Artificial Neural Network

Peng Zhichao 1155062015

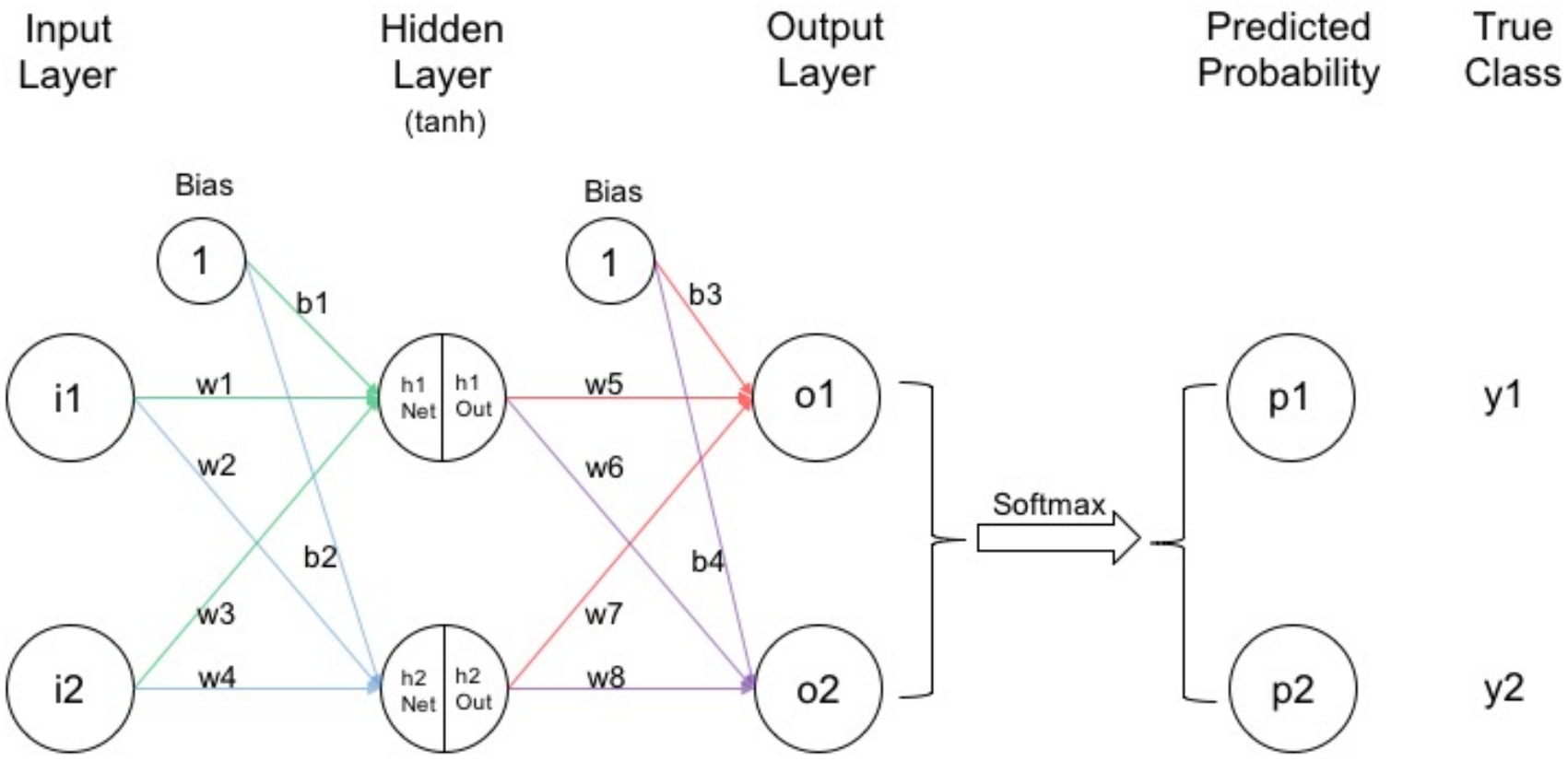
Project 1: Fraud detection of insurance claims

October, 2018

Outline

- Structure of Neural Network
- Activation Function
- Loss Function with L1 penalty
- Backpropagation
- Stochastic Gradient Descent and Mini-batch Gradient Descent
- Model Tuning and Parameter selection
- Result

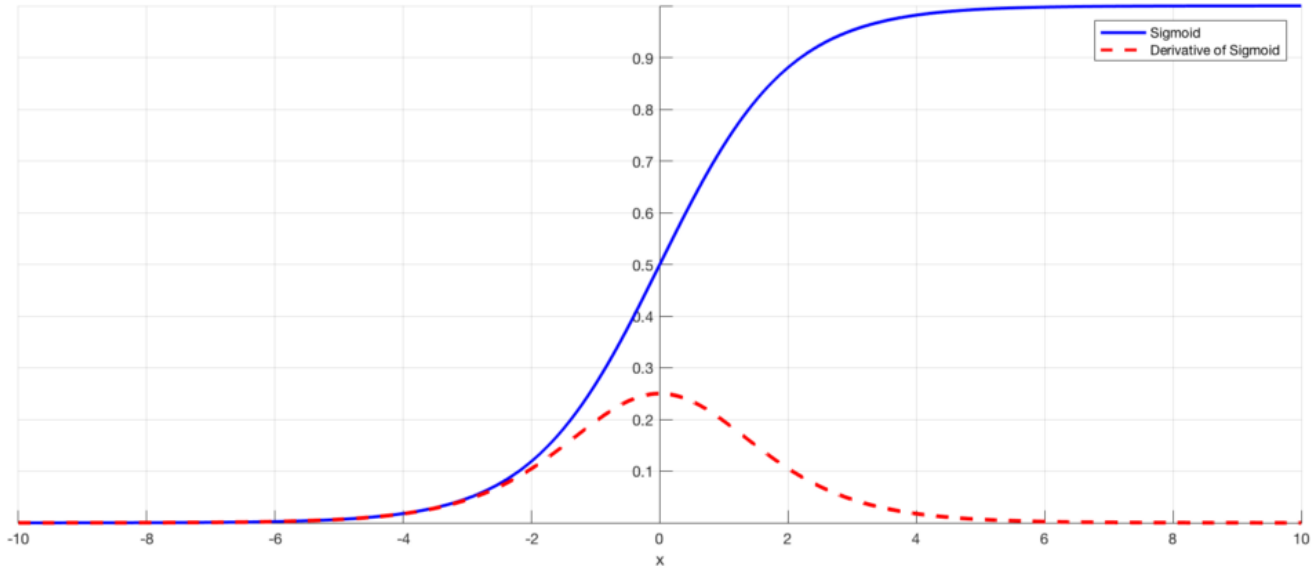
Structure of Neural Network

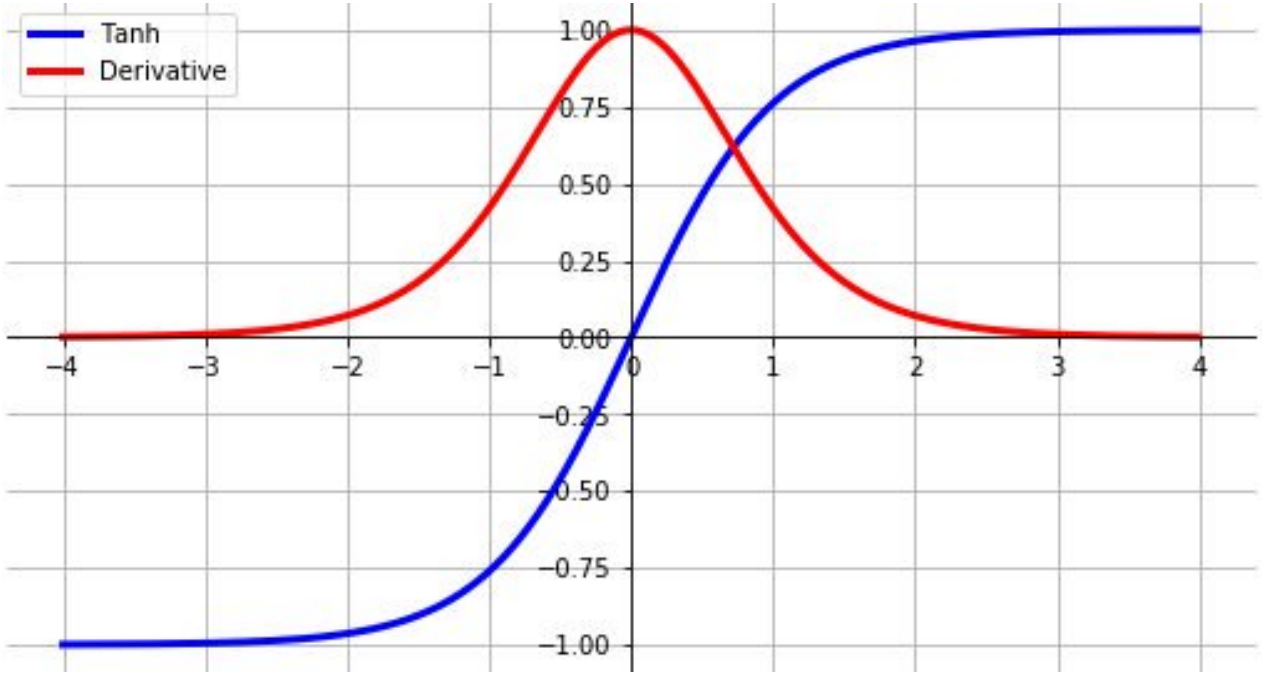


Activation Function

Name	Function	Domain	Range
Logistic	$f(x) = \frac{1}{1+e^{-x}}$	$\mathbb{R}$	$(0, 1)$
Tanh	$f(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$	$\mathbb{R}$	$(-1, 1)$
Sofmax	$f_i(\vec{x}) = \frac{e^{x_i}}{\sum_{k=1}^{K} e^{x_k}}$	$\mathbb{R}^K$	$(0, 1)$

Activation Function



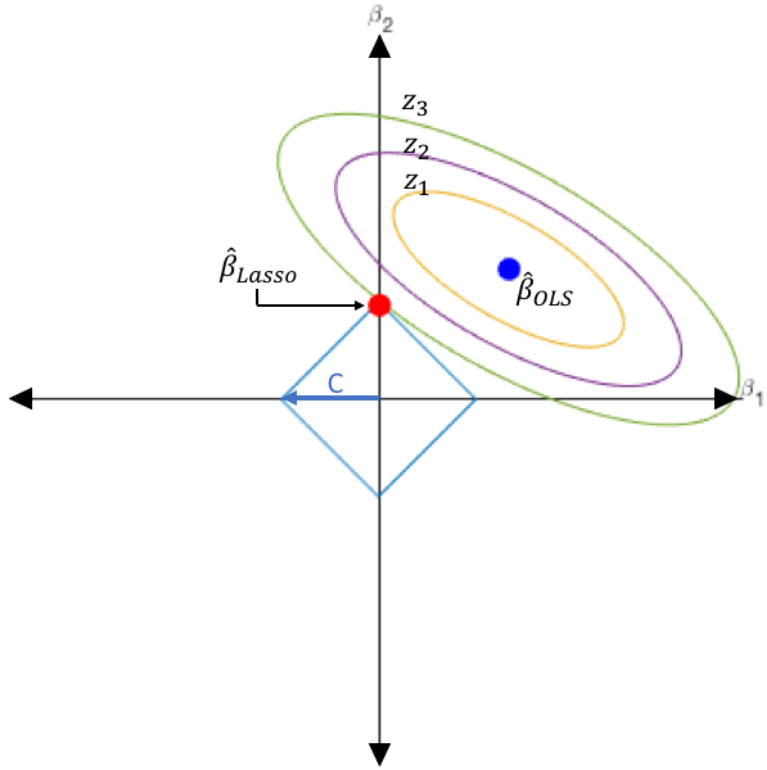


Loss Function with L1 penalty

$$\text{Log Loss: } L(p, y) = -y log(p) - (1 - y) log(1 - p)$$

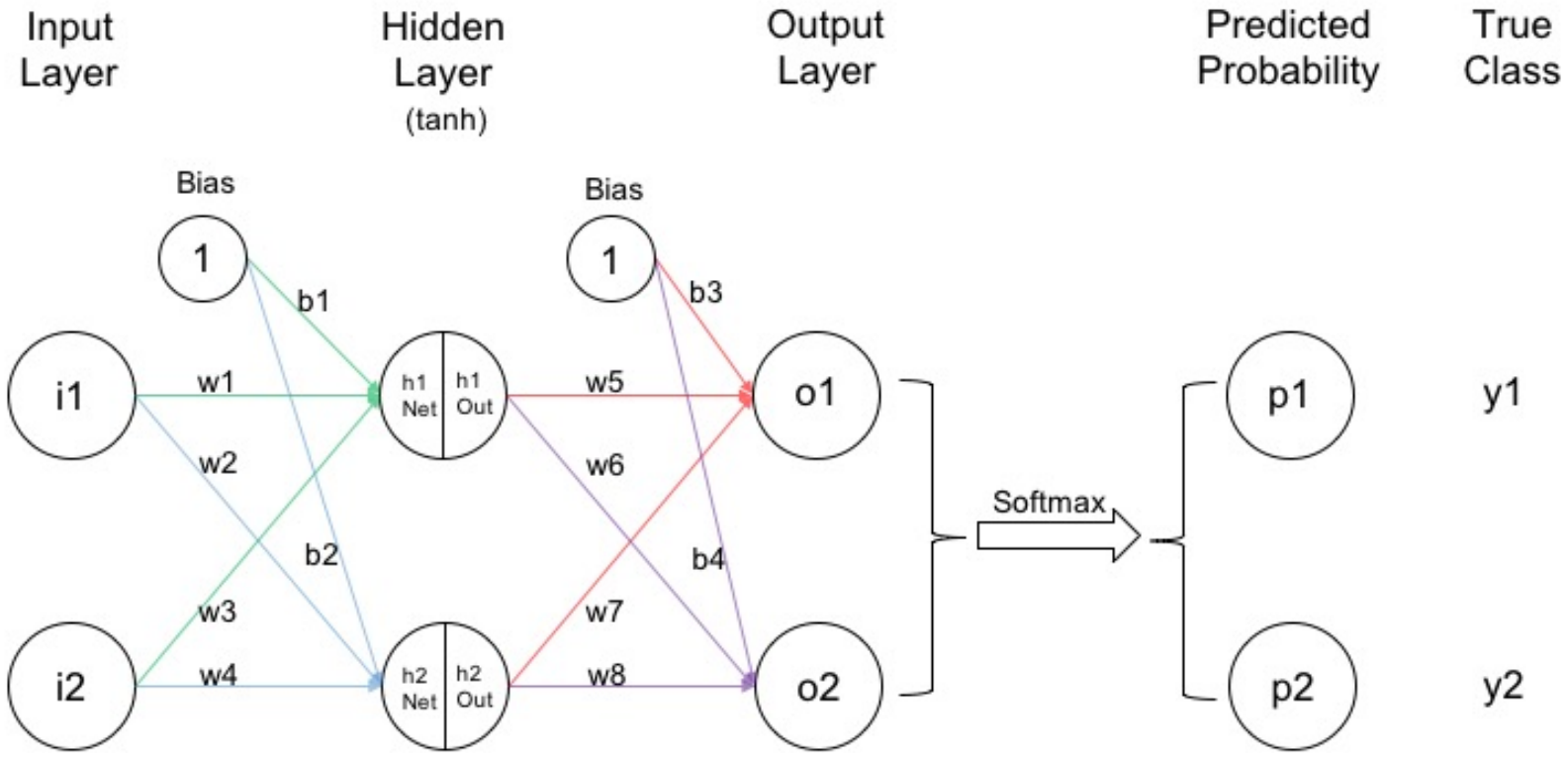
$$\text{Cross Entropy: } L(P, Y) = -\sum_k y_k log(p_k)$$

$$L = -\frac{1}{n} \sum_{allx} \sum_k y_k log(p_k) + \lambda \sum_l \sum_i \sum_j |w_{ij}^{(l)}|$$



Backpropagation

$$w_{ij}(t+1) = w_{ij}(t) - \eta \frac{\partial L}{\partial w_{ij}}$$



$$\frac{\partial L}{\partial w_5} = (p_1 - y_1) h_{1Out} = \delta_5 \times h_{1Out}$$

$$\frac{\partial L}{\partial w_1} = [(p_1 - y_1) w_5 + (p_2 - y_2) w_6][1 - tanh^2(h_{1Net})] i_1 = \delta_1 \times i_1$$

where $\delta_5 = (p_1 - y_1)$ and $\delta_1 = [(p_1 - y_1) w_5 + (p_2 - y_2) w_6][1 - tanh^2(h_{1Net})]$.

Stochastic Gradient Descent and Mini-batch Gradient Descent

Stochastic Gradient Descent(SGD): Randomly draw one piece of training example out of the training set at each iteration and perform backpropagation solely based on it.

Mini-batch Gradient Descent: Take a small batch of training example to perform the backpropagation update.

Model Tuning and Parameter selection

- Use *neuralnetwork* function in *ANN2* package. ANN2 is capable of performing SGD.
- 20% of the original train set is randomly selected out to be validation set while the remaining will still be the training set.
- Structure: 59-5-2
- Batchsize=32
- Learning rate=$1e-4$
- L1(λ)=0.03