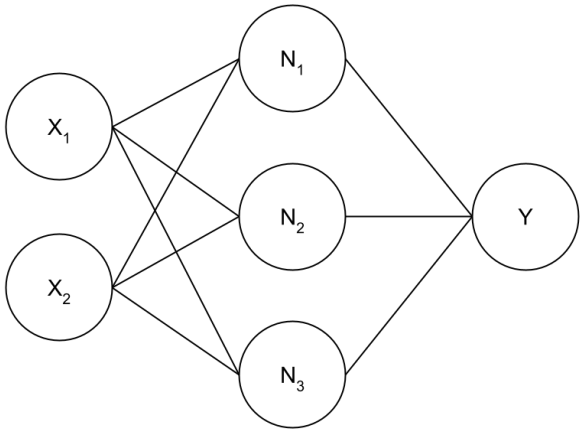
ASTR 5900 Neural Network Homework

Michael Bartlett

March 2024

1 Question 1

Below is a diagram of a simple neural network with 2 input nodes, 1 hidden layer with 3 nodes, and an output layer with 1 node. For this problem, $w_{a,b}$ represents the weight between nodes a and b (e.g. $w_{X1,N1}$ gives the weight between nodes X_1 and N_1) and b_a represents the bias for node a.



A Implement this neural network using hyperbolic tangent as the activation function for the hidden layer and output layer, and using the following weights:

$w_{X1,N1}$	-0.91
$w_{X1,N2}$	0.09
$w_{X1,N3}$	0.72
$w_{X2,N1}$	0.72
$w_{X2,N2}$	0.34
$w_{X2,N3}$	0.62
$w_{N1,Y}$	-0.56
$w_{N2,Y}$	0.94
$w_{N3,Y}$	-0.47
b_{N1}	0.29
b_{N2}	0.05
b_{N3}	-0.99
b_Y	-0.25

B What does this network predict for the points $(0, 0)$, $(7.5, 2.5)$, and $(-5, -2)$?

C Make a heat map with a color bar that shows the networks predicted values for all points $-10 < X_1 < 10$ and $-10 < X_2 < 10$.

D Try changing some of the weights and biases and see how that impacts your heat map. Plot an example that you find interesting.

2 Question 2

For this problem, we are going to calculate the gradient of the loss function of the above neural network with respect to $w_{N1,Y}$. Assume that the loss function is given by $L(\hat{y}, y) = (\hat{y} - y)^2$ where $\hat{y}$ is the predicted value by the neural network and y is the true label. Leave the loss function as a generic σ.

A Using the chain rule, we know that $\frac{\partial L}{\partial w_{N1,y}} = \frac{\partial L}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial w_{N1,Y}}$. What is $\frac{\partial L}{\partial \hat{y}}$?

B Recall that $\hat{y} = \sigma(w_{N1,Y}N_1 + w_{N2,Y}N_2 + w_{N3,Y}N_3 + b_Y)$. Using $\frac{\partial \sigma(w_{N1,Y}N_1 + w_{N2,Y}N_2 + w_{N3,Y}N_3 + b_Y)}{\partial(w_{N1,Y}N_1 + w_{N2,Y}N_2 + w_{N3,Y}N_3 + b_Y)} = \sigma'$ what is $\frac{\partial \hat{y}}{\partial w_{N1,Y}}$?

C Combining the answers from the previous two parts, what is $\frac{\partial L}{\partial w_{N1,y}}$?

D Assume that for a given input using the weights from problem 1, that we get a prediction from the neural that $\hat{y} = 0.5$ while the true value is $y = 1$. Also assume that for this input that $N_1 = 0.8$ and $\sigma' = 3$. Using a learning rate of $l = 0.1$ what should the updated value of $w_{N1,Y}$ be?

3 Question 3

The magic04 data set is simulated data that models what a gamma telescope detects when gamma rays hit the atmosphere (you can read more about it here.

magic04.data is a located in the slack under assignments. Using this data set, utilize TensorFlow to create a neural net with an input layer, 1 hidden layer with 100 nodes, and 1 output layer with 2 nodes. Select values for the other parameters that you feel are appropriate.

A After training your neural network, what accuracy do you achieve on the test set?

B To see how different parameters can impact the accuracy of the model, choose one of the following: number of hidden layers, number of nodes per hidden layer, number of training epochs, activation function used, optimizer used. Now retrain the same neural network from part A, but vary the selected parameter and plot how it affects accuracy on the test set. Repeat this for another one of the chosen parameters.

C Pre-processing the data can also have an impact on the accuracy of the model. Try normalizing the data before training. You can do this manually in several different way (re-scaling the sample, shifting elemnets so they have a mean of zero, making each element a unit vector etc. I recommend using scikit-learn normalize). Now retrain your model from part A. Did this impact your accuracy?