

Notes on a 2-Layer Feed Forward Neural Network for Regression Tasks

Jonathan Navarrete

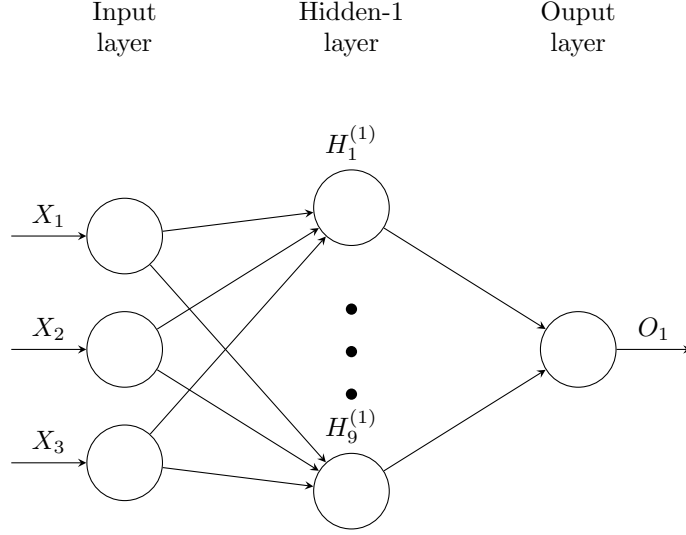
2019
January

1 Introduction

The following notes relate to a simple feed forward neural network trained for regression tasks. The neural network has an input layer (the inputs) and two hidden layers. The first hidden layer receives the input data and transforms it into a nonlinear space. The data is feed forward into the second hidden layer for the output. The neural network is trained using vanilla stochastic gradient descent (SGD).

The neural network is implemented in a Python script (`nnet.py`).

2 Architecture



2.1 Data

The neural network takes 3 inputs, where the design matrix \mathbf{X} is $N \times 3$; however, this can be generalized to any number of M inputs for a $N \times M$ design matrix. The neural network is then fed one observation at a time, \mathbf{x}_i^\top for $i = 1, 2, \dots, N$

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \dots & \dots & \dots \\ x_{N1} & x_{N2} & x_{N3} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \dots \\ \mathbf{x}_N^\top \end{bmatrix}$$

There is one output target

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_N \end{bmatrix}$$

2.2 First Hidden Layer

In the first hidden layer there are 9 hidden units $H_1^{(1)}, H_2^{(1)}, \dots, H_9^{(1)}$. Each hidden unit sees each of the three 3 inputs, aggregates the inputs with

weights and passes that hidden output to an activation function. The activation function used in the first hidden layer is logistic,

$$H_k^{(1)} = f(z) = \frac{1}{1 + e^{-z}}$$

where $z = w_1^{(1)} x_{i1} + w_2^{(1)} x_{i2} + w_3^{(1)} x_{i3}$.

For each hidden unit, we have three weights $\mathbf{w}^{(1)}$

2.3 Second Hidden Layer

In the second hidden layer there is 1 hidden units (the output unit). Because there is only one output, y_i , there is only the need for one hidden unit.

The second hidden layer expects 9 inputs from the first hidden layer. The activation function is simply the identity function $g(z) = z$, where $z = w_1^{(2)} x_{i1} + w_2^{(2)} x_{i2} + w_9^{(2)} x_{i3}$.

3 Feed Forward

For each observation i , we take input \mathbf{x}_i^\top and pass it to the first layer.

$$\mathbf{x}_i^\top \mathbf{w}^{(1)}$$