CS 6375 Machine Learning - Neural Networks

Names of students in your group

Karan Kanani (kyk170030)

Tejas Ravi Rao (txr171830)

Number of free late days used: 1

Please list clearly all the sources/references that you have used in this assignment.

Assumptions

* Bias neurons are ignored
* Last column in the dataset is the class label
* Training and test datasets have the same format

The dataset is read from following URLs

* Car Evaluation Dataset: <https://archive.ics.uci.edu/ml/datasets/Car+Evaluation>

The datasets are preprocessed by changing the categorical variables to numerical values, standardization and scaling of attributes.

Our neural model is built with one input layer, 2 hidden layers [4,2] and one output layer.

Number of nodes in each hidden layer can be changed.

We have evaluated our neural net by changing the different parameters such as

* Learning rate
* Number of nodes in hidden layer
* Number of iterations

**Case 1:**

Changing/variable parameter in NN-> Learning rate

Here, we are setting number of iterations = 1000

Learning rate values -> 0.01, 0.05, 0.1,0.2

Number of nodes in hidden layer 1 = 4

Number of nodes in hidden layer 1 = 2

We have documented the result by changing the learning rate in the following table (refer next page)

Following table has entries for training and test error for Car dataset for activation functions sigmoid, tanh, relu for different values of learning rate.

We have run our neural network code for various learning rates -> 0.01, 0.05, 0.1, 0,2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Parameter ƞ** | **Dataset** | **Activation function** | **Training Error** | **Test Error** |
| 0.01 | Car | Sigmoid | 0.272090 | 0.269976 |
|  |  | tanh | 1.522799 | 0.455382 |
|  |  | Relu | 0.505780 | 0.487171 |
| 0.05 | Car | Sigmoid | 0.269640 | 0.266690 |
|  |  | tanh | 0.467631 | 0.465807 |
|  |  | Relu | 0.505780 | 0.487171 |
| 0.1 | Car | Sigmoid | 0.352680 | 0.351893 |
|  |  | tanh | 1.509130 | 0.452331 |
|  |  | Relu | 0.505780 | 0.487171 |
| 0.2 | Car | Sigmoid | 0.355858 | 0.350545 |
|  |  | tanh | 1.530533 | 0.538030 |
|  |  | Relu | 0.505780 | 0.487171 |

**Analysis**

We have changed our learning rates from 0.01 to 0.2. If the learning rate is low, then there is significant difference between Training error and test error. As we are increasing the learning rate, we notice that for sigmoid and relu activation functions, difference between training and test error is very less.

On comparison of 3 activation functions, we find that tanh performs significantly better when compared to the other two.

**Case 2:**

Changing/variable parameter in NN-> **Number of nodes in hidden layer**

Here, we are setting number of iterations = 1000

Learning rate -> 0.05

Number of nodes in hidden layer 1 = 5

Number of nodes in hidden layer 1 = 3 **OR**

Number of nodes in hidden layer 1 = 2

Number of nodes in hidden layer 1 = 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Parameter** | **Dataset** | **Activation function** | **Training Error** | **Test Error** |
|  | Car | Sigmoid | 0.364704 | 0.327431 |
| h1 = 5 |  | tanh | 0.473137 | 0.430940 |
| h2 = 3 |  | Relu | 0.522452 | 0.466834 |
| h1 = 2 | Car | Sigmoid | 0.374288 | 0.341835 |
| h2 = 2 |  | tanh | 1.404088 | 0.462794 |
|  |  | Relu | 0.522452 | 0.466834 |

We experimented with number of neurons in hidden layers such as h1=2, h2= 2 OR

h1 = 4, h2 = 2 OR h1=5, h2 = 3. Initially we started increasing number of nodes in the hidden layer, and even after decreasing the hidden layer nodes, there was no case of overfitting observed as the test error was always lesser as compared to the training error.

**Case 3:**

Changing/variable Parameter in NN-> Number of iterations

Here, we are setting number of iterations = 500 **OR** 2000

Learning rate -> 0.05

Number of nodes in hidden layer 1 = 4

Number of nodes in hidden layer 1 = 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Parameter** | **Dataset** | **Activation function** | **Training Error** | **Test Error** |
| Number of | Car | Sigmoid | 0.270745 | 0.268289 |
| iterations |  | tanh | 0.461642 | 1.448623 |
| 500 |  | Relu | 0.505780 | 0.487171 |
|  | Car | Sigmoid | 0.266924 | 0.265723 |
| 2000 |  | tanh | 1.548990 | 0.738289 |
|  |  | Relu | 0.543610 | 0.447120 |

**Analysis**

We saw that at some threshold with lower number of iterations, there was a case wherein the test error was greater than that of the training error showing overfitting. At higher number of iterations the model was significantly better specially with the tanh activation function.