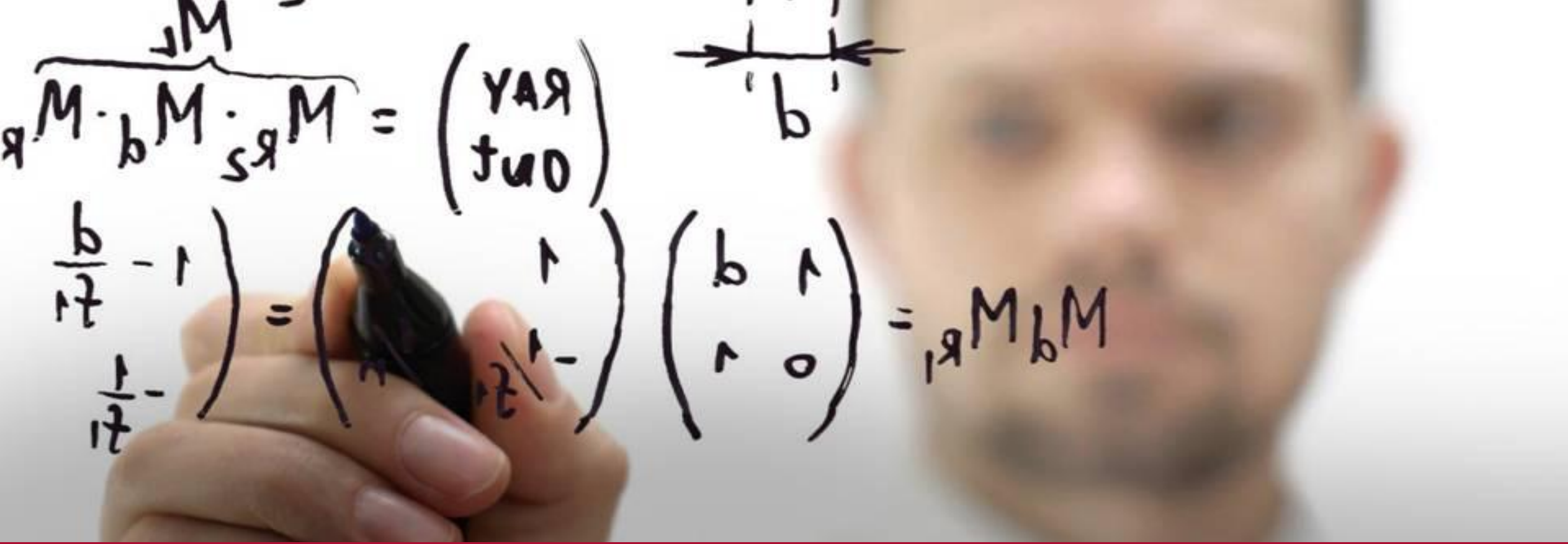
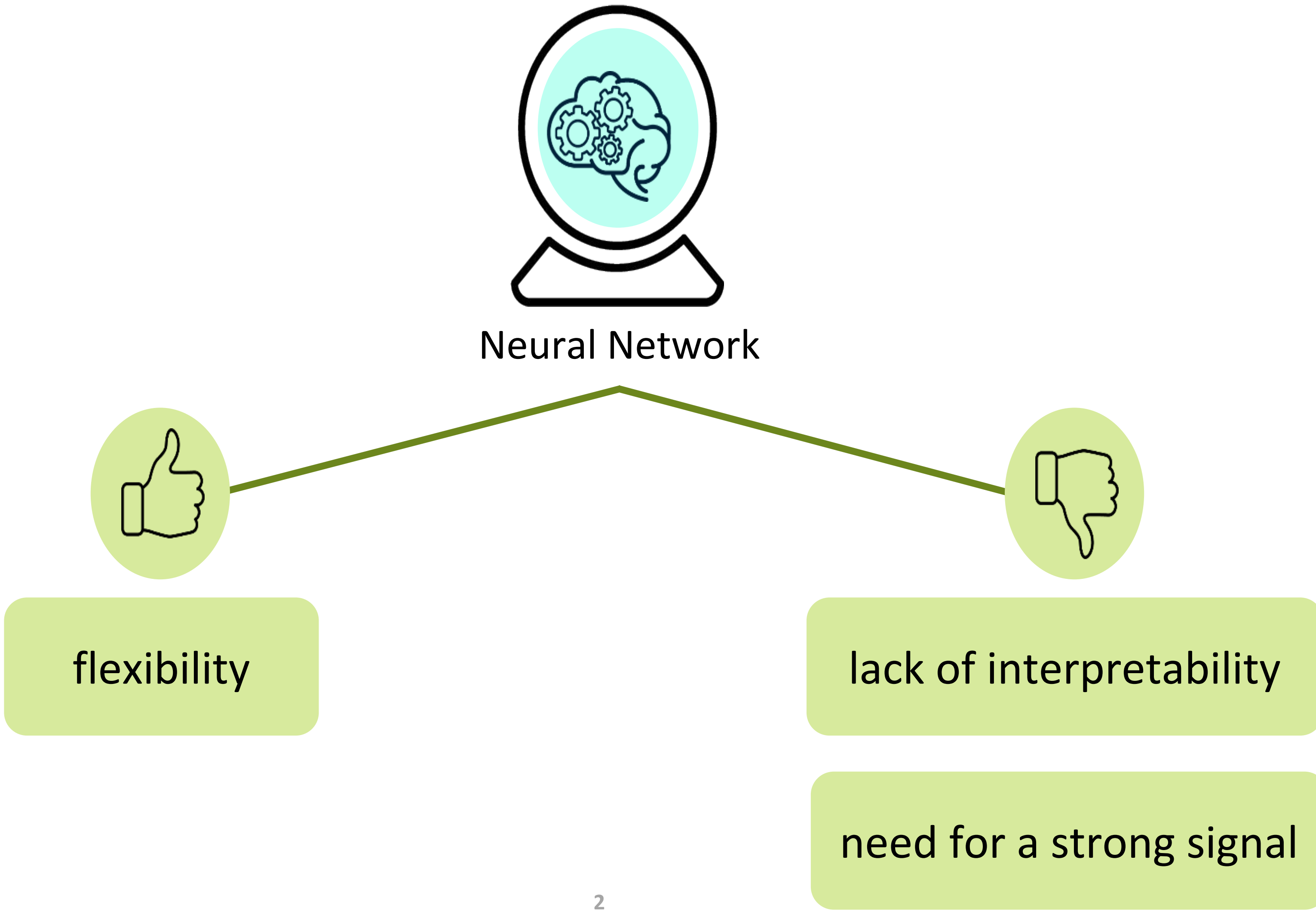


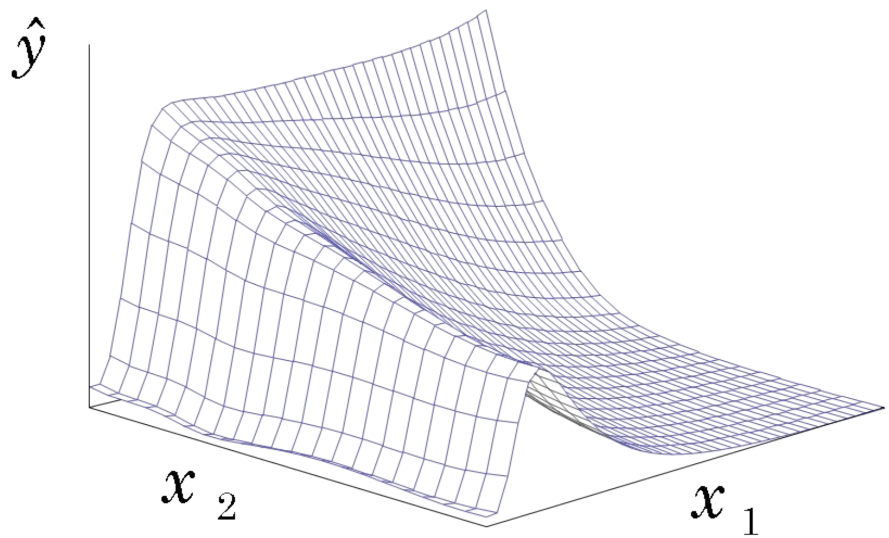
Lesson XII : Neural Network

Advantages and Disadvantages of Neural Networks



Universal Approximator

Given enough neurons and time, a neural network can model any input/output relationship, to any degree of precision.



Neural Network Prediction Formula

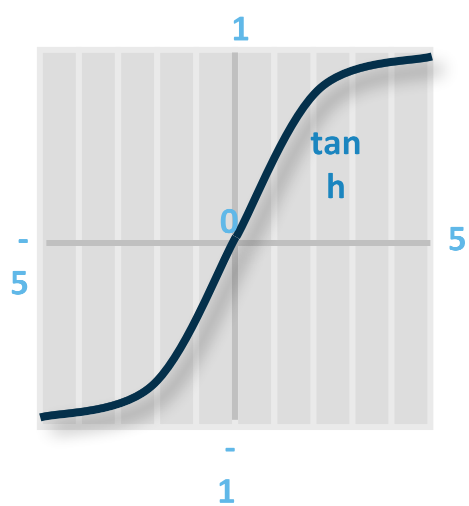
hidden unit

prediction estimate

$$\hat{y} = \hat{w}_{00} + \hat{w}_{01} \cdot H_1 + \hat{w}_{02} \cdot H_2 + \hat{w}_{03} \cdot H_3$$

bias estimate

weight estimate



$$H_1 = \tanh(\hat{w}_{10} + \hat{w}_{11} x_1 + \hat{w}_{12} x_2)$$

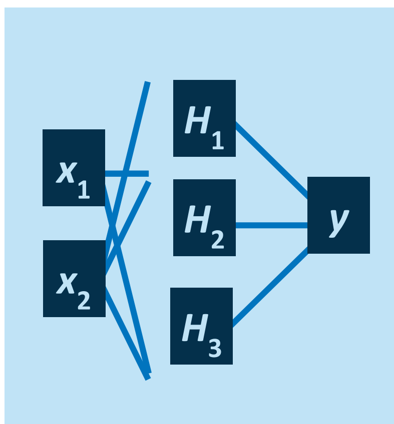
$$H_2 = \tanh(\hat{w}_{20} + \hat{w}_{21} x_1 + \hat{w}_{22} x_2)$$

$$H_3 = \tanh(\hat{w}_{30} + \hat{w}_{31} x_1 + \hat{w}_{32} x_2)$$

activation function

Neural Network Diagram

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \hat{w}_{00} + \hat{w}_{01} \cdot H_1 + \hat{w}_{02} \cdot H_2 + \hat{w}_{03} \cdot H_3$$



$$H_1 = \tanh(\hat{w}_{10} + \hat{w}_{11}\, x_1 + \hat{w}_{12}\, x_2)$$

$$H_2 = \tanh(\hat{w}_{20} + \hat{w}_{21}\, x_1 + \hat{w}_{22}\, x_2)$$

$$H_3 = \tanh(\hat{w}_{30} + \hat{w}_{31}\, x_1 + \hat{w}_{32}\, x_2)$$

input layer | hidden layer | target layer

Prediction Illustration: Neural Networks

1. Standardize (scale) the input variables.

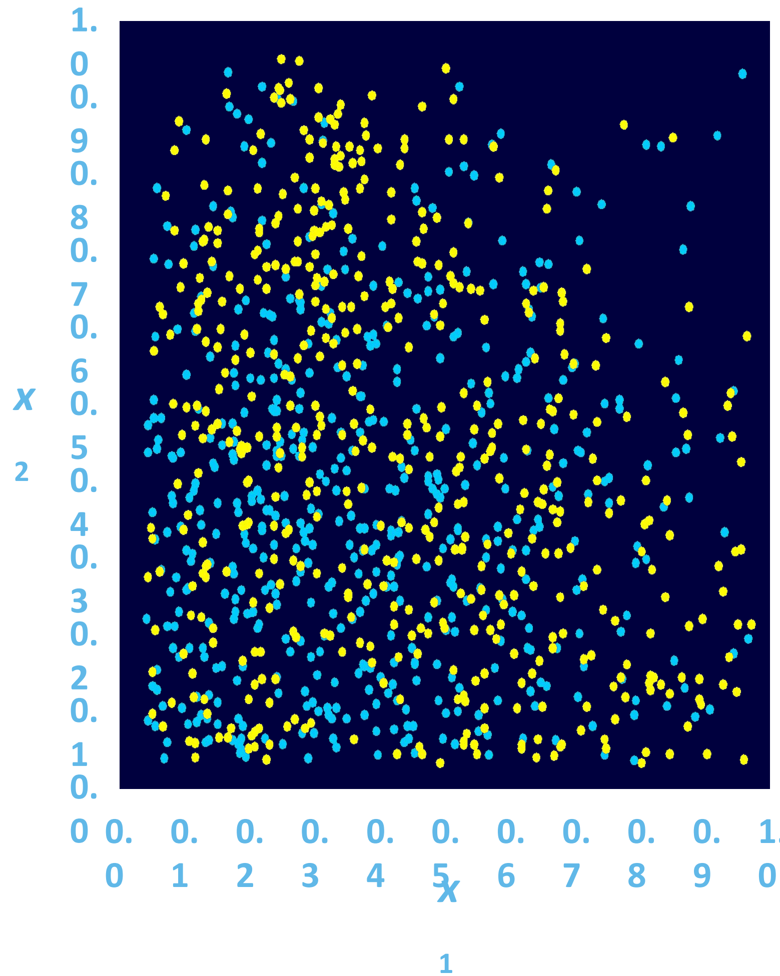
logit

$$\mathrm{logit}(\hat{p}) = \hat{w}_{00} + \hat{w}_{01} \cdot H_1 + \hat{w}_{02} \cdot H_2 + \hat{w}_{03} \cdot H_3$$

$$H_1 = \tanh(\hat{w}_{10} + \hat{w}_{11}\, x_1 + \hat{w}_{12}\, x_2)$$

$$H_2 = \tanh(\hat{w}_{20} + \hat{w}_{21}\, x_1 + \hat{w}_{22}\, x_2)$$

$$H_3 = \tanh(\hat{w}_{30} + \hat{w}_{31}\, x_1 + \hat{w}_{32}\, x_2)$$



Standardization Methods

Midrange

- $midrange = \frac{(Max+Min)}{2}$
- $x_{midrange} = \frac{(x-midrange)}{range/2}$
- Midrange is 0. Half range is 1.

Z-Score

- $\mu = 0$ and $\sigma = 1$
- $x_{std} = z = \frac{x-\mu}{\sigma}$

Standardization can be defined for hidden and target layers.

Prediction Illustration: Neural Networks

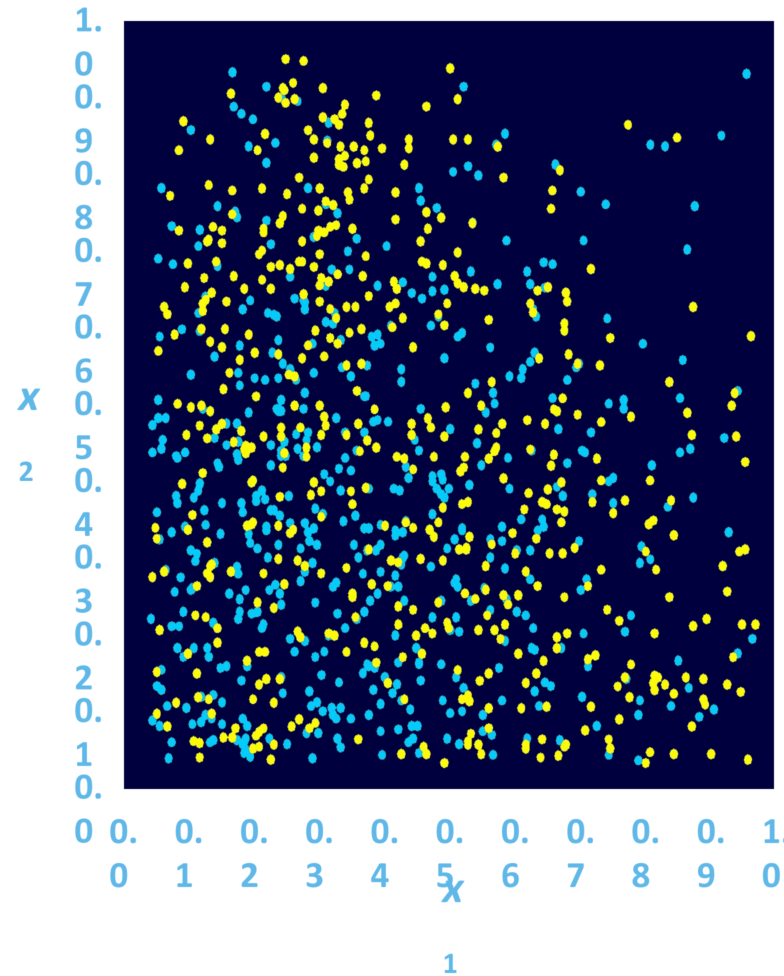
2. Find the weight estimates.

$$\text{logit}(\hat{p}) = \boxed{\hat{w}_{00}} + \boxed{\hat{w}_{01}} H_1 + \boxed{\hat{w}_{02}} H_2 + \boxed{\hat{w}_{03}} H_3$$

$$H_1 = \tanh(\boxed{\hat{w}_{10}} + \boxed{\hat{w}_{11}} x_1 + \boxed{\hat{w}_{12}} x_2)$$

$$H_2 = \tanh(\boxed{\hat{w}_{20}} + \boxed{\hat{w}_{21}} x_1 + \boxed{\hat{w}_{22}} x_2)$$

$$H_3 = \tanh(\boxed{\hat{w}_{30}} + \boxed{\hat{w}_{31}} x_1 + \boxed{\hat{w}_{32}} x_2)$$



Prediction Illustration: Neural Networks

2. Find the weight estimates.

$$\text{logit}(\hat{p}) = \hat{w}_{00} + \hat{w}_{01} H_1 + \hat{w}_{02} H_2 + \hat{w}_{03} H_3$$

$$H_1 = \tanh(\hat{w}_{10} + \hat{w}_{11} x_1 + \hat{w}_{12} x_2)$$

$$H_2 = \tanh(\hat{w}_{20} + \hat{w}_{21} x_1 + \hat{w}_{22} x_2)$$

$$H_3 = \tanh(\hat{w}_{30} + \hat{w}_{31} x_1 + \hat{w}_{32} x_2)$$

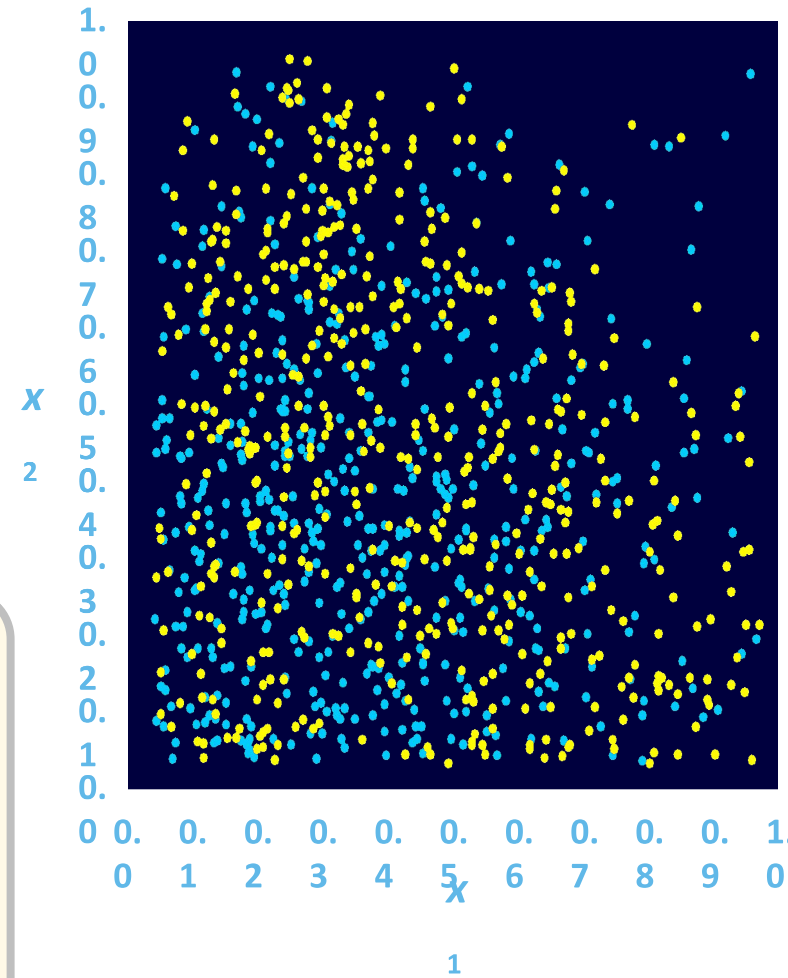
Binary Target

minimize

$$-2 \cdot \Big[\overbrace{\sum \log(\hat{p}_i) + \sum \log(1 - \hat{p}_i)}^{\text{log-likelihood function}} \Big]$$

primary outcome training cases

secondary outcome training cases



Prediction Illustration: Neural Networks

3. Obtain a prediction.

$$\text{logit}(\hat{p}) = -0.5 + -2.6\ H_1 + -1.9\ H_2 + -0.63\ H_3$$

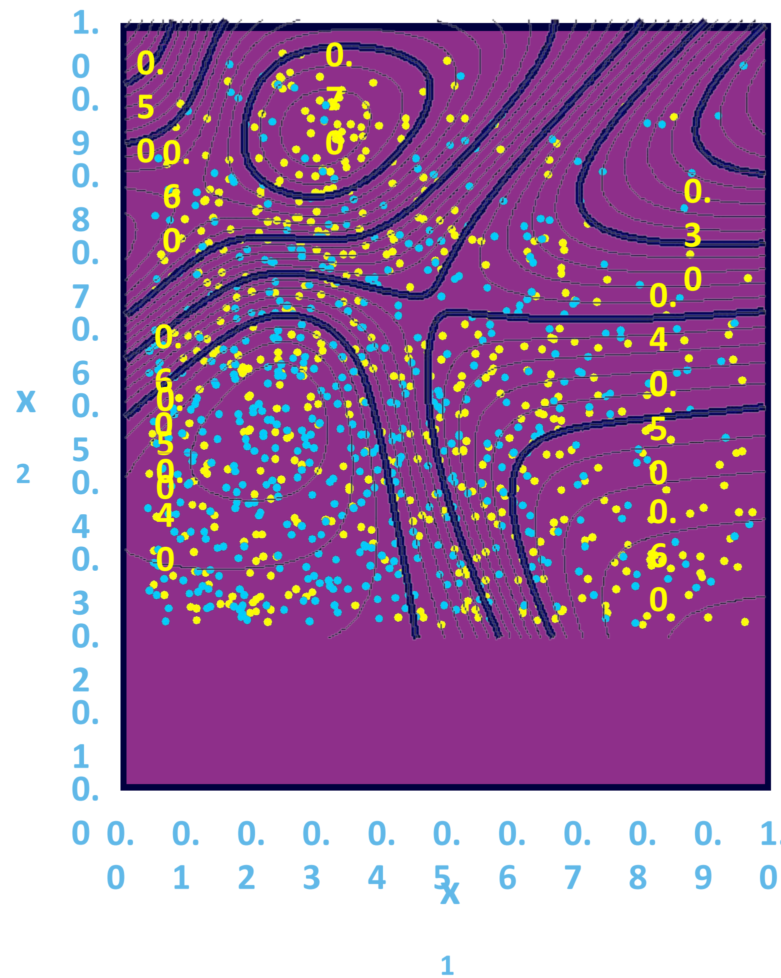
$$H_1 = \tanh(-1.8 + 0.25\ x_1 + -1.8\ x_2)$$

$$H_2 = \tanh(\ 2.7 + 2.7\ x_1 + -5.3\ x_2)$$

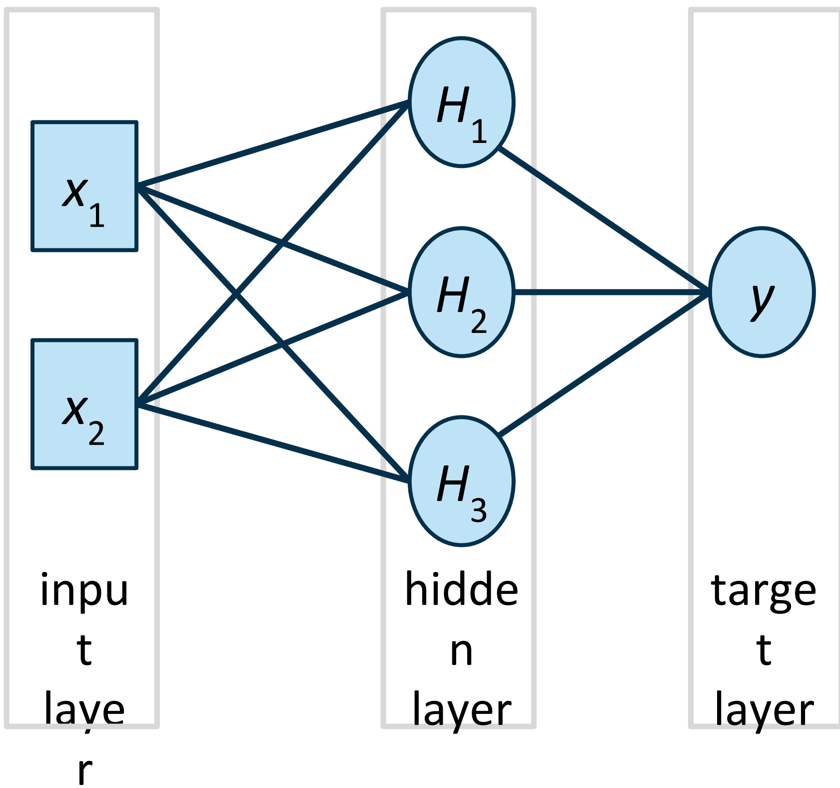
$$H_3 = \tanh(-5.0 + 8.1\ x_1 + 4.3\ x_2)$$

Logistic Function

$$\hat{p} = \frac{1}{1 + e^{-\text{logit}(\hat{p})}}$$



Network Architecture



different connection types

number of layers

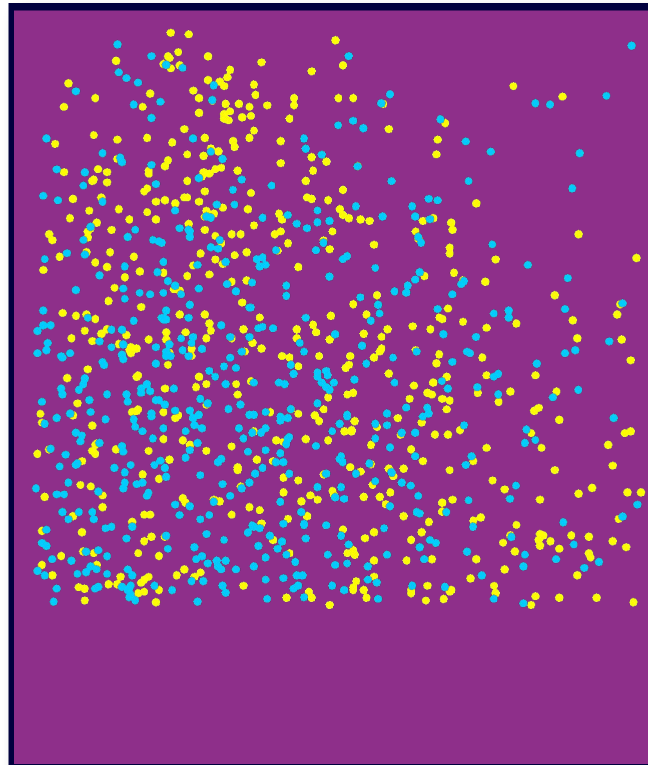
activation functions

number of neurons in each layer

Early Stopping

Initial hidden unit weights

$$\text{logit}(\hat{p}) = 0 + 0 \cdot H_1 + 0 \cdot H_2 + 0 \cdot H_3$$



Early Stopping

Initial hidden unit weights

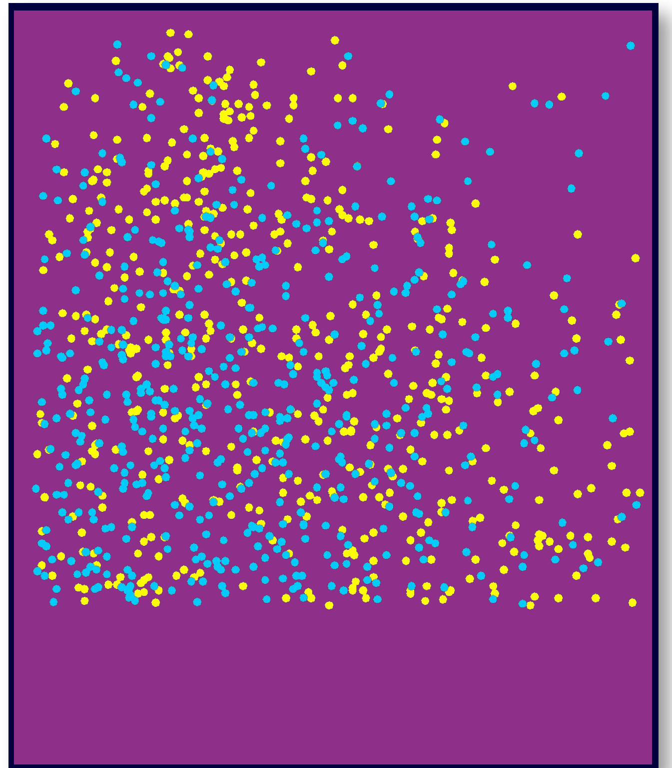
logit($\hat{p}$) = $0 + 0 \cdot H_1 + 0 \cdot H_2 + 0 \cdot H_3$

$H_1 = \tanh(-1.5 - .03x_1 - .07x_2)$

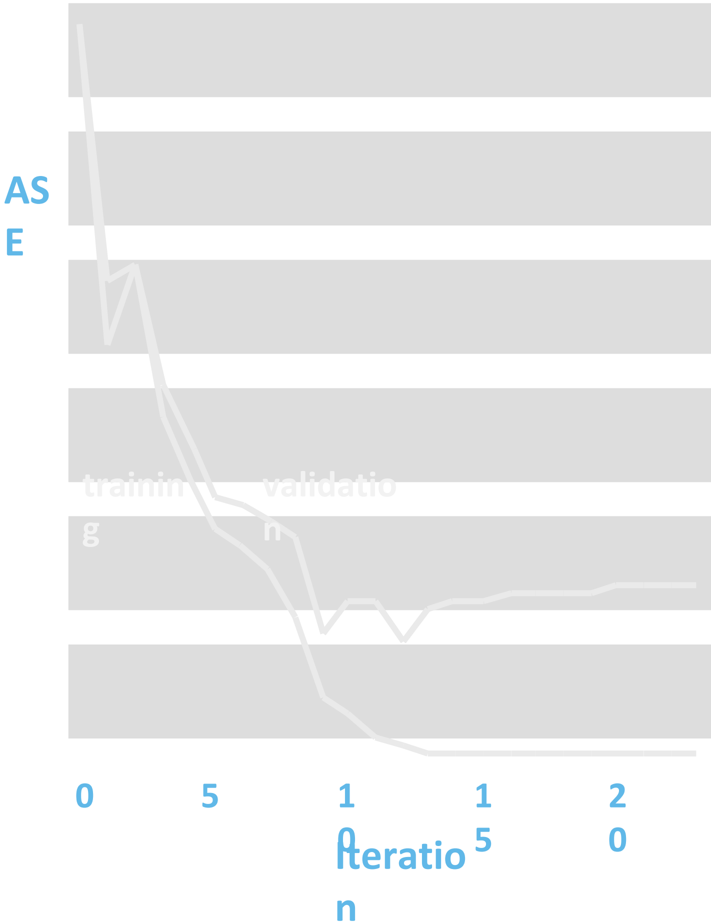
$H_2 = \tanh(.79 - .17x_1 - .16x_2)$

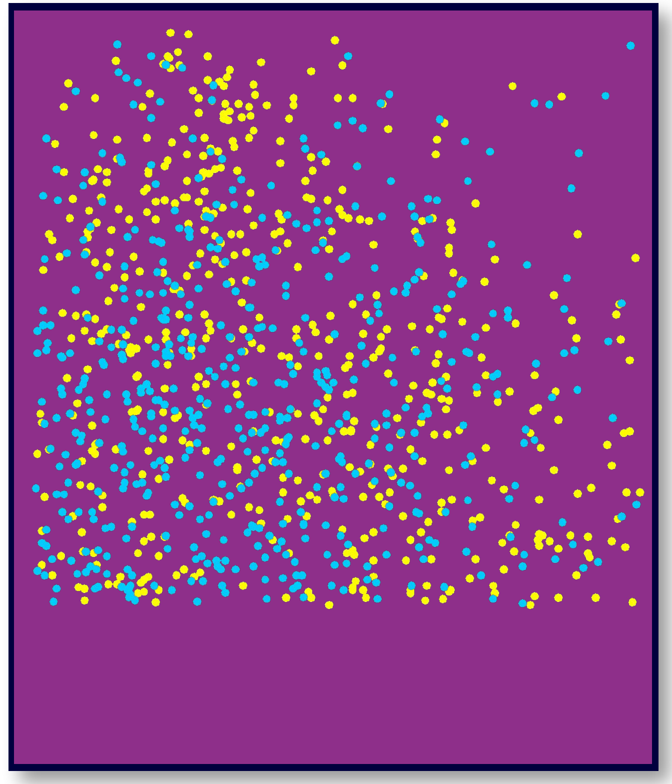
$H_3 = \tanh(.57 + .05x_1 +.35x_2)$

Random initial input weights and biases

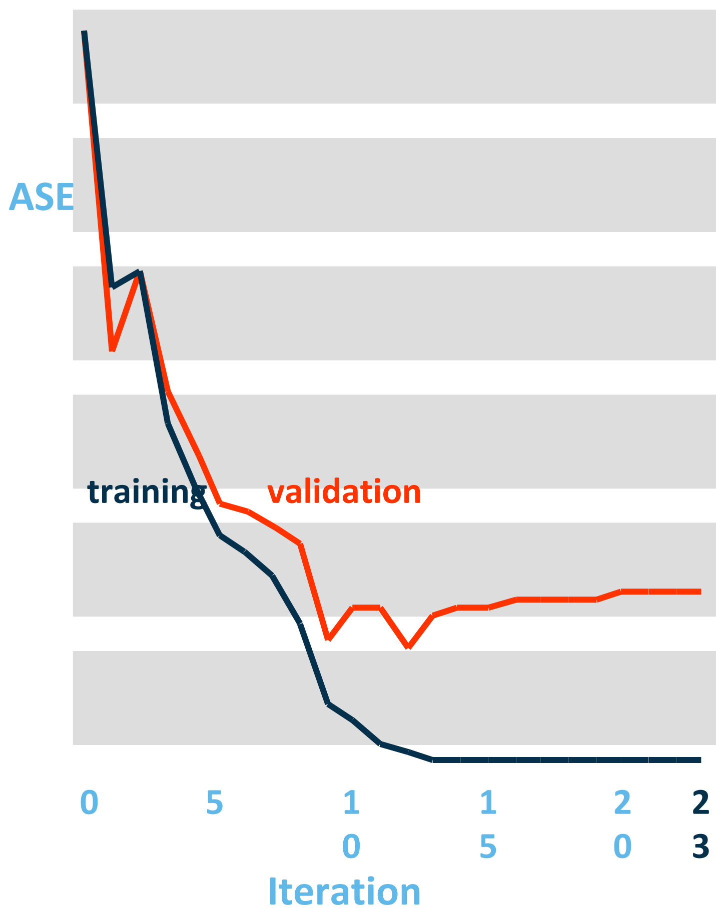


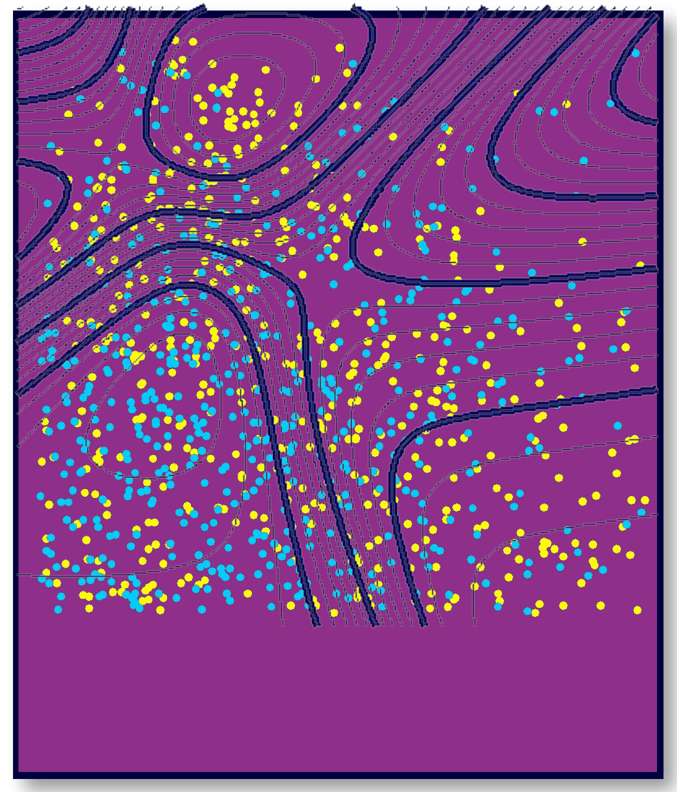
Early Stopping



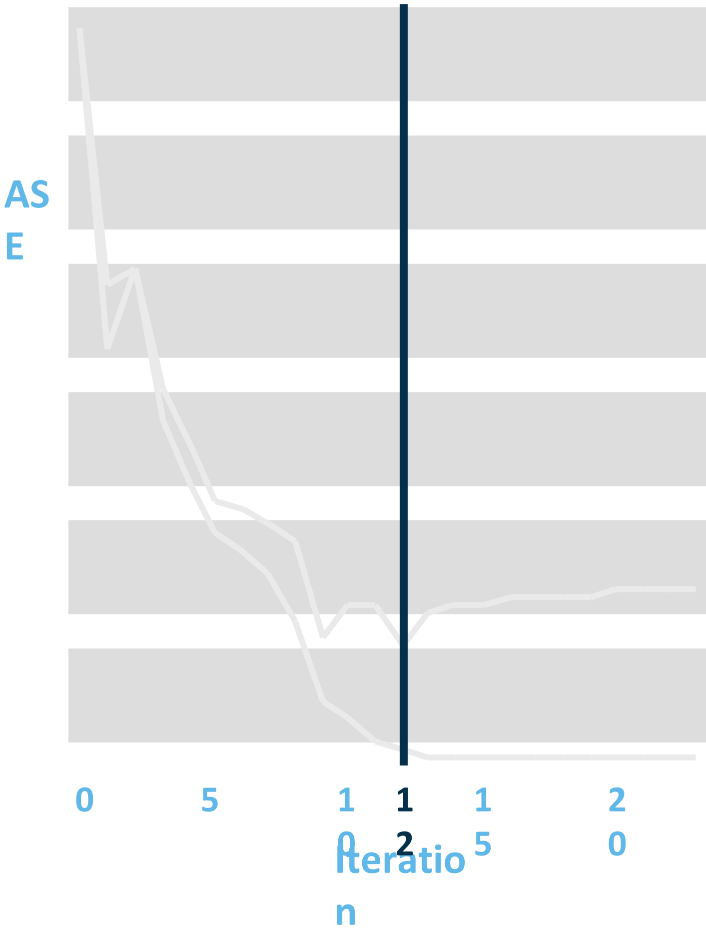


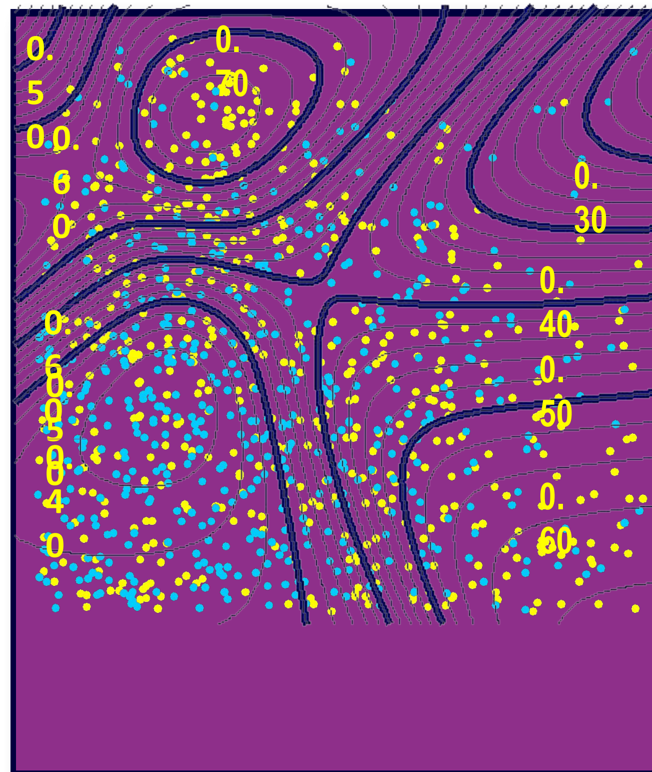
Early Stopping





Early Stopping





Parameter Estimation: Example

$$\text{logit}(\hat{p}) = \hat{w}_{00} + \hat{w}_{01} H_1 + \hat{w}_{02} H_2 + \hat{w}_{03} H_3$$

$$H_1 = \tanh(\hat{w}_{10} + \hat{w}_{11} x_1 + \hat{w}_{12} x_2)$$

$$H_2 = \tanh(\hat{w}_{20} + \hat{w}_{21} x_1 + \hat{w}_{22} x_2)$$

$$H_3 = \tanh(\hat{w}_{30} + \hat{w}_{31} x_1 + \hat{w}_{32} x_2)$$

The error function is always a deviance function.

Binary Target

minimize

$$-2 \cdot \left[\overbrace{\sum \log(\hat{p}_i) + \sum \log(1 - \hat{p}_i)}^{\text{log-likelihood function}} \right]$$

primary outcome training cases

secondary outcome training cases

Bernoulli function

Construct a NN using SKlearn

ADSUP Lesson 11 SVM

1- Construct a NN model for Organics Dataset

Practice

This practice reinforces the concepts in relation to neural network modeling. You use Model Studio to build a neural network model based on hyperparameter autotune.

Science for real life

IDS Analytics is a firm of experts specialized in applying quantitative methods to growth and optimization issues of enterprises, particularly in risk, revenue management and marketing domains.

We help enterprises both substantially improve commercial success and build capacity to repeat this performance continuously. In true "scientific" spirit, IDS analytics provides a comprehensive and complementing range of services across management advisory, capability building and managed services disciplines.

www.ids-analytics.com

OFİS
Telefon +90 (216) 343 6623
Tophanelioğlu Cad. Korukent Köşk
No:6/5 Koşuyolu
İstanbul TÜRKİYE