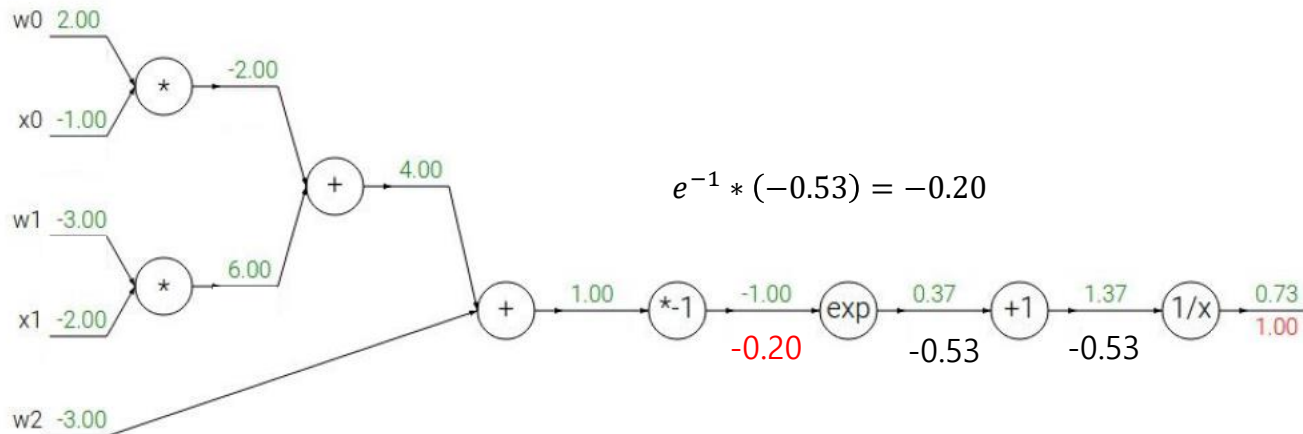
Another example: $f(w, x) = \frac{1}{1 + e^{-(w_0x_0 + w_1x_1 + w_2)}}$



$f(x) = e^x$	→	$\frac{df}{dx} = e^x$		$f(x) = \frac{1}{x}$	→	$\frac{df}{dx} = -1/x^2$
$f_a(x) = ax$	→	$\frac{df}{dx} = a$		$f_c(x) = c + x$	→	$\frac{df}{dx} = 1$

Forward & Back Prop.

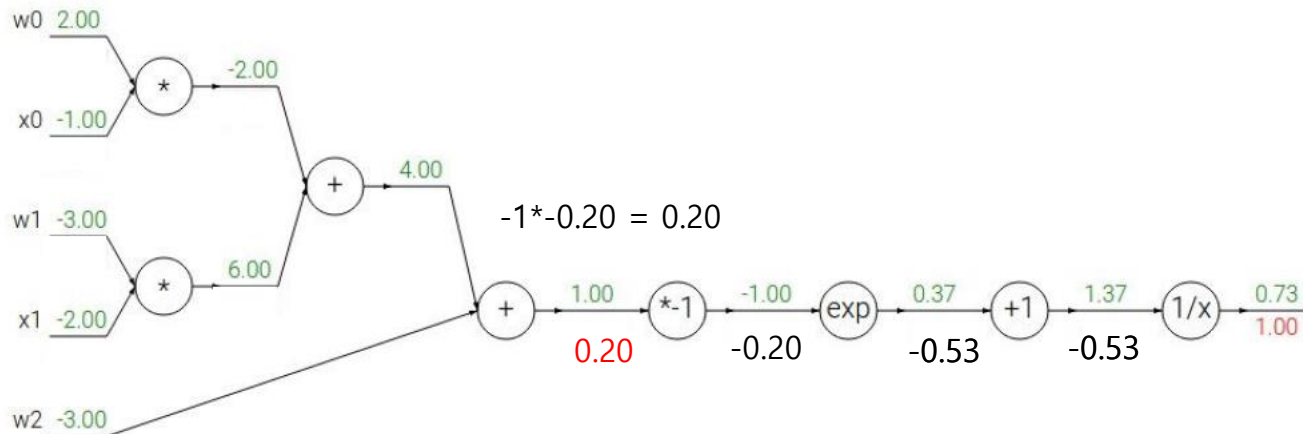
Another example: $f(w, x) = \frac{1}{1 + e^{-(w_0x_0 + w_1x_1 + w_2)}}$



$f(x) = e^x$	→	$\frac{df}{dx} = e^x$		$f(x) = \frac{1}{x}$	→	$\frac{df}{dx} = -1/x^2$
$f_a(x) = ax$	→	$\frac{df}{dx} = a$		$f_c(x) = c + x$	→	$\frac{df}{dx} = 1$

Forward & Back Prop.

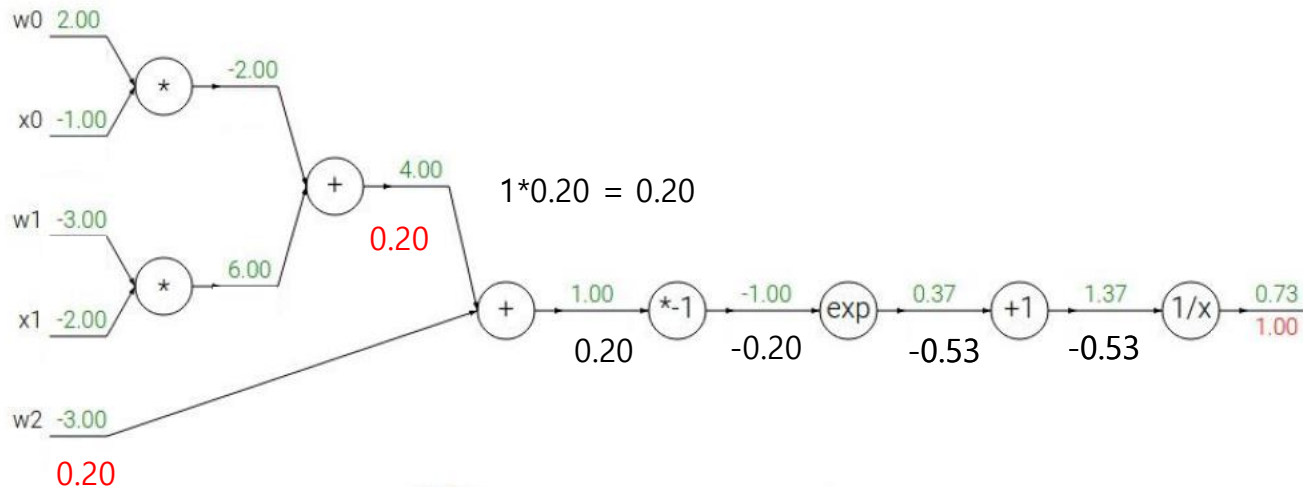
Another example: $f(w, x) = \frac{1}{1 + e^{-(w_0x_0 + w_1x_1 + w_2)}}$



$f(x) = e^x$	→	$\frac{df}{dx} = e^x$		$f(x) = \frac{1}{x}$	→	$\frac{df}{dx} = -1/x^2$
$f_a(x) = ax$	→	$\frac{df}{dx} = a$		$f_c(x) = c + x$	→	$\frac{df}{dx} = 1$

Forward & Back Prop.

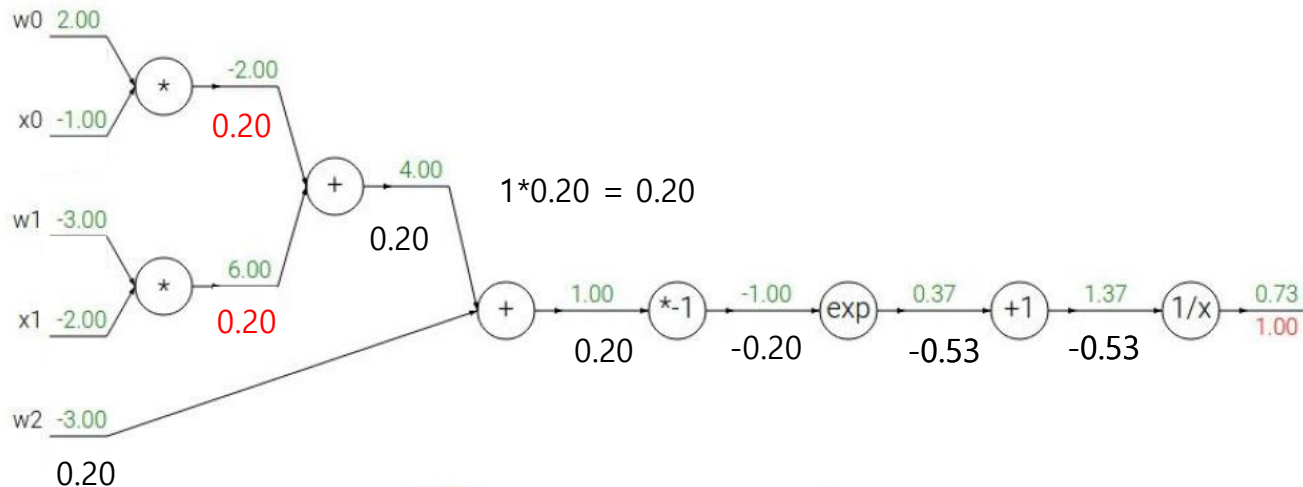
Another example: $f(w, x) = \frac{1}{1 + e^{-(w_0x_0 + w_1x_1 + w_2)}}$



$f(x) = e^x$	→	$\frac{df}{dx} = e^x$		$f(x) = \frac{1}{x}$	→	$\frac{df}{dx} = -1/x^2$
$f_a(x) = ax$	→	$\frac{df}{dx} = a$		$f_c(x) = c + x$	→	$\frac{df}{dx} = 1$

Forward & Back Prop.

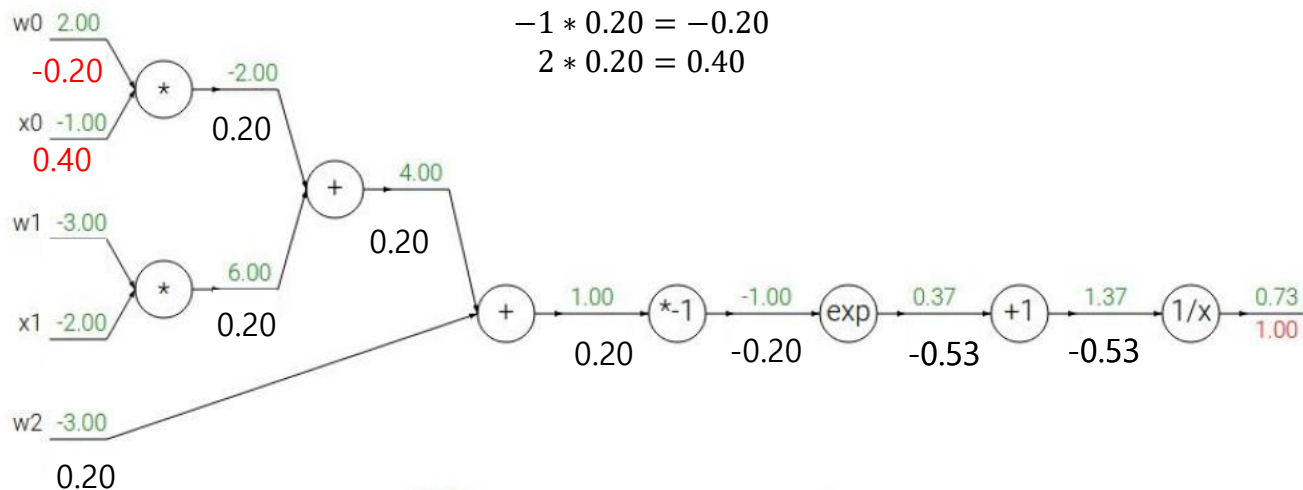
Another example: $f(w, x) = \frac{1}{1 + e^{-(w_0x_0 + w_1x_1 + w_2)}}$



$f(x) = e^x$	\rightarrow	$\frac{df}{dx} = e^x$		$f(x) = \frac{1}{x}$	\rightarrow	$\frac{df}{dx} = -1/x^2$
$f_a(x) = ax$	\rightarrow	$\frac{df}{dx} = a$		$f_c(x) = c + x$	\rightarrow	$\frac{df}{dx} = 1$

Forward & Back Prop.

Another example: $f(w, x) = \frac{1}{1 + e^{-(w_0x_0 + w_1x_1 + w_2)}}$



$$\begin{aligned} -1 * 0.20 &= -0.20 \\ 2 * 0.20 &= 0.40 \end{aligned}$$

$$f(x) = e^x$$

→

$$\frac{df}{dx} = e^x$$

$$f(x) = \frac{1}{x}$$

→

$$\frac{df}{dx} = -1/x^2$$

$$f_a(x) = ax$$

→

$$\frac{df}{dx} = a$$

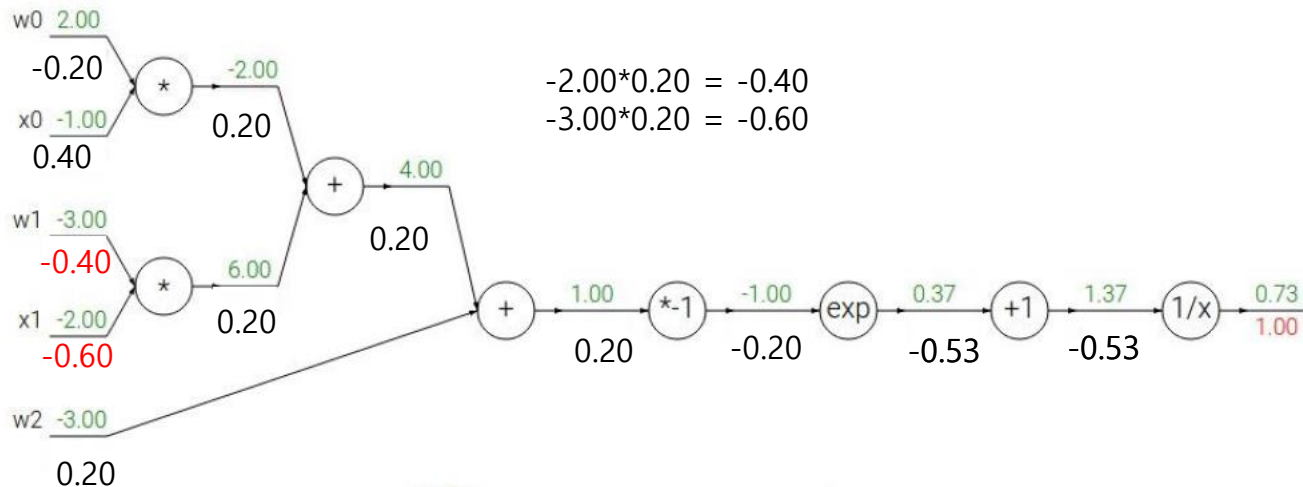
$$f_c(x) = c + x$$

→

$$\frac{df}{dx} = 1$$

Forward & Back Prop.

Another example: $f(w, x) = \frac{1}{1 + e^{-(w_0x_0 + w_1x_1 + w_2x_2)}}$



$f(x) = e^x$	\rightarrow	$\frac{df}{dx} = e^x$		$f(x) = \frac{1}{x}$	\rightarrow	$\frac{df}{dx} = -1/x^2$
$f_a(x) = ax$	\rightarrow	$\frac{df}{dx} = a$		$f_c(x) = c + x$	\rightarrow	$\frac{df}{dx} = 1$

Neural Network

A mostly complete chart of Neural Networks

©2016 Fjodor van Veen - asimovinstitute.org

- Backfed Input Cell
- Input Cell
- △ Noisy Input Cell
- Hidden Cell
- Probablistic Hidden Cell
- △ Spiking Hidden Cell
- Output Cell
- Match Input Output Cell
- Recurrent Cell
- Memory Cell
- △ Different Memory Cell
- Kernel
- Convolution or Pool

Perceptron (P)



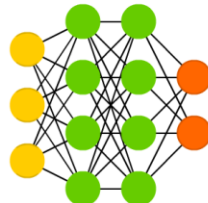
Feed Forward (FF)



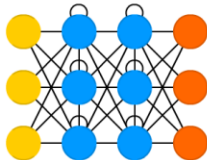
Radial Basis Network (RBF)



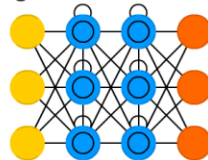
Deep Feed Forward (DFF)



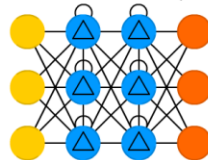
Recurrent Neural Network (RNN)



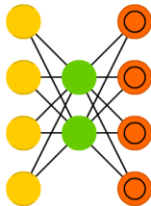
Long / Short Term Memory (LSTM)



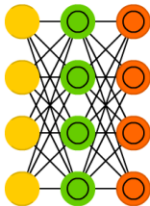
Gated Recurrent Unit (GRU)



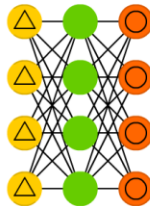
Auto Encoder (AE)



Variational AE (VAE)



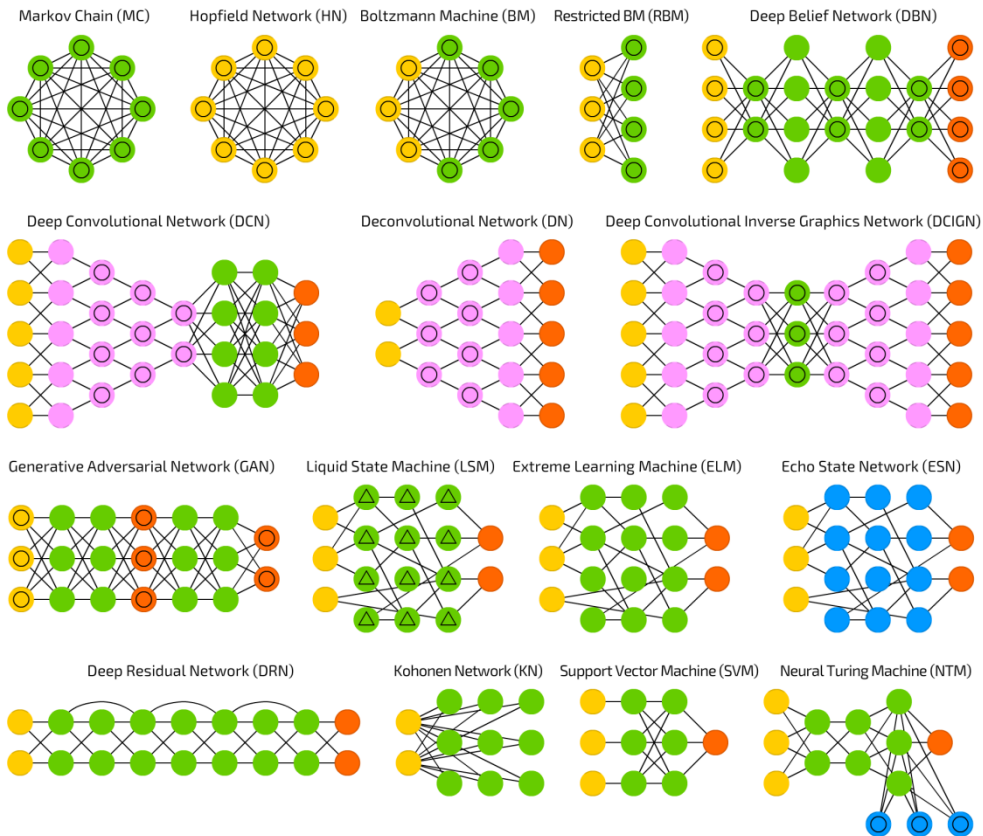
Denosing AE (DAE)



Sparse AE (SAE)



Neural Network



Forward & Back Prop.

```
1 import numpy as np
2 import torch
3 import torch.nn as nn
4 import torch.optim as optim
5 import torch.nn.init as init
6 from torch.autograd import Variable
7 from visdom import Visdom
8 viz = Visdom()
9
10 # data generation
11
12 num_data = 1000
13 num_epoch = 5000
14
15 noise = init.normal(torch.FloatTensor(num_data,1),std=0.5)
16 x = init.uniform(torch.Tensor(num_data,1),-15,15)
17 y = 8*(x**2) + 7*x + 3
18 x_noise = x + noise
19 y_noise = 8*(x_noise**2) + 7*x_noise + 3
```

Forward & Back Prop.

필요한 라이브러리

```
1 import numpy as np
2 import torch
3 import torch.nn as nn
4 import torch.optim as optim
5 import torch.nn.init as init
6 from torch.autograd import Variable
7 from visdom import Visdom
8 viz = Visdom()
9
10 # data generation
11
12 num_data = 1000
13 num_epoch = 5000
14
15 noise = init.normal(torch.FloatTensor(num_data,1),std=0.5)
16 x = init.uniform(torch.Tensor(num_data,1),-15,15)
17 y = 8*(x**2) + 7*x + 3
18 x_noise = x + noise
19 y_noise = 8*(x_noise**2) + 7*x_noise + 3
```

Forward & Back Prop.

필요한 라이브러리

```
1 import numpy as np
2 import torch
3 import torch.nn as nn
4 import torch.optim as optim
5 import torch.nn.init as init
6 from torch.autograd import Variable
7 from visdom import Visdom
8 viz = Visdom()
9
```

데이터 생성

```
10 # data generation
11
12 num_data = 1000
13 num_epoch = 5000
14
15 noise = init.normal(torch.FloatTensor(num_data,1),std=0.5)
16 x = init.uniform(torch.Tensor(num_data,1),-15,15)
17 y = 8*(x**2) + 7*x + 3
18 x_noise = x + noise
19 y_noise = 8*(x_noise**2) + 7*x_noise + 3
```

Forward & Back Prop.

```
44 # model & optimizer
45
46 model = nn.Sequential(
47     nn.Linear(1,10),
48     nn.Sigmoid(),
49     nn.Linear(10,6),
50     nn.Sigmoid(),
51     nn.Linear(6,1),
52 ).cuda()
53
54 loss_func = nn.L1Loss()
55 optimizer = optim.SGD(model.parameters(),lr=0.001)
56
57 # train
58 loss_arr = []
59 label = Variable(y_noise.cuda())
60 for i in range(num_epoch):
61     output = model(Variable(x.cuda()))
62     optimizer.zero_grad()
63
64     loss = loss_func(output,label)
65     loss.backward()
66     optimizer.step()
67     #print(loss)
68     loss_arr.append(loss.cpu().data.numpy()[0])
69
70 param_list = List(model.parameters())
71 print(param_list[0].data,param_list[1].data)
```

Forward & Back Prop.

Neural Network 모델 생성

```
44 # model & optimizer
45
46 model = nn.Sequential(
47     nn.Linear(1,10),
48     nn.Sigmoid(),
49     nn.Linear(10,6),
50     nn.Sigmoid(),
51     nn.Linear(6,1),
52 ).cuda()
53
54 loss_func = nn.L1Loss()
55 optimizer = optim.SGD(model.parameters(),lr=0.001)
56
57 # train
58 loss_arr = []
59 label = Variable(y_noise.cuda())
60 for i in range(num_epoch):
61     output = model(Variable(x.cuda()))
62     optimizer.zero_grad()
63
64     loss = loss_func(output,label)
65     loss.backward()
66     optimizer.step()
67     #print(loss)
68     loss_arr.append(loss.cpu().data.numpy()[0])
69
70 param_list = List(model.parameters())
71 print(param_list[0].data,param_list[1].data)
```

Forward & Back Prop.

Neural Network 모델 생성

loss function 및
gradient descent
optimizer 생성

```
44 # model & optimizer
45
46 model = nn.Sequential(
47     nn.Linear(1,10),
48     nn.Sigmoid(),
49     nn.Linear(10,6),
50     nn.Sigmoid(),
51     nn.Linear(6,1),
52 ).cuda()
53
54 loss_func = nn.L1Loss()
55 optimizer = optim.SGD(model.parameters(), lr=0.001)
56
57 # train
58 loss_arr = []
59 label = Variable(y_noise.cuda())
60 for i in range(num_epoch):
61     output = model(Variable(x.cuda()))
62     optimizer.zero_grad()
63
64     loss = loss_func(output, label)
65     loss.backward()
66     optimizer.step()
67     #print(loss)
68     loss_arr.append(loss.cpu().data.numpy()[0])
69
70 param_list = List(model.parameters())
71 print(param_list[0].data, param_list[1].data)
```

Forward & Back Prop.

Neural Network 모델 생성

loss function 및
gradient descent
optimizer 생성

<training 단계>

1. 모델로 결과값 추정
2. loss 및 gradient 계산
3. 모델 업데이트

```
44 # model & optimizer
45
46 model = nn.Sequential(
47     nn.Linear(1,10),
48     nn.Sigmoid(),
49     nn.Linear(10,6),
50     nn.Sigmoid(),
51     nn.Linear(6,1),
52 ).cuda()
53
54 loss_func = nn.L1Loss()
55 optimizer = optim.SGD(model.parameters(), lr=0.001)
56
57 # train
58 loss_arr = []
59 label = Variable(y_noise.cuda())
60 for i in range(num_epoch):
61     output = model(Variable(x.cuda()))
62     optimizer.zero_grad()
63
64     loss = loss_func(output, label)
65     loss.backward()
66     optimizer.step()
67     #print(loss)
68     loss_arr.append(loss.cpu().data.numpy()[0])
69
70 param_list = List(model.parameters())
71 print(param_list[0].data, param_list[1].data)
```

Forward & Back Prop.

Neural Network 모델 생성

loss function 및
gradient descent
optimizer 생성

<training 단계>

1. 모델로 결과값 추정
2. loss 및 gradient 계산
3. 모델 업데이트

training 이후 파라미터 값 확인

```
44 # model & optimizer
45
46 model = nn.Sequential(
47     nn.Linear(1,10),
48     nn.Sigmoid(),
49     nn.Linear(10,6),
50     nn.Sigmoid(),
51     nn.Linear(6,1),
52 ).cuda()
53
54 loss_func = nn.L1Loss()
55 optimizer = optim.SGD(model.parameters(), lr=0.001)
56
57 # train
58 loss_arr = []
59 label = Variable(y_noise.cuda())
60 for i in range(num_epoch):
61     output = model(Variable(x.cuda()))
62     optimizer.zero_grad()
63
64     loss = loss_func(output, label)
65     loss.backward()
66     optimizer.step()
67     #print(loss)
68     loss_arr.append(loss.cpu().data.numpy()[0])
69
70 param_list = list(model.parameters())
71 print(param_list[0].data, param_list[1].data)
```

Q&A
