Training Neural Networks

- Activation Function

Sigmoid $\sigma(x) = \frac{1}{1+e^{-x}}$

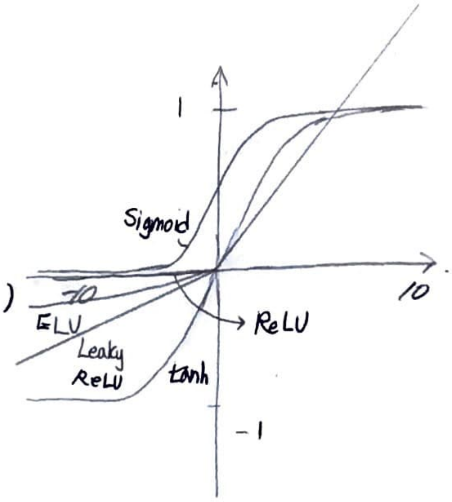
Leaky ReLU $\max(0.1x, x)$

tanh

Maxout $\max(w_1^T x + b_1, w_2^T x + b_2)$

ReLU $\max(0, x)$

ELU $\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$



(use ReLU)

- Data Preprocessing

Normalization → PCA. Whitening of the data

- just subtract the mean image
  subtract per-channel mean

- Weight Initialization

Small random numbers : bad at deeper network

zero bad at backpropagation

random /sqrt 그나마 better

- Batch Normalization

usually after FC or Convolutional layers

$$\hat{x}^{(k)} = \frac{x^{(k)} - E[x^{(k)}]}{\sqrt{Var[x^{(k)}]}} \xrightarrow{\text{recover}} y^{(k)} = \sqrt{Var[x^{(k)}]}\,\hat{x}^{(k)} + E[x^{(k)}]$$

- Babysitting the Learning Process

- Hyperparameter Optimization

Cross-Validation Strategy

random sample hyperparams, in log space when appropriate

- Optimizations: problems with SGD　　　파라미터 update 정리.

SGD　$x_{t+1} = x_t - \alpha \nabla f(x_t)$

↓

SGD + Momentum　$v_{t+1} = \rho v_t + \nabla f(x_t)$

$x_{t+1} = x_t - \alpha v_{t+1}$

SGD + Nesterov Momentum　$v_{t+1} = \rho v_t - \alpha \nabla f(x_t + \rho v_t)$

$x_{t+1} = x_t + v_{t+1}$

AdaGrad

$h_{t+1} = h_t + \frac{\partial}{\partial W} L \odot \frac{\partial}{\partial W} L$

$W_{t+1} = W_t - \eta \frac{1}{\sqrt{h}} \frac{\partial}{\partial W} L$

RMSProp　$h_{t+1} = \rho h_t + (1-\rho) \frac{\partial}{\partial W} L_i \odot \frac{\partial}{\partial W} L_i$

Adam
default.

1차 momentum m　　2차 momentum v

$m_1 \leftarrow \beta_1 m_0 + (1-\beta_1) g_1$

$\hat{m}_1 \leftarrow \frac{m_1}{1-\beta_1^1} = \frac{\beta_1 m_0}{1-\beta_1^1} + \frac{(1-\beta_1) g_1}{1-\beta_1^1} = 0 + g_1 \; (\because m_0 = 0)$　이를 반복

- Model Ensembles

1. train multiple independent models
2. at test time average their result.

- Regularization
  - add term to loss

    L2 regularization
    L1 regularization
    Elastic Net (L1+L2)

  - Dropout　randomly set neurons to zero
  - Data Augmentation
    random crops and scales
    horizontal flips