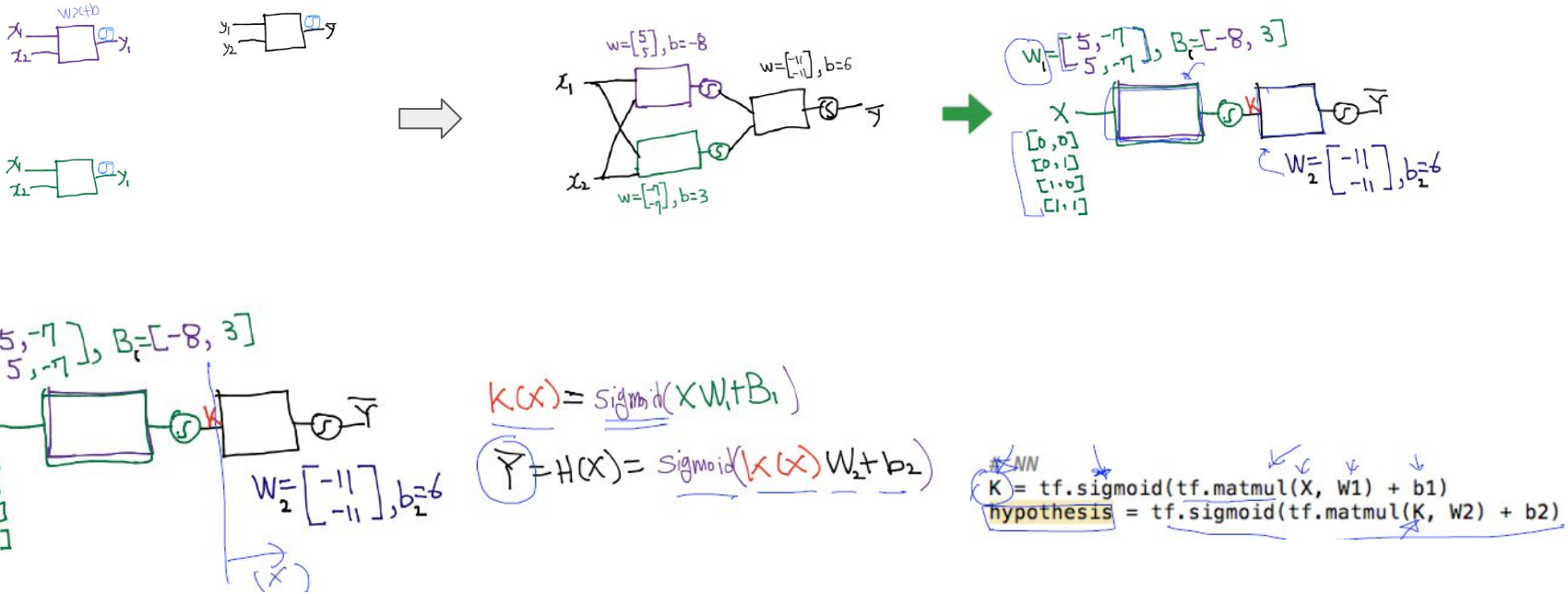
머신러닝 스터디 5th week 보조 자료

20180209 김성헌

summary

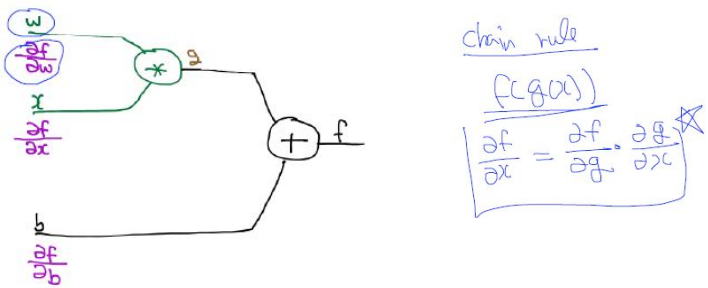
- ch.9-1 Neural Nets for XOR
  - 

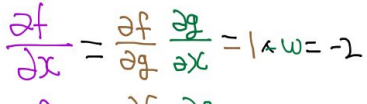
summary

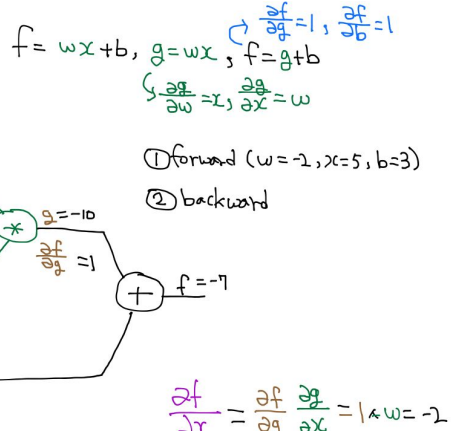
- ch.9-2 Backpropagation

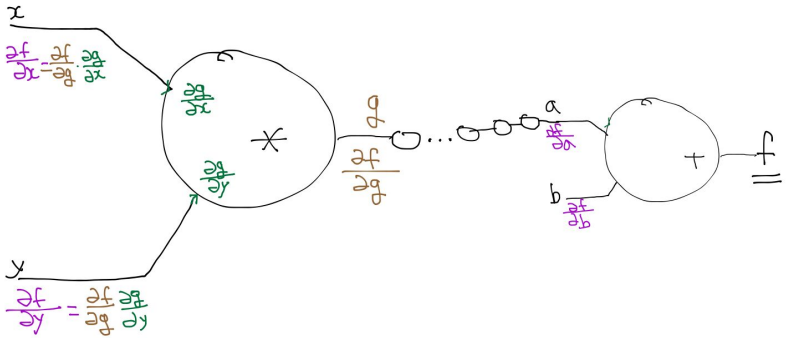
$$\frac{d}{dx}f(x) = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

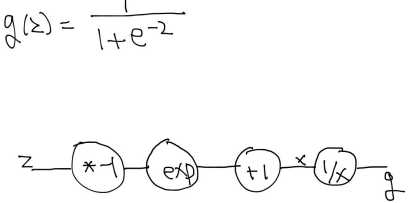
$f(x) = 3$ $f(x) = 2x$ $f(x) = 3$ $f(x) = 2x$ $f(x) = x + x$ $f(x) = x$ $f(x, y) = xy, \frac{\partial f}{\partial x}$ $f(x) = x + 3$ $f(x) = 2x$ $f(x, y) = x + y, \frac{\partial f}{\partial x}$ $f(x, y) = xy, \frac{\partial f}{\partial y}$ $f(x, y) = x + y, \frac{\partial f}{\partial y}$



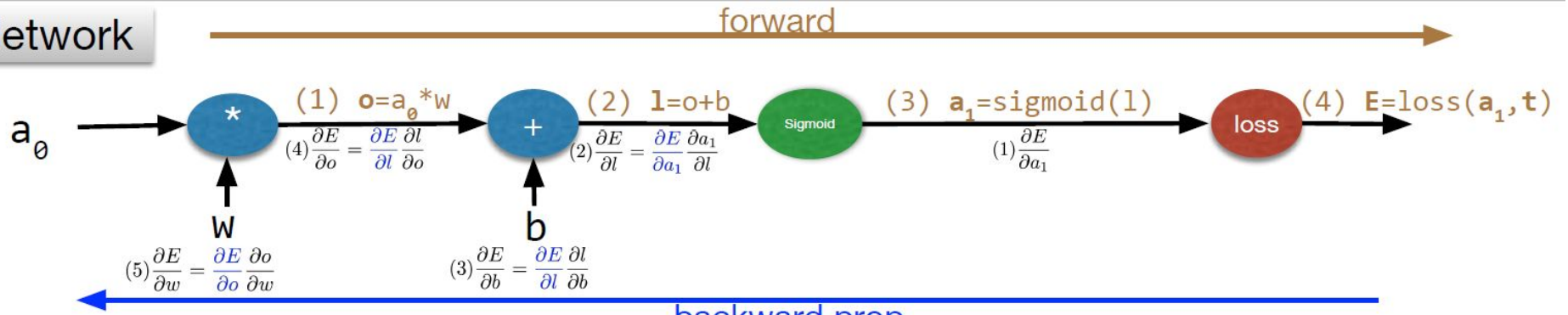








Network



Derivatives (chain rule)

$$E = -\sum tlog(a) + (1-t)log(1-a), \quad \frac{\partial E}{\partial a} = \frac{a-t}{a(1-a)}$$

$$a = sigmoid(l) = \frac{1}{1+e^{-l}}, \quad \frac{\partial a}{\partial l} = a(1-a)$$

$$(1)\frac{\partial E}{\partial a_1} = \frac{a_1 - t}{a_1(1-a_1)}$$

$$(2)\frac{\partial E}{\partial l} = \frac{\partial E}{\partial a_1}\frac{\partial a_1}{\partial l} = \frac{a_1 - t}{a_1(1-a_1)} * a_1(1-a_1) = a_1 - t$$

$$(3)\frac{\partial E}{\partial b} = \frac{\partial E}{\partial l}\frac{\partial l}{\partial b} = \frac{\partial E}{\partial l} * 1 = a_1 - t$$

$$(4)\frac{\partial E}{\partial o} = \frac{\partial E}{\partial l}\frac{\partial l}{\partial o} = \frac{\partial E}{\partial l} * 1 = a_1 - t$$

$$(5)\frac{\partial E}{\partial w} = \frac{\partial E}{\partial o}\frac{\partial o}{\partial w} = a_0^T\frac{\partial E}{\partial o} = a_0^T(a_1 - t)$$

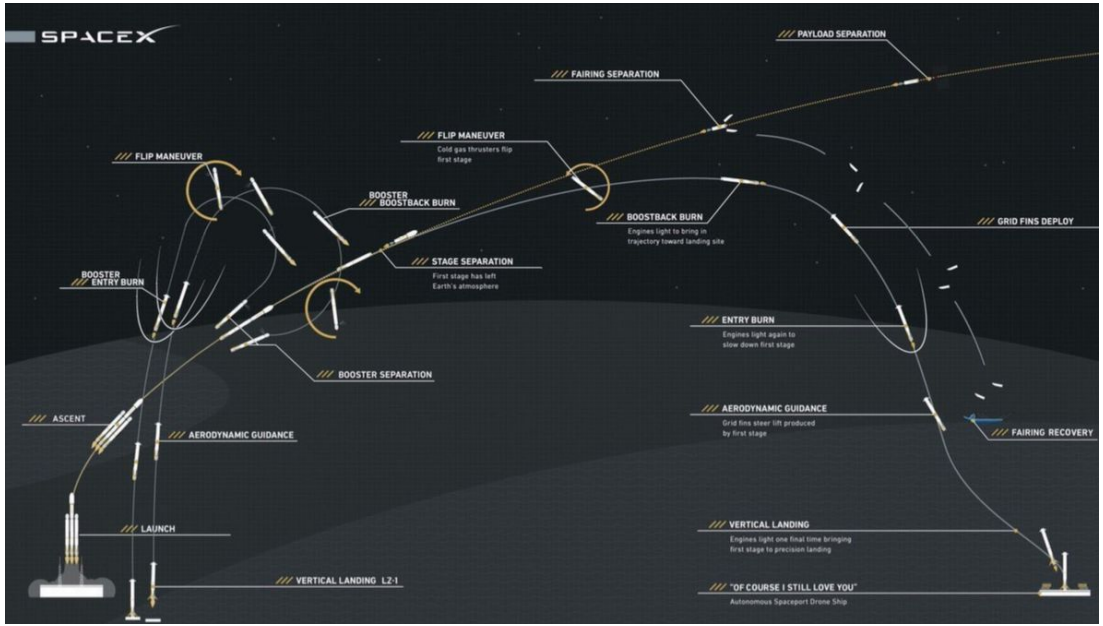
```
d_a1 = (a1 - t) / (a1 * (1. - a1) + 1e-7)
d_sigma = a1 * (1 - a1) # sigma prime
d_l = d_a1 * d_sigma  # (a1 - t)
d_b = d_l * 1
d_o = d_l * 1
d_W = tf.matmul(tf.transpose(a0), d_o)
# Updating network using gradients
learning_rate = 0.01
train_step = [
   tf.assign(W, W - learning_rate * d_W),
   tf.assign(b, b - learning_rate * d_b)]
```

Neural Nets 보충

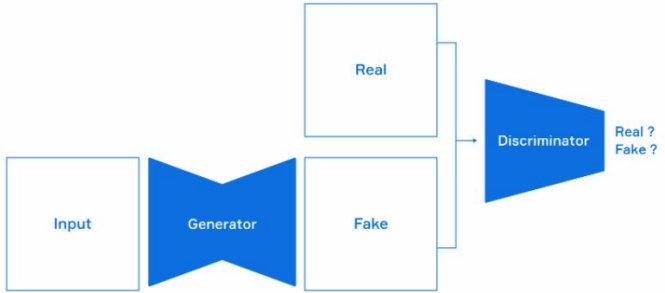
- NN 사용 목적
  - non-linear decision boundary 가 필요
- 활성화 함수
  - 각 뉴런의 계산값이 기준치 이상이면 신호를 출력(1)하고 그렇지 않으면 출력하지 않음(0)
  - 활성화 함수는 반드시 비선형이어야 한다.!!!
    - 왜?
      - h(x)=cx를 사용한 3층 네트워크가 있다고 하자.. y(x)=h(h(h(x))) -> y(x)=c*c*c*x -> y(x)=ax와 똑같은 형식. 즉 1층 네트워크. 층을 쌓는 의미가 없음.

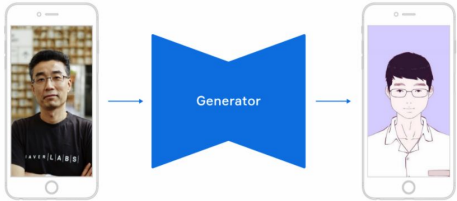
쉬어가는 주제

- 스페이스X 팔콘헤비
  - 발사된 로켓이 안전하게 착륙
  - http://res.thegear.co.kr/images/20180207/20180207104333641416.gif
  - 2018년 2월 6일 시험발사
  - 엘론 머스크가 타던 스포츠카에 마네킹”스타맨”을 앉혀서 화성에 근접한 태양공전궤도에 도달
  - 2024년 유인화성탐사선 발사 계획을 위한 첫 시험발사



faceswap 기술





- GAN 기반 얼굴교체 기술
  - GAN(Generative adversarial network)
    - 2014년 NIPS 학회에서 발표
    - 1시간만에 GAN 완전정복 : https://www.youtube.com/watch?v=odpjk7_tGY0
    - 네이버 웹툰 '마주쳤다'에도 GAN 적용 : https://www.naverlabs.com/storyDetail/44
- for 니콜라스 케이지
  - https://gfycat.com/yearlynewbarb
  - https://movieweb.com/nicolas-cage-deep-learning-technology-face-replacement/
- deepfakes
  - 유명인 얼굴을 합성한 가짜 porn 제작
  - TensorFlow사용하며 일반 소비자용 GPU로 몇시간 만에 처리 가능 (CPU는 며칠
  - https://github.com/joshua-wu/deepfakes_faceswap
- MNIST를 이용한 GAN
  - http://jsideas.net/python/2017/07/01/GAN.html
- faceswap-GAN
  - https://github.com/shaoanlu/faceswap-GAN

GAN

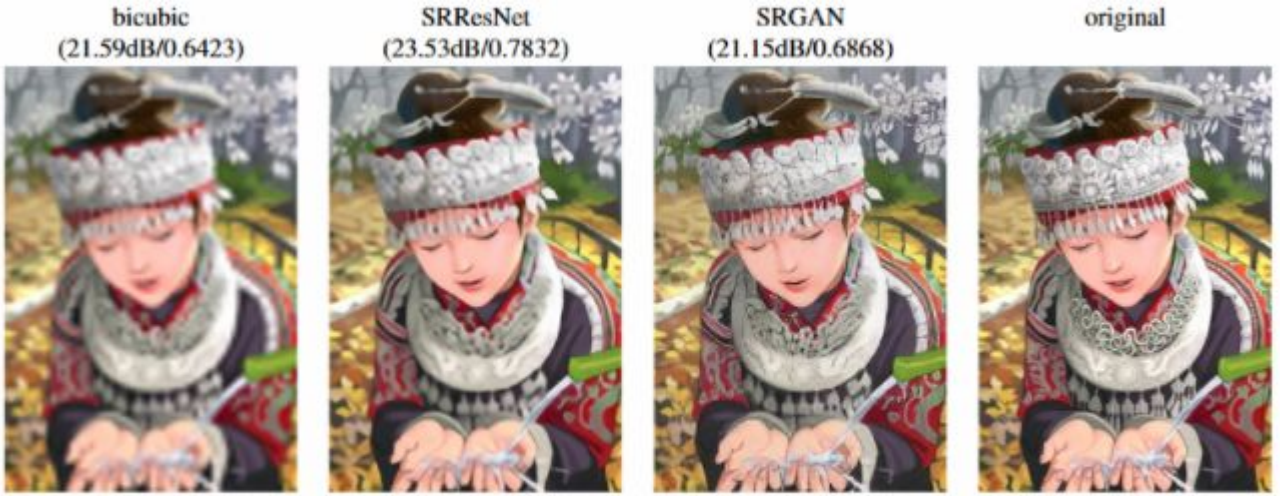


Figure 2: From left to right: bicubic interpolation, deep residual network optimized for MSE, deep residual generative adversarial network optimized for a loss more sensitive to human perception, original HR image. Corresponding PSNR and SSIM are shown in brackets. [4× upscaling]

Figure 16: Example results of our method on automatically detected edges→shoes, compared to ground truth.