Welcome to

DS3010:
DS-III: Computational Data Intelligence
Activation Functions in DNN

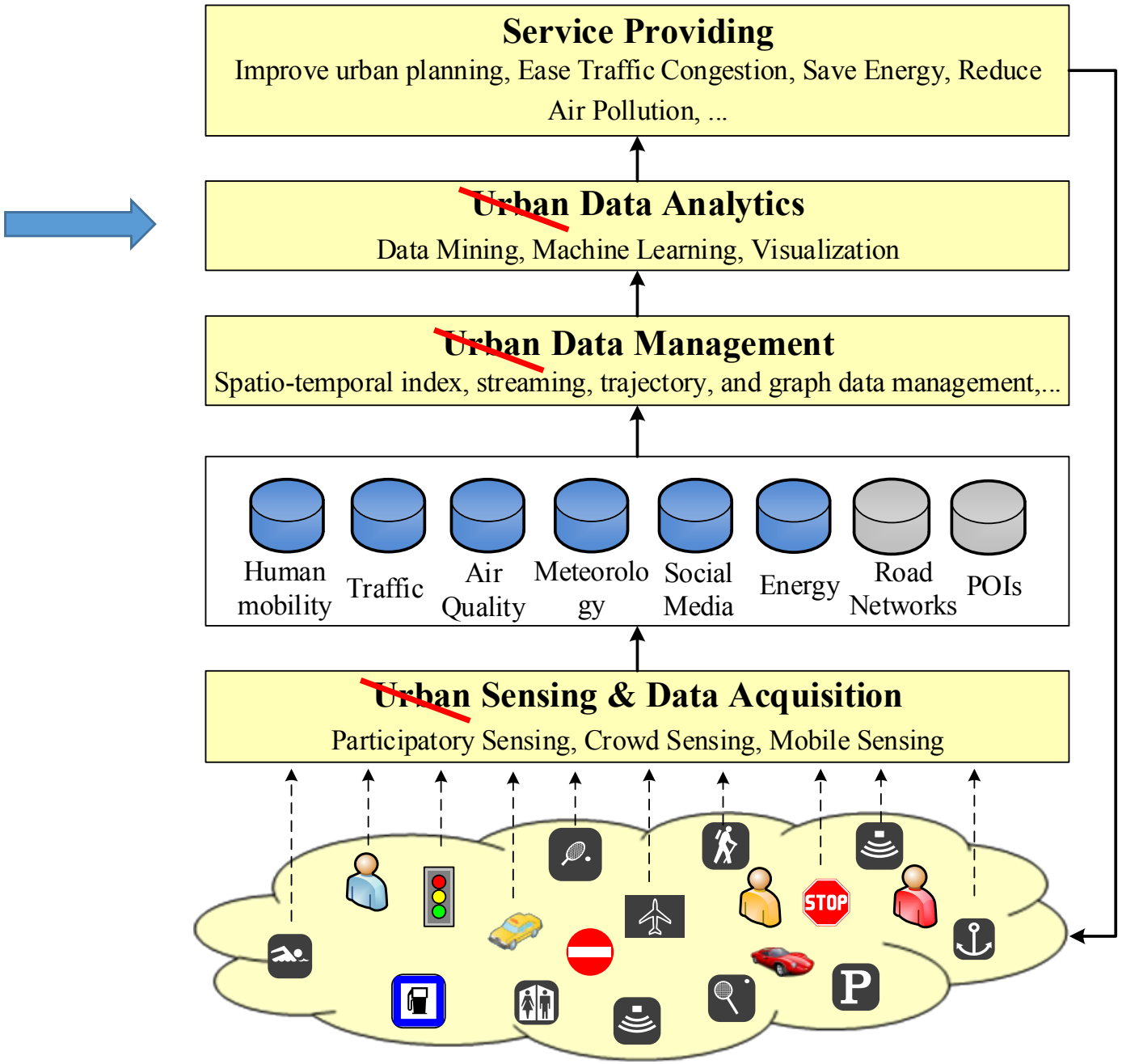
Prof. Yanhua Li

Time: 11:00am – 12:50pm M & R
Location: HL 114
D-term 2022

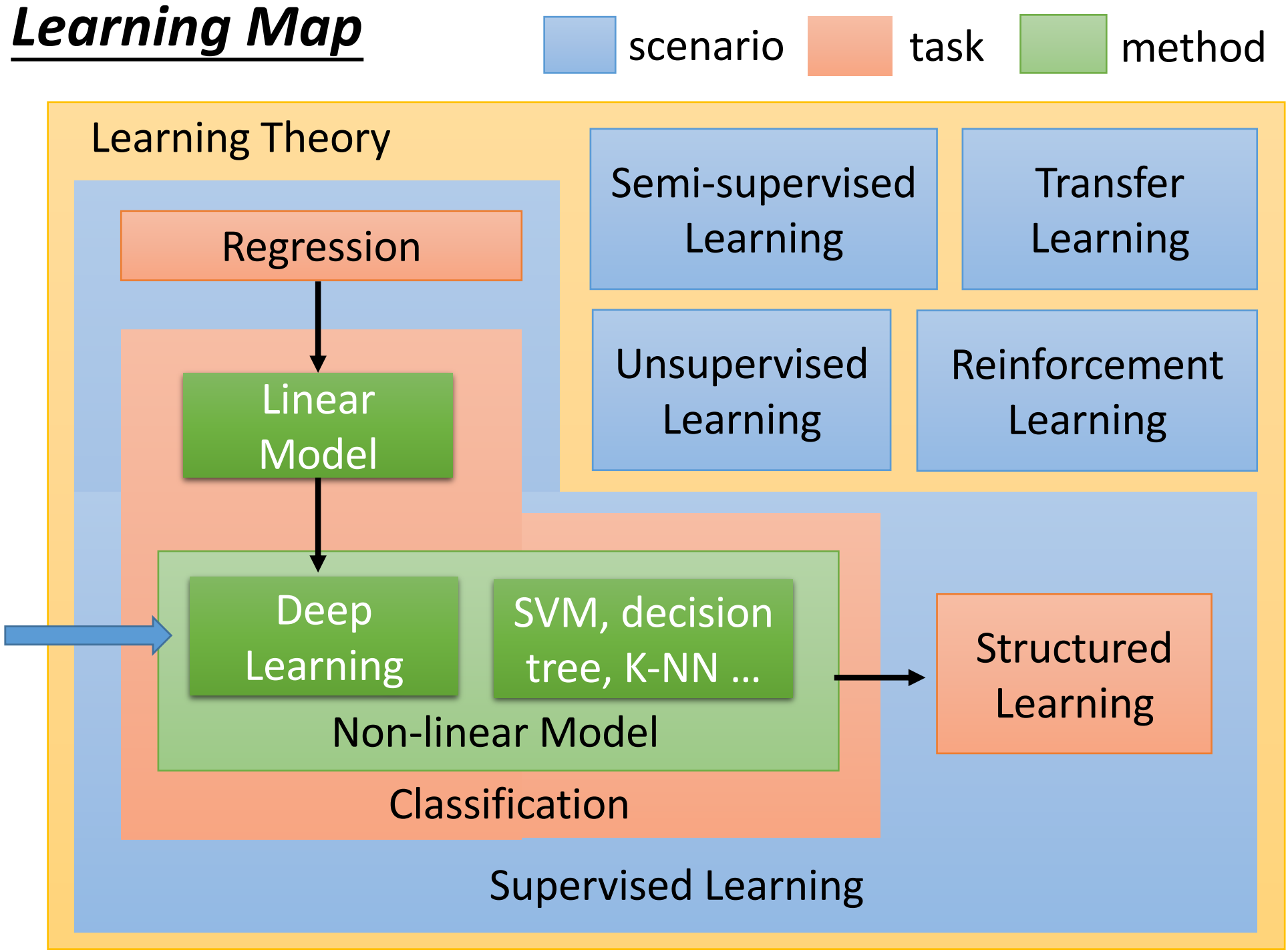
Quiz #2

- Thursday 4/14
- 15 mins, 11-11:15AM
- The quiz 2 will be available at 10:58AM for your convenience.
- Topics:
- Question 1: Logistic regression (its limitation)
- Question 2: Logistic regression for multi-class classification

Data pipeline

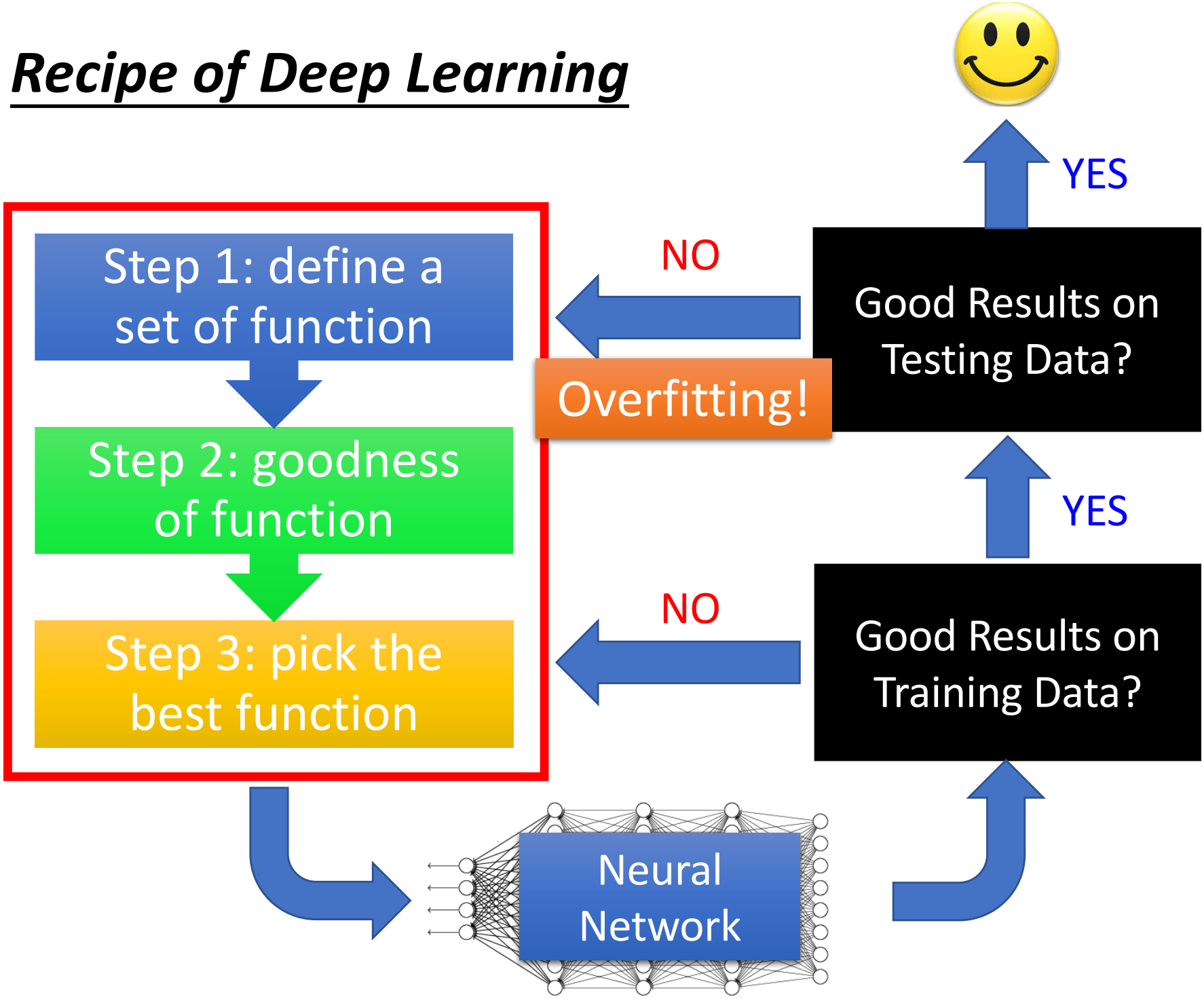


Urban Computing: concepts, methodologies, and applications.
Zheng, Y., et al. *ACM transactions on Intelligent Systems and Technology*.

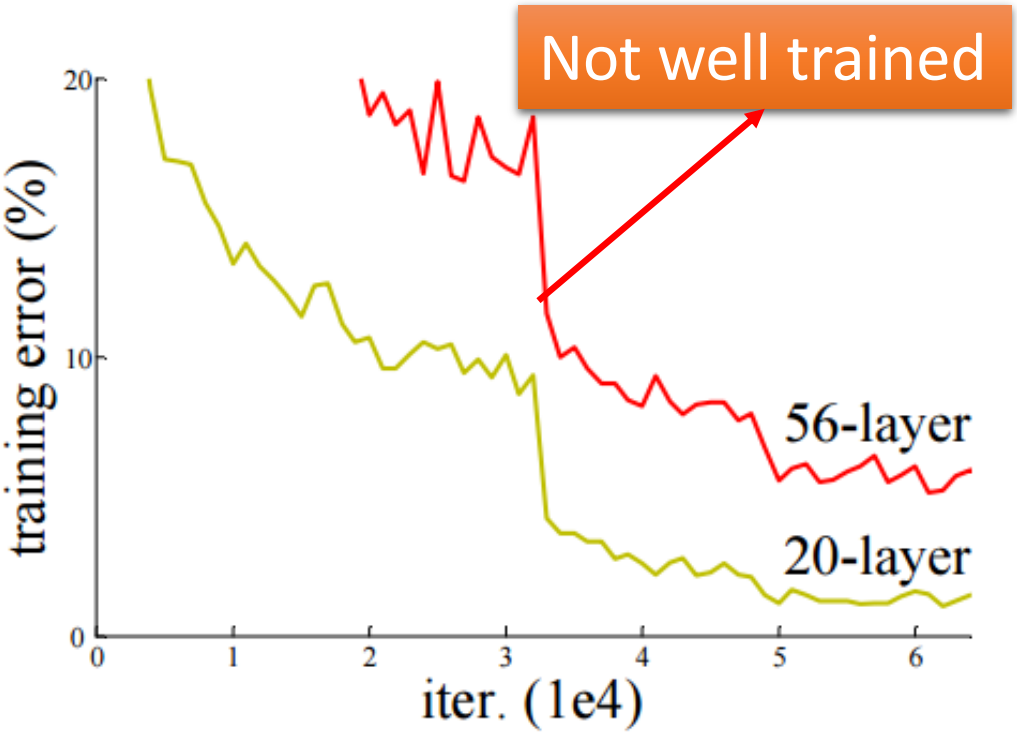

Learning Map
scenario
task
method
Learning Theory
Semi-supervised Learning
Transfer Learning
Regression
Unsupervised Learning
Reinforcement Learning
Linear Model
Deep Learning
SVM, decision tree, K-NN …
Non-linear Model
Structured Learning
Classification
Supervised Learning

Activation Functions for Deep Learning

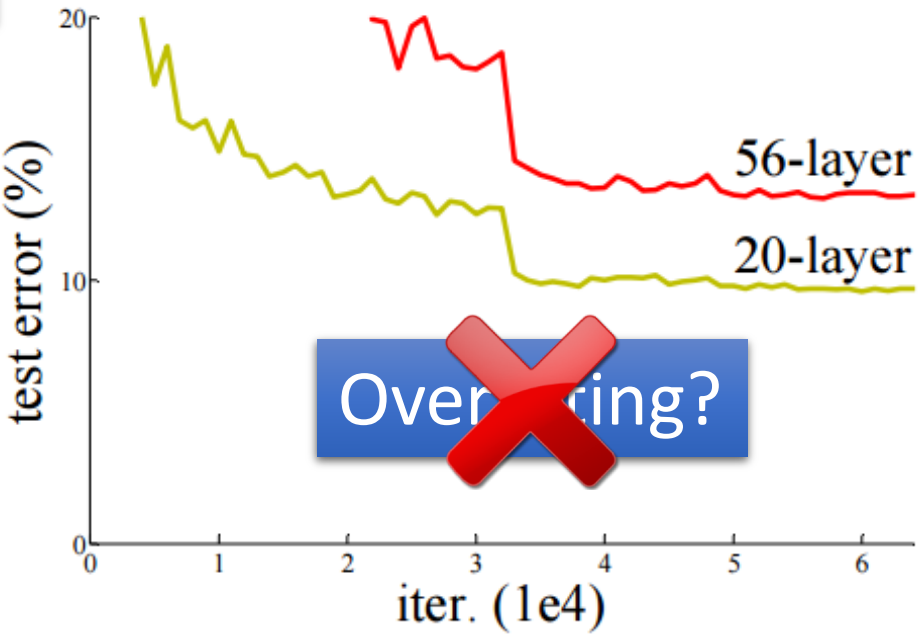
Recipe of Deep Learning



Do not always blame Overfitting



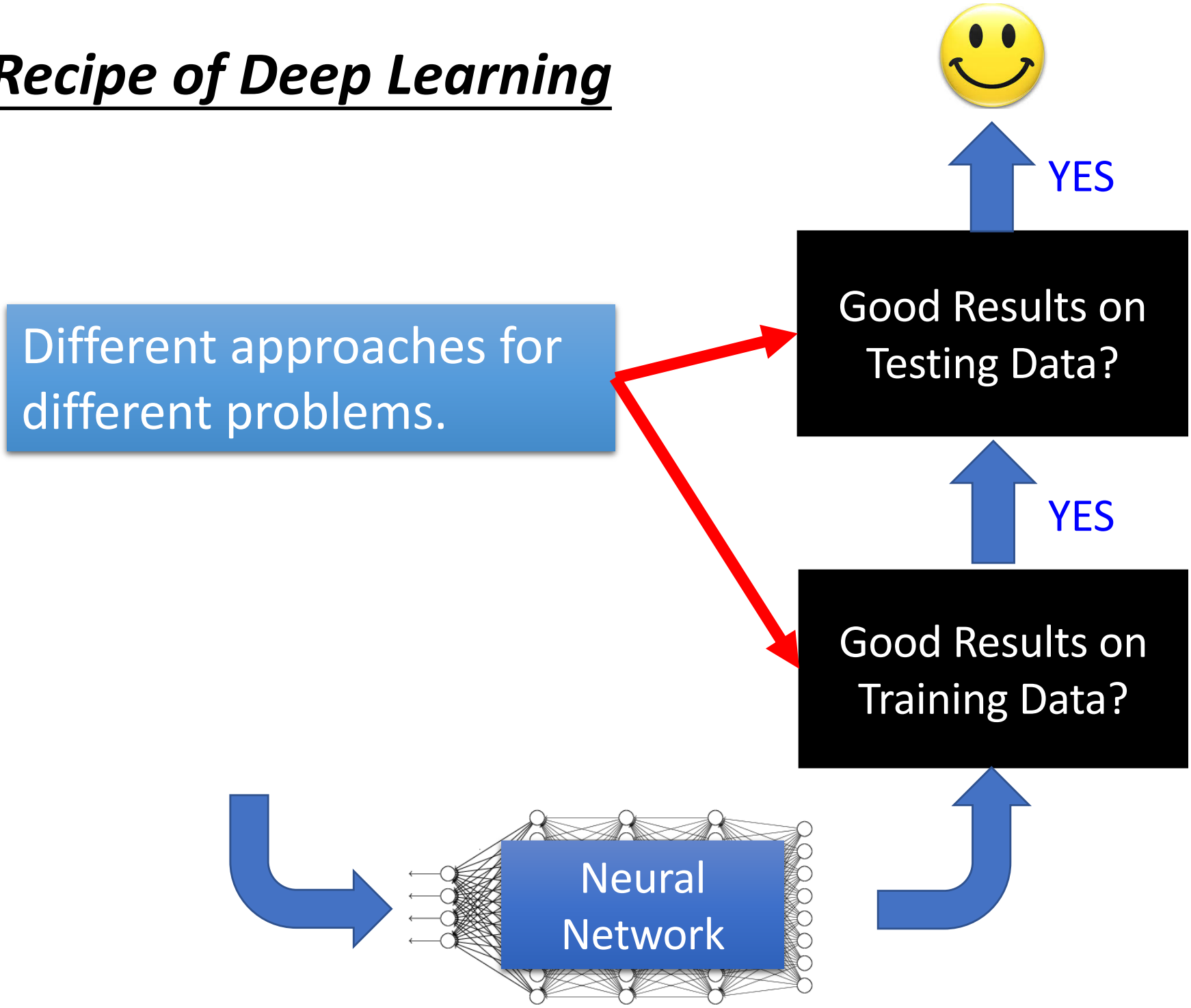
Training Data



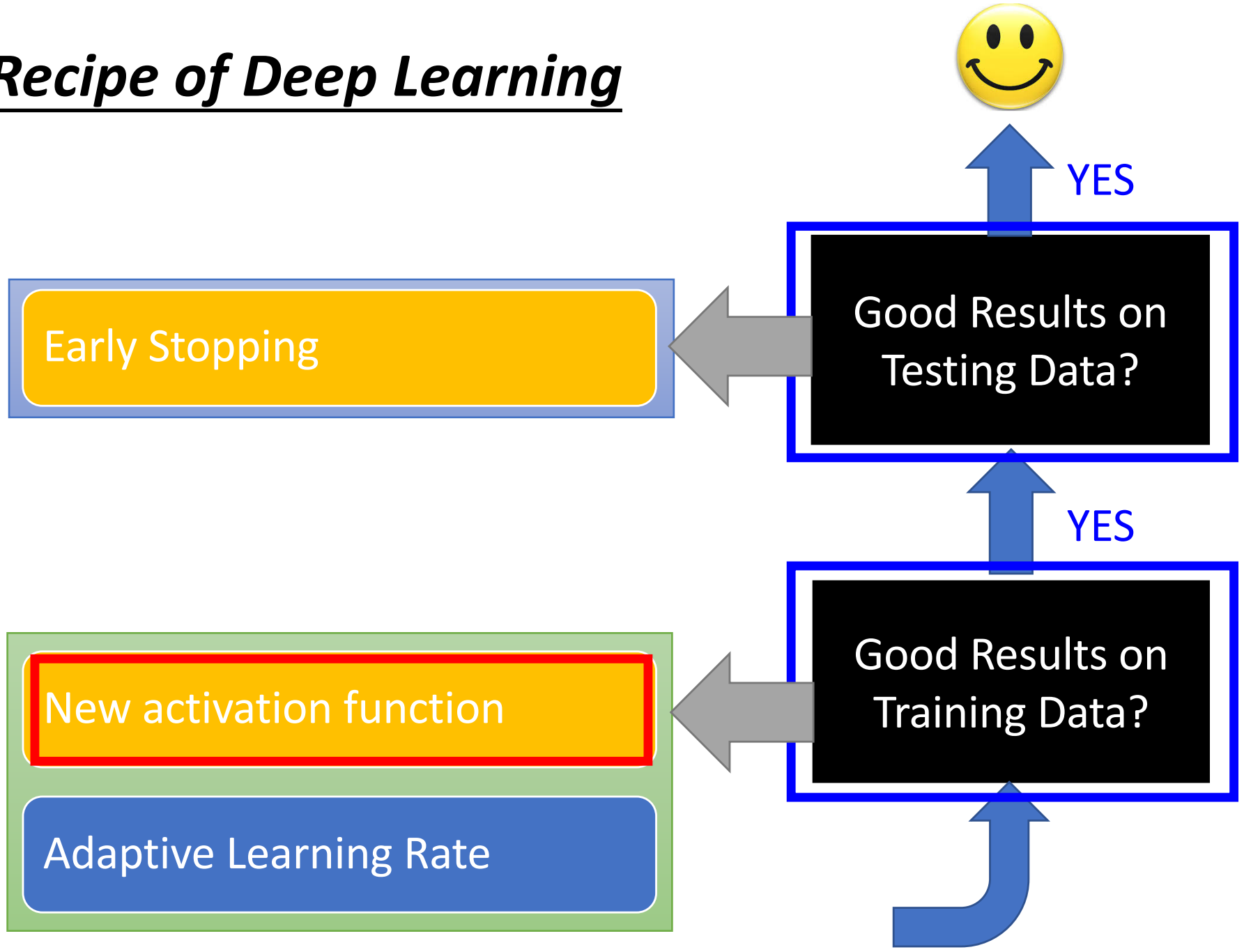
Testing Data

Deep Residual Learning for Image Recognition
http://arxiv.org/abs/1512.03385

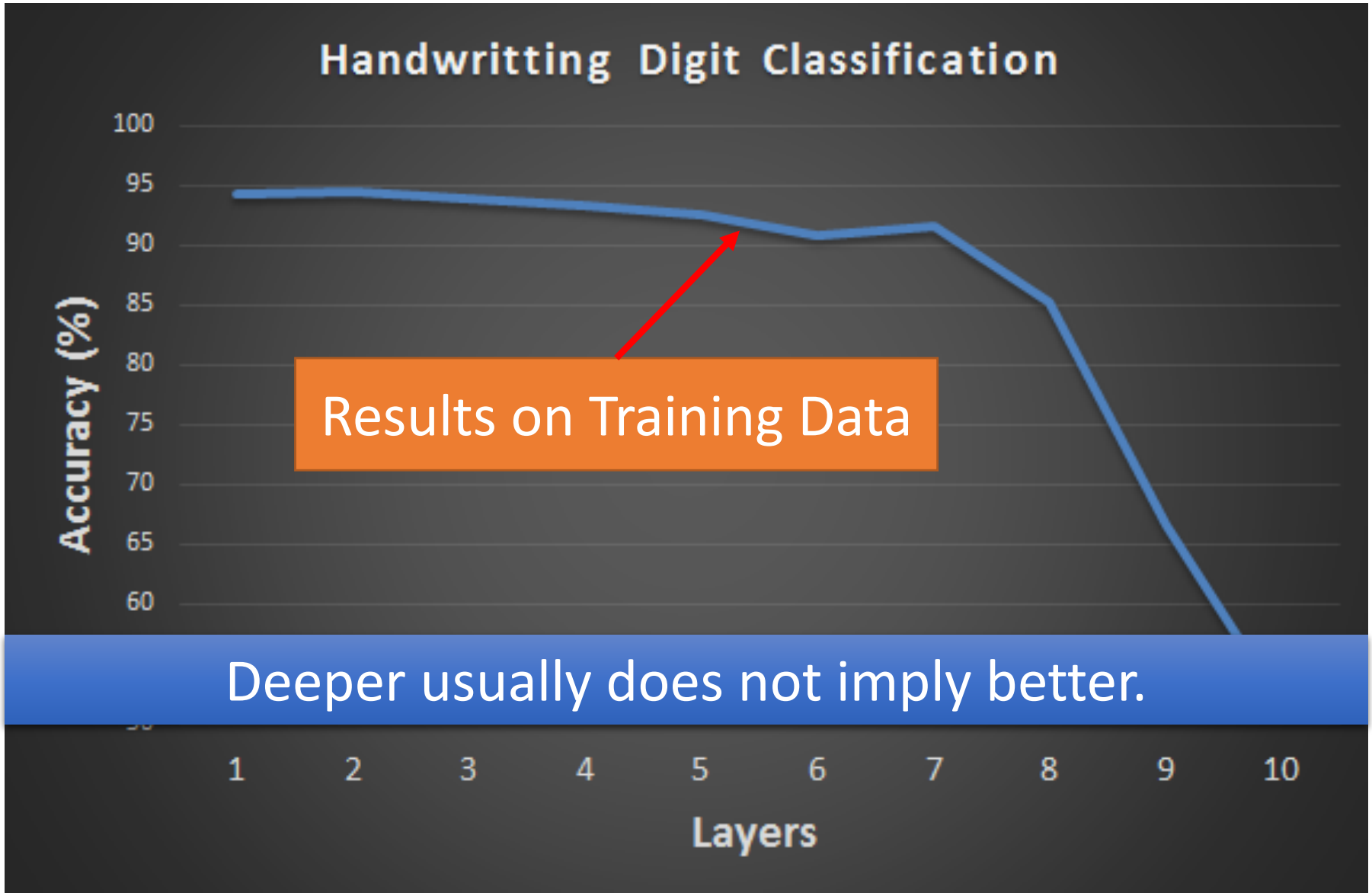
Recipe of Deep Learning



Recipe of Deep Learning

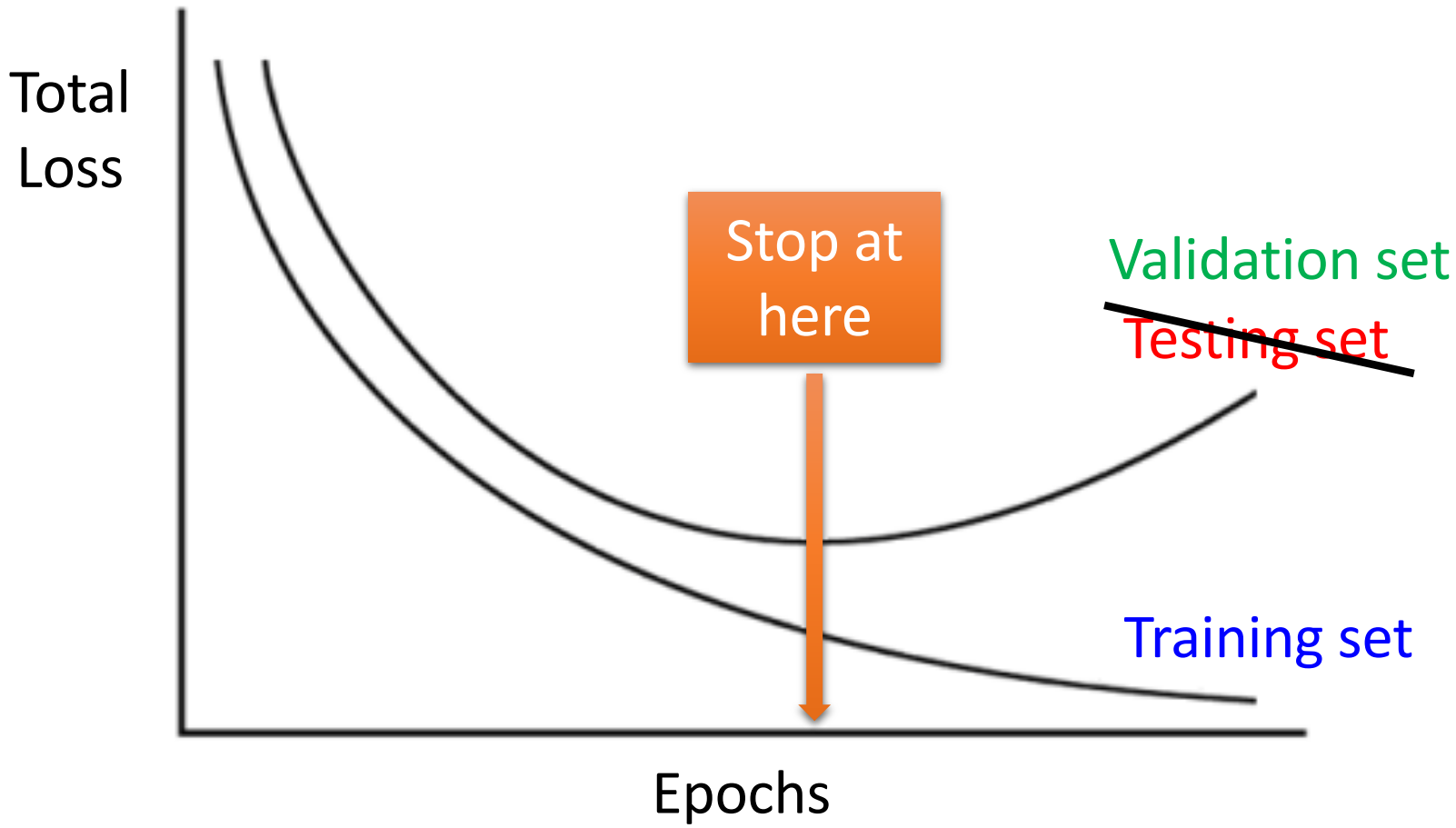


Hard to get the power of Deep …



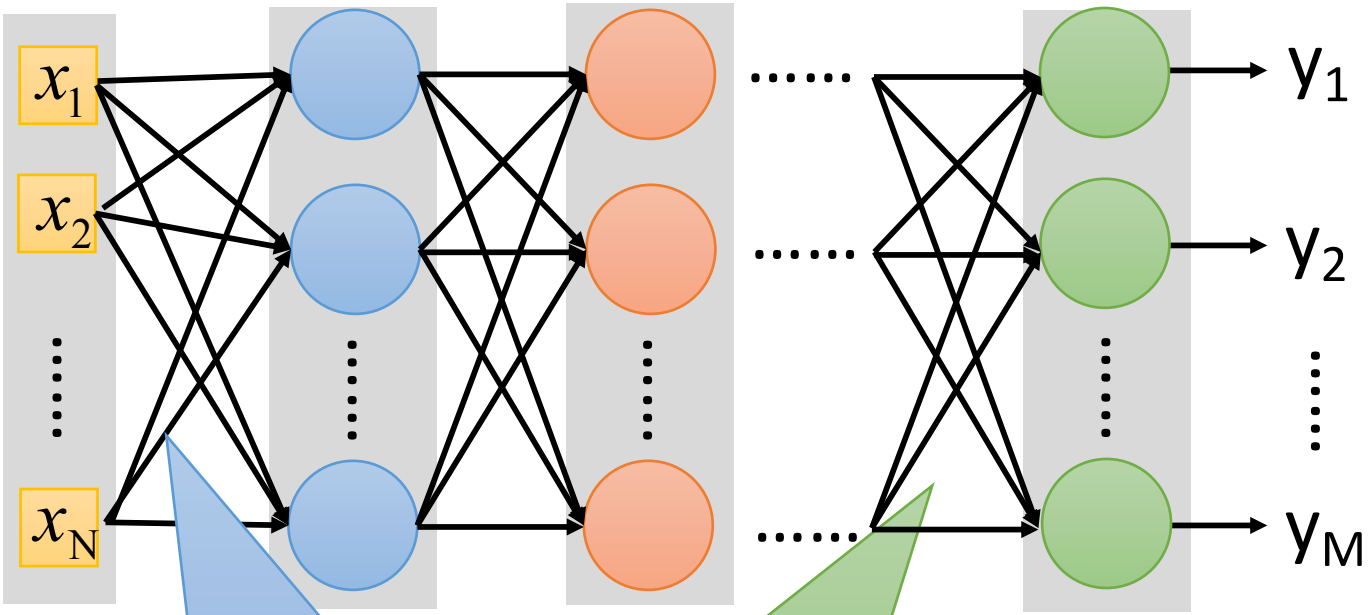
Early Stop

Early Stopping



Activation functions

Vanishing Gradient Problem



Smaller gradients

Learn very slow

Almost random

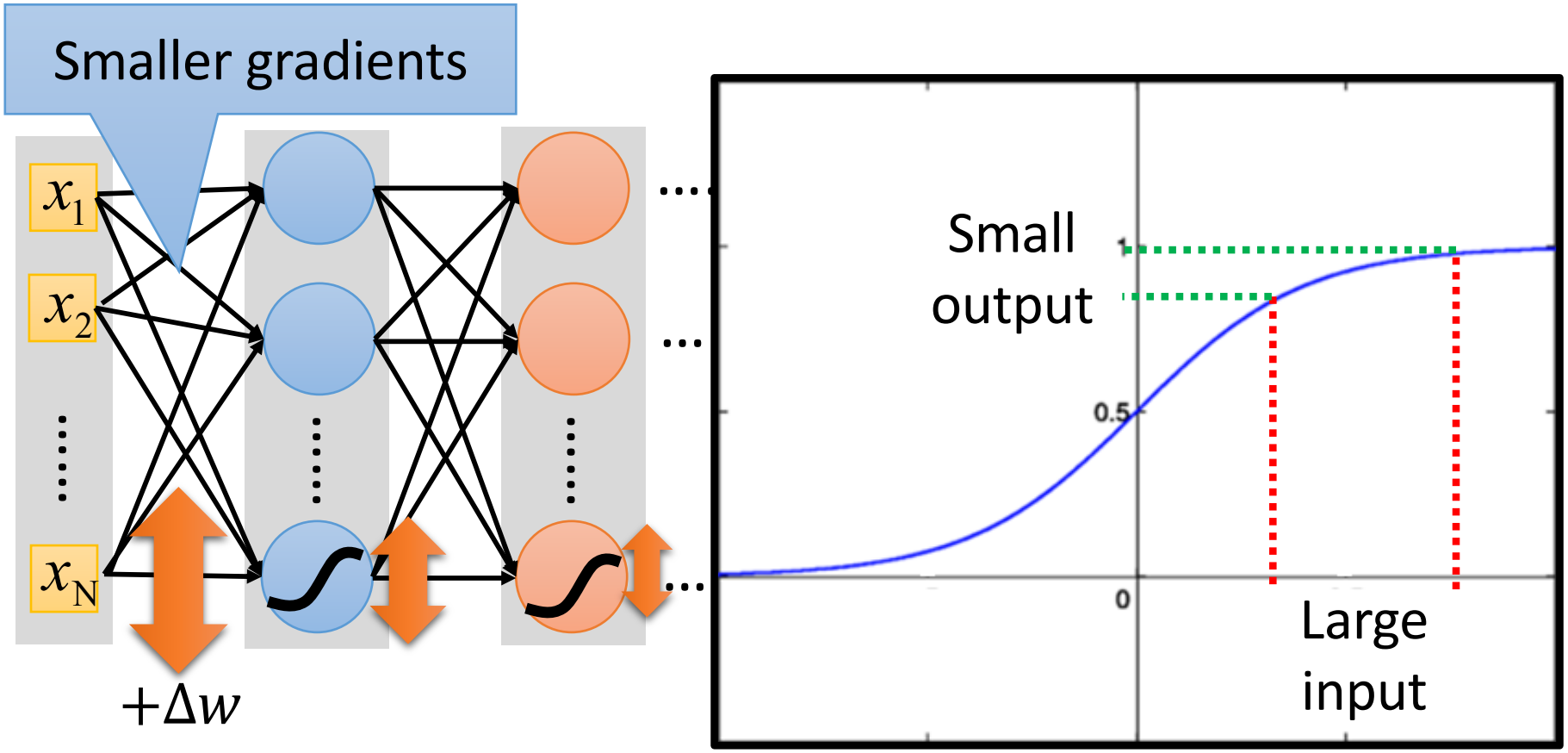
Larger gradients

Learn very fast

Already converge

based on random!?

Vanishing Gradient Problem

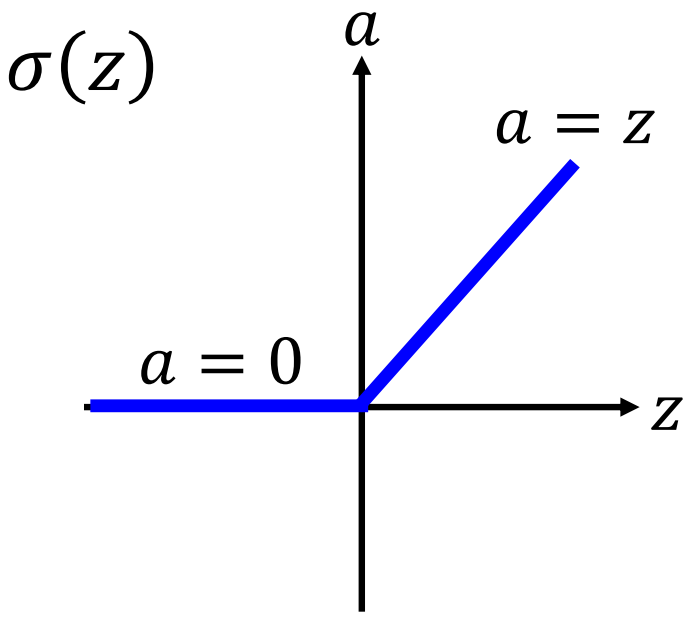


Intuitive way to compute the derivatives …

$$\frac{\partial l}{\partial w} =? \frac{\Delta l}{\Delta w}$$

ReLU

- Rectified Linear Unit (ReLU)



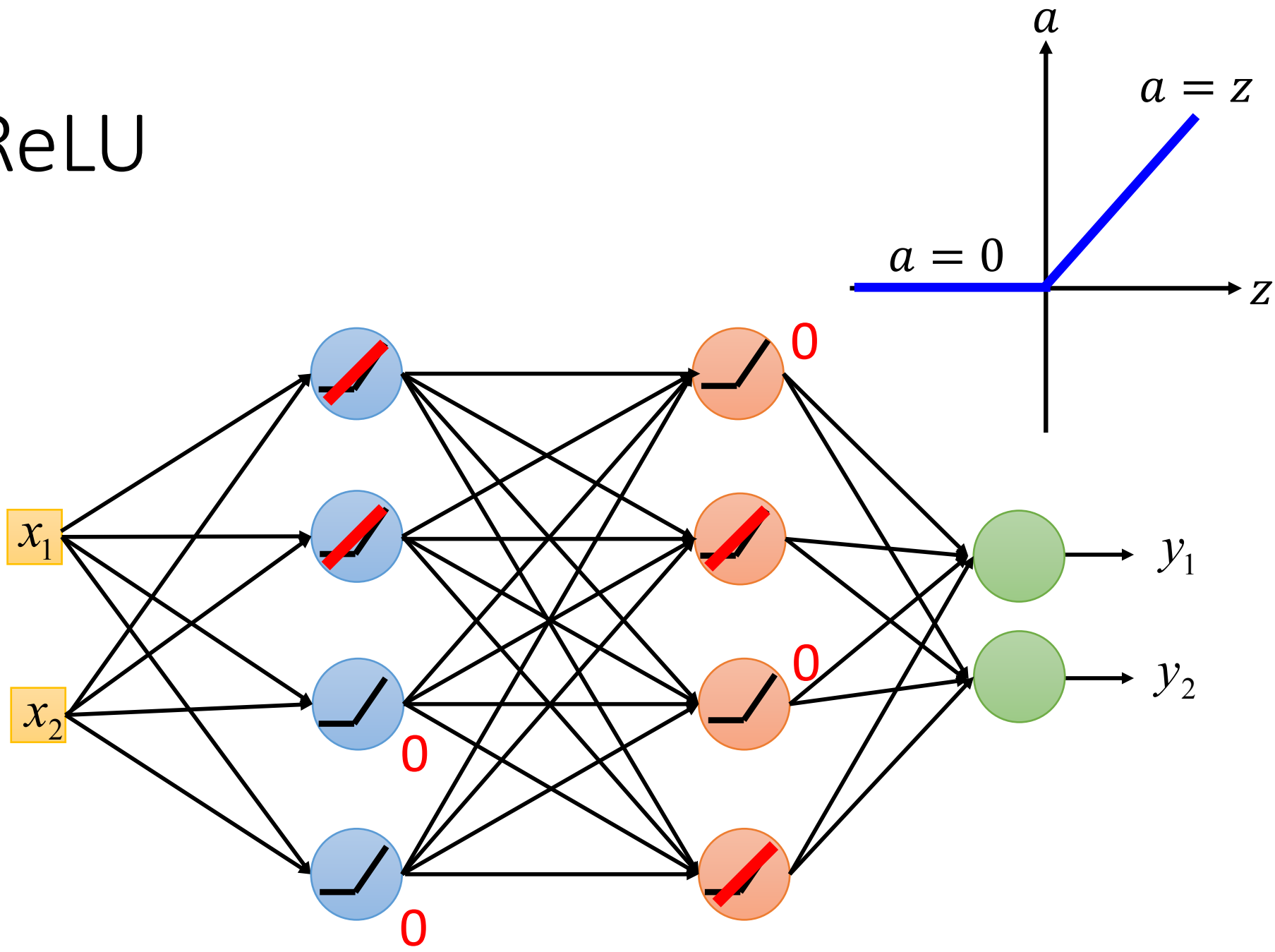
[Xavier Glorot, AISTATS'11]
[Andrew L. Maas, ICML'13]
[Kaiming He, arXiv'15]

Reason:

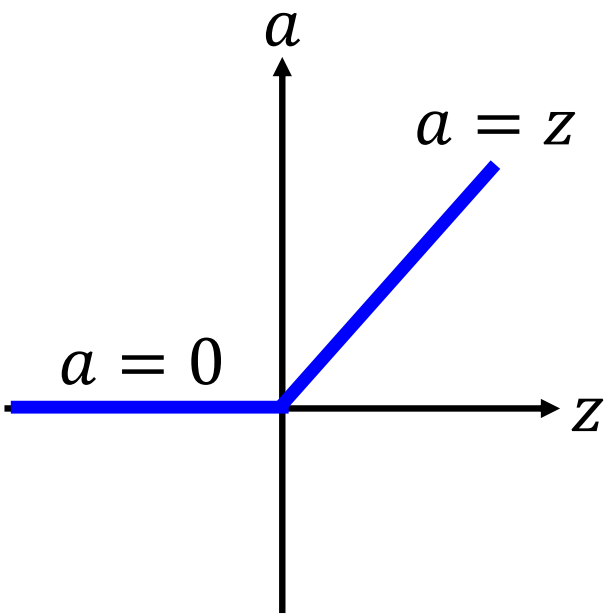
1. Fast to compute

2. Vanishing gradient problem

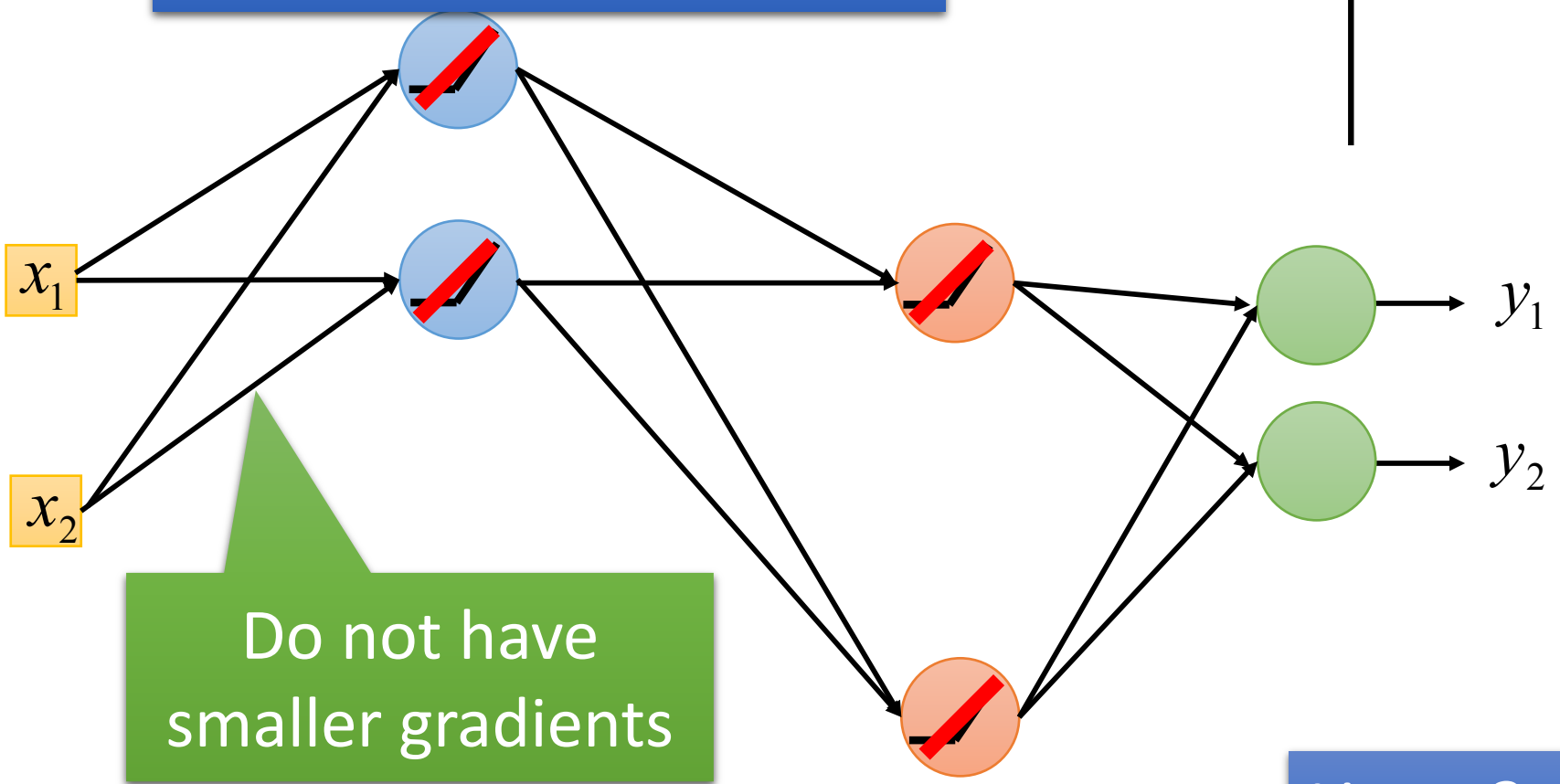
ReLU



ReLU



A Thinner linear network

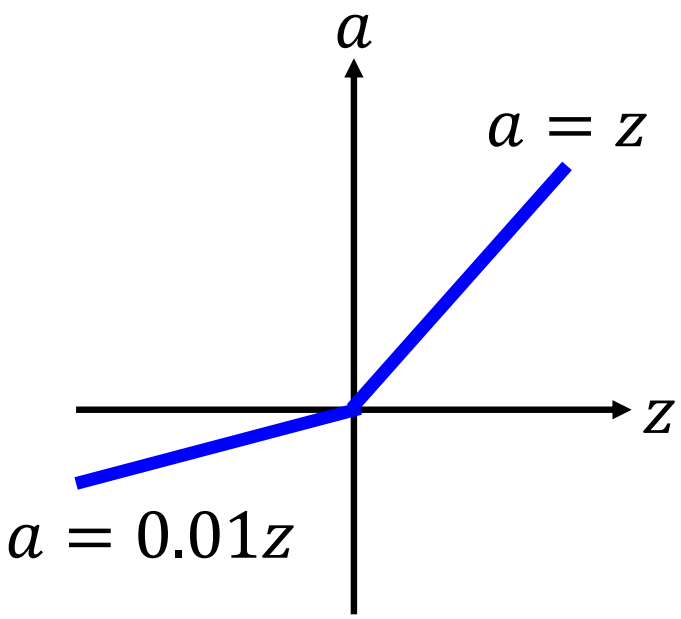


Do not have smaller gradients

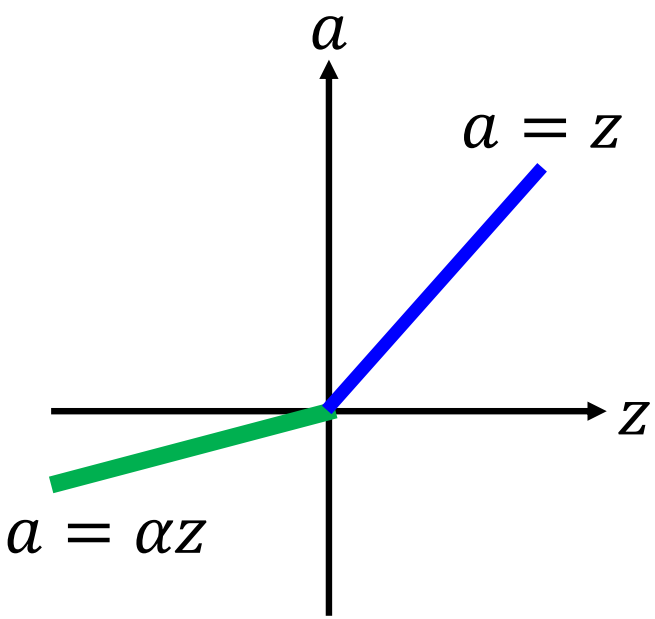
Linear?
Gradient?

ReLU - variant

Leaky ReLU



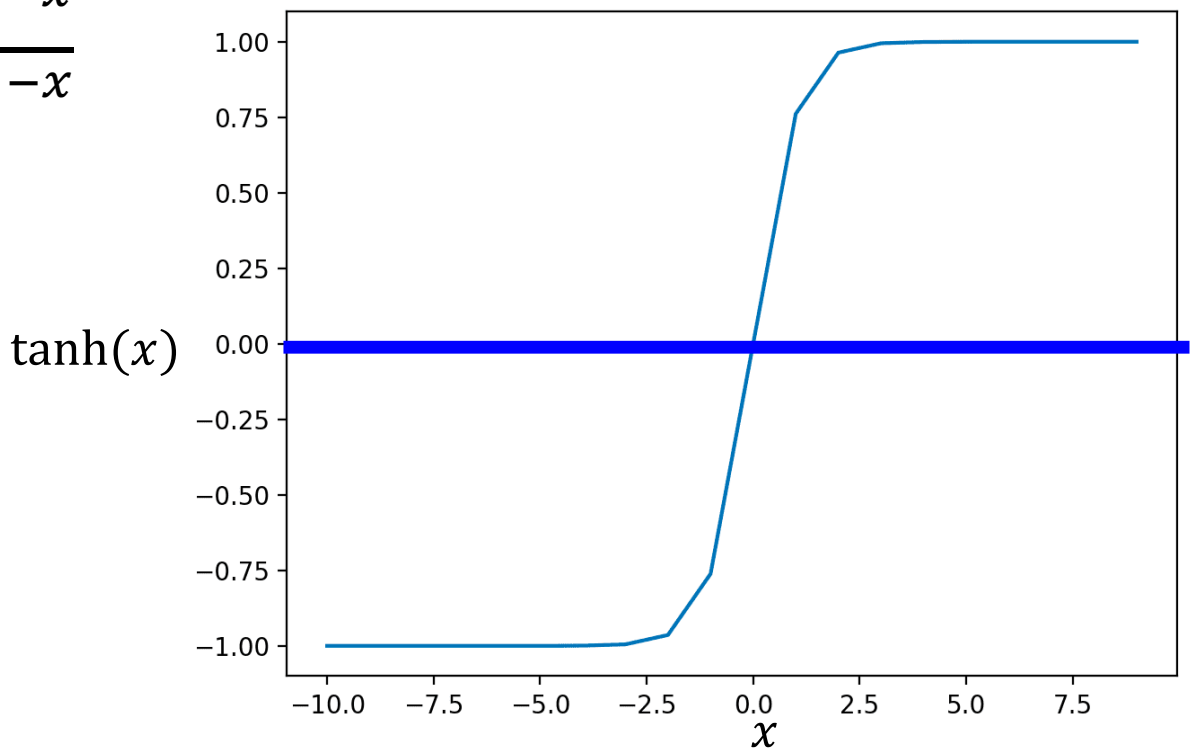
Parametric ReLU



α also learned by gradient descent

tanh (tangent) activation function

- Tanh: The hyperbolic tangent activation function
- Similar to the sigmoid function with an S-shape.
- **Input:** Any real value, **Output:** a value in [-1, 1].
- $\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$

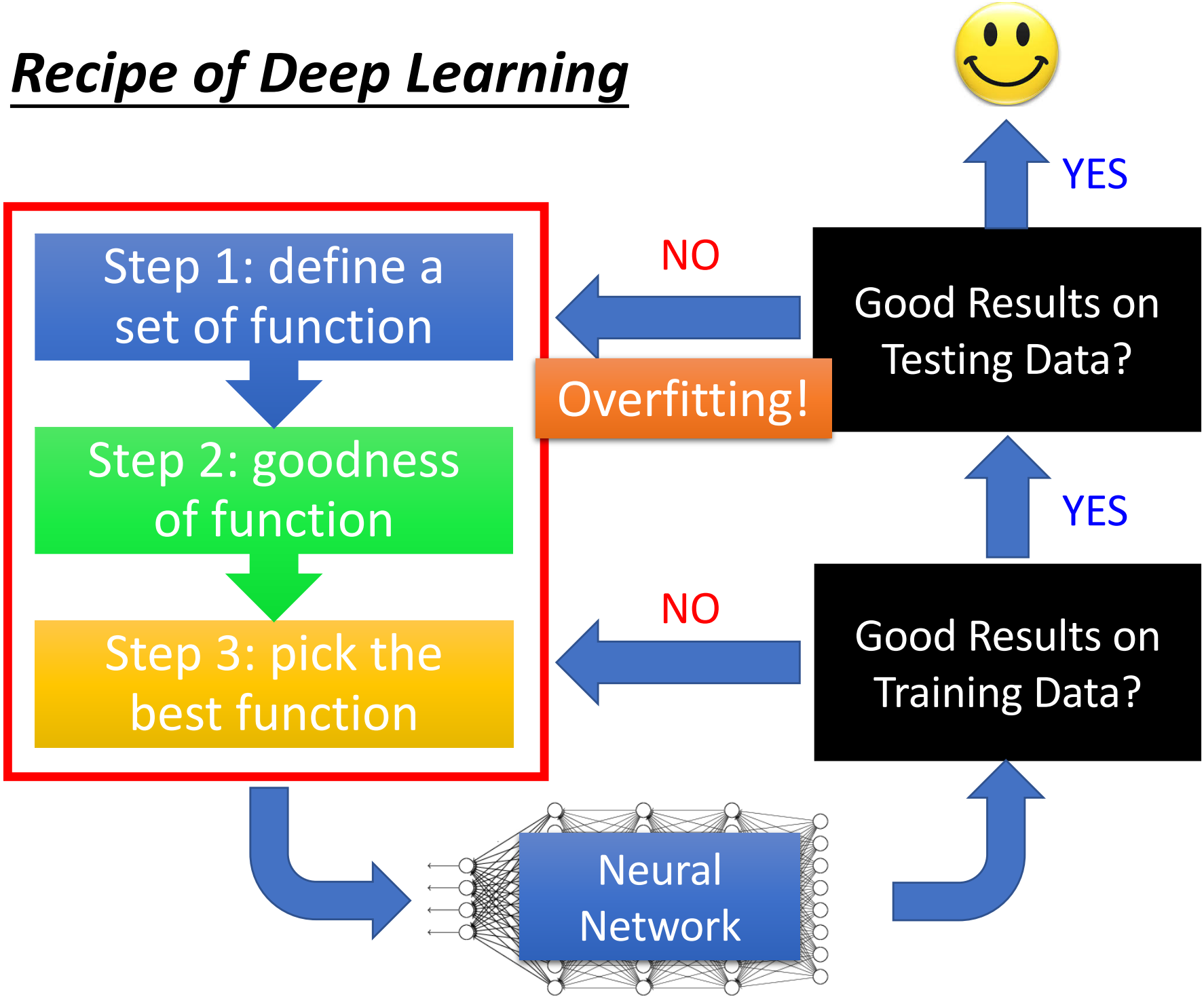


Activation functions in Python

- activation{*'identity', 'logistic', 'tanh', 'relu'}, default='relu'*
- 'identity': no-op activation, f(x) = x
- 'logistic', sigmoid function
- 'tanh', the hyperbolic tan function,
- 'relu', the rectified linear unit function

https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html#sklearn.neural_network.MLPClassifier

Recipe of Deep Learning



Questions