Impact of Activation Functions and Learning Rates on Non-Linear Pattern Learning in Neural Networks

2024-12-06

```
##   pattern activation learning.rate test.error
## 1  spiral       ReLU         0.001      0.496
## 2  spiral       ReLU         0.003      0.496
## 3  spiral       ReLU         0.010      0.465
## 4  spiral       Tanh         0.001      0.493
## 5  spiral       Tanh         0.003      0.487
## 6  spiral       Tanh         0.010      0.421
```

1 Abstract

This study investigates the effects of different activation functions (ReLU, Tanh, Sigmoid, and Linear) and learning rates (low - 0.001, medium - 0.003, high - 0.01) on the ability of a neural network to learn non-linear data patterns. The experiment involves generation of synthetic datasets featuring `spiral`, `concentric`, and `Gaussian` data patterns for binary classification tasks using neural networks.

The test error, representing misclassification rates, was assessed across 36 treatment combinations. Results were analyzed using ANOVA to identify significant factors influencing performance.

2 Introduction

The use of neural networks in classifying complex, non-linear data patterns is a central challenge in machine learning. This experiment explores how different activation functions and learning rates affect a neural network's ability to accurately classify such patterns. Specifically, we aim to determine the optimal hyper-parameters for training a neural network with fixed architecture and its implications for learning efficiency.

The report is structured as follows: **Section 2** describes the experimental design, **Section 3** outlines the statistical analysis, and **Section 4** presents the conclusions.

3 Details of the Experimental Design

3.1 Factors and Levels

Factor	Description	Levels
Activation Function	Non-linearity handling functions at each node	ReLU, Tanh, Sigmoid Linear
Learning Rate	Rate at which model weights are updated	Low (0.001), Medium (0.005), High (0.01)

Factor	Description	Levels
Data Patterns	Types of patterns used for classification	Spiral, Gaussian Mixture, Concentric Circles

3.2 Experimental Plan

The experiment involves a fully connected neural network with 4-8-4 neurons in a hidden layer configuration. The network uses a sigmoid activation function in the output layer for binary classification. The data features used for training the neural network include `X_1`, `X_2`, `X_1^2`, `X_2^2`, and `X_1 * X_2` (interaction term). A 70-30 data split is used, where 30% of the data is used for testing. The neural network is trained for 100 epochs, and the test error (misclassification error) is calculated after each epoch. The goal is to analyze the impact of different treatment combinations (activation function, learning rate, and data pattern) on the test error.

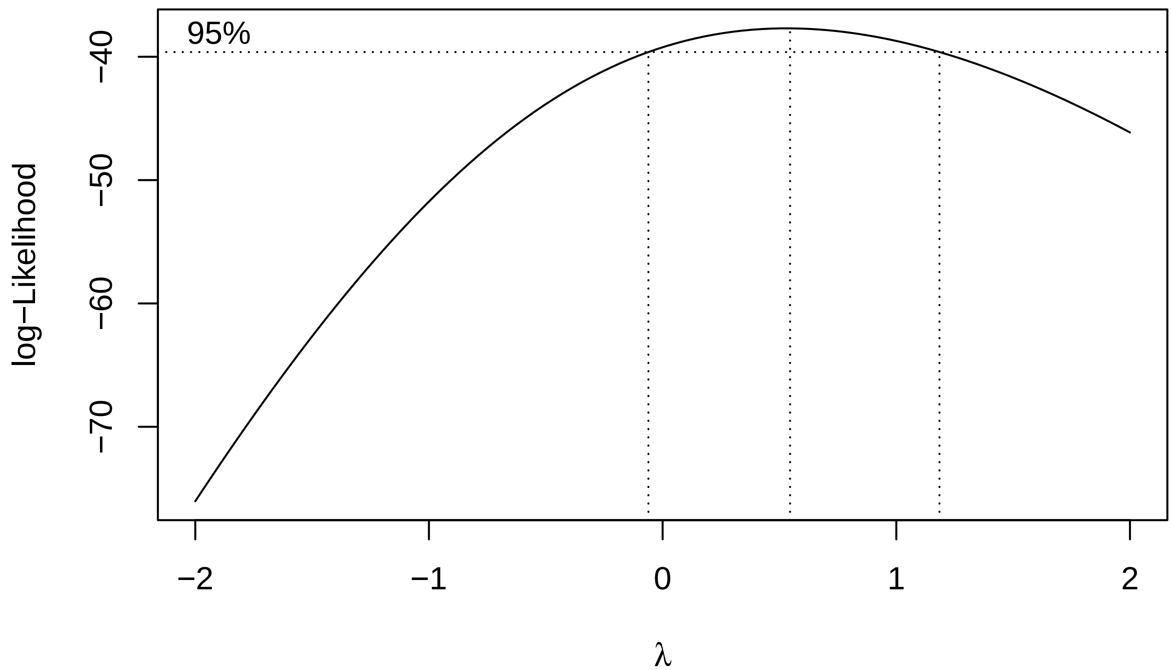
3.3 Randomization and Blocking

Randomization is applied to the treatment combinations to avoid bias. Data patterns serve as the blocking factor, ensuring that observed variations in the test error are not confounded by the complexity of the patterns themselves. The randomization of treatment combinations across runs ensures that results are not influenced by the order of the treatments.

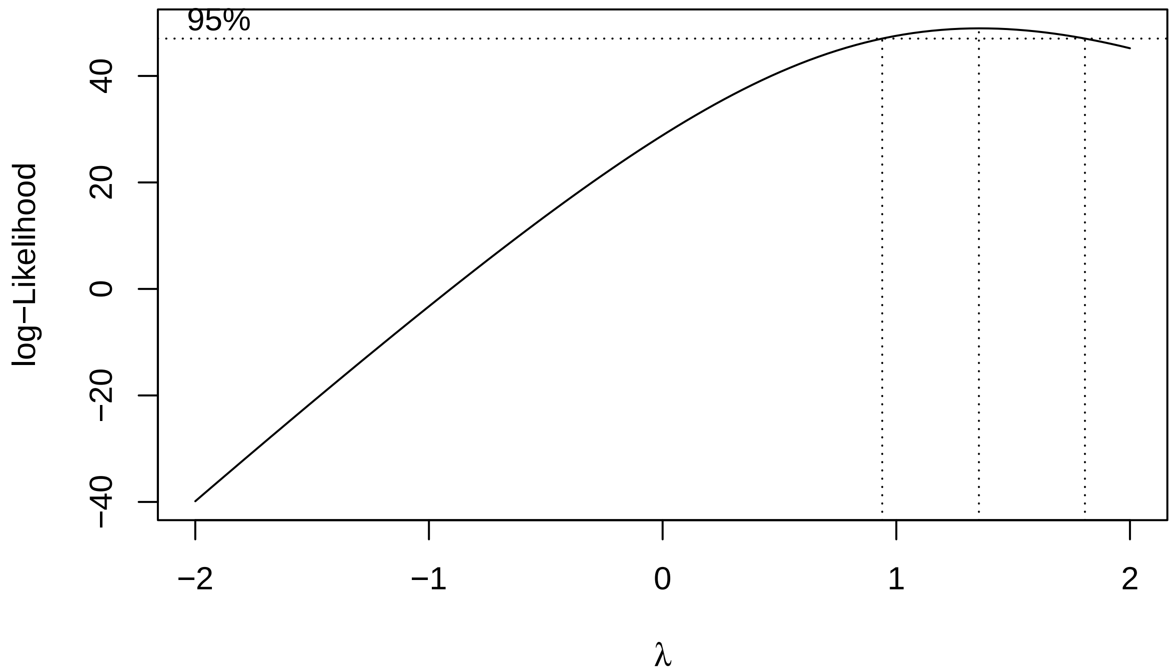
4 Statistical analysis

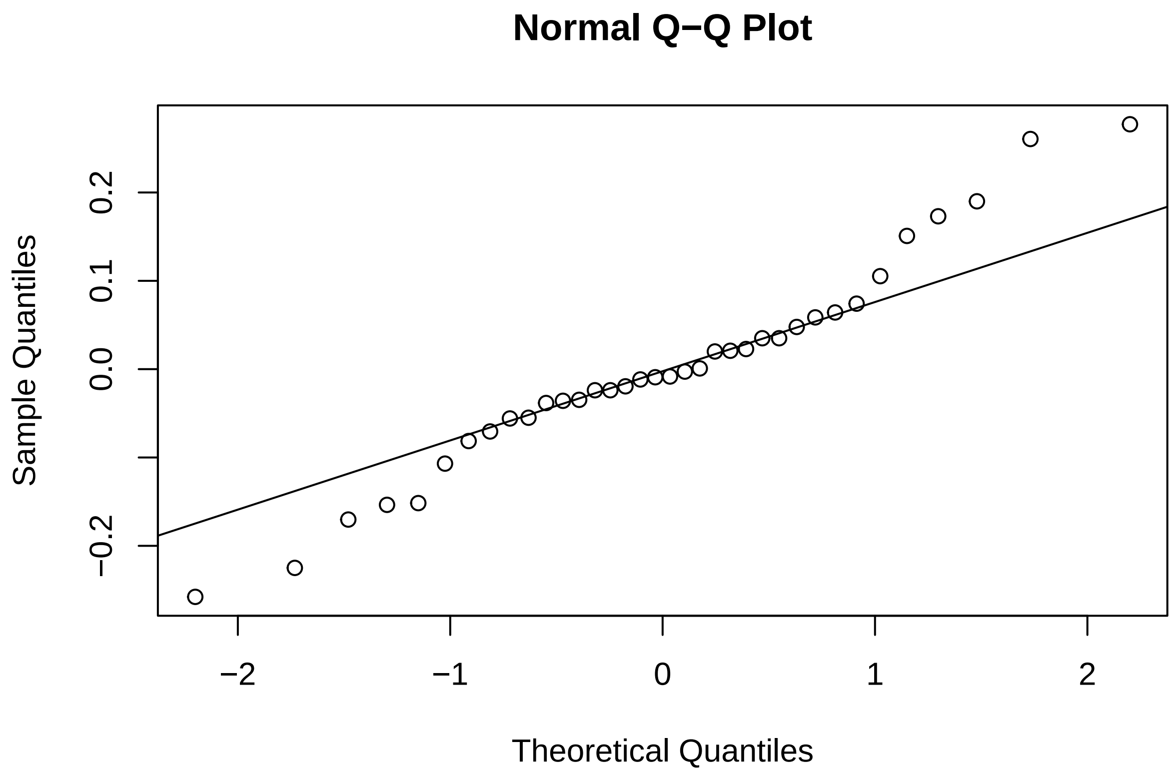
```
## 'data.frame':    36 obs. of  4 variables:
##  $ pattern      : Factor w/ 3 levels "concentric","gaussian",..: 3 3 3 3 3 3 3 3 3 3 ...
##  $ activation   : Factor w/ 4 levels "Linear","ReLU",..: 2 2 2 4 4 4 3 3 3 1 ...
##  $ learning.rate: Factor w/ 3 levels "0.001","0.003",..: 1 2 3 1 2 3 1 2 3 1 ...
##  $ test.error   : num  0.496 0.496 0.465 0.493 0.487 0.421 0.5 0.5 0.5 0.469 ...
```

```
##                          Df Sum Sq Mean Sq F value   Pr(>F)    
## pattern                   2  50.06  25.029 611.225 8.44e-13 ***
## activation                3  40.36  13.452 328.509 8.84e-12 ***
## learning.rate             2   0.32   0.162   3.962   0.0477 *  
## pattern:activation        6  17.57   2.928  71.499 1.08e-08 ***
## activation:learning.rate  6   0.40   0.066   1.613   0.2262    
## pattern:learning.rate     4   0.10   0.025   0.599   0.6708    
## Residuals                12   0.49   0.041                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

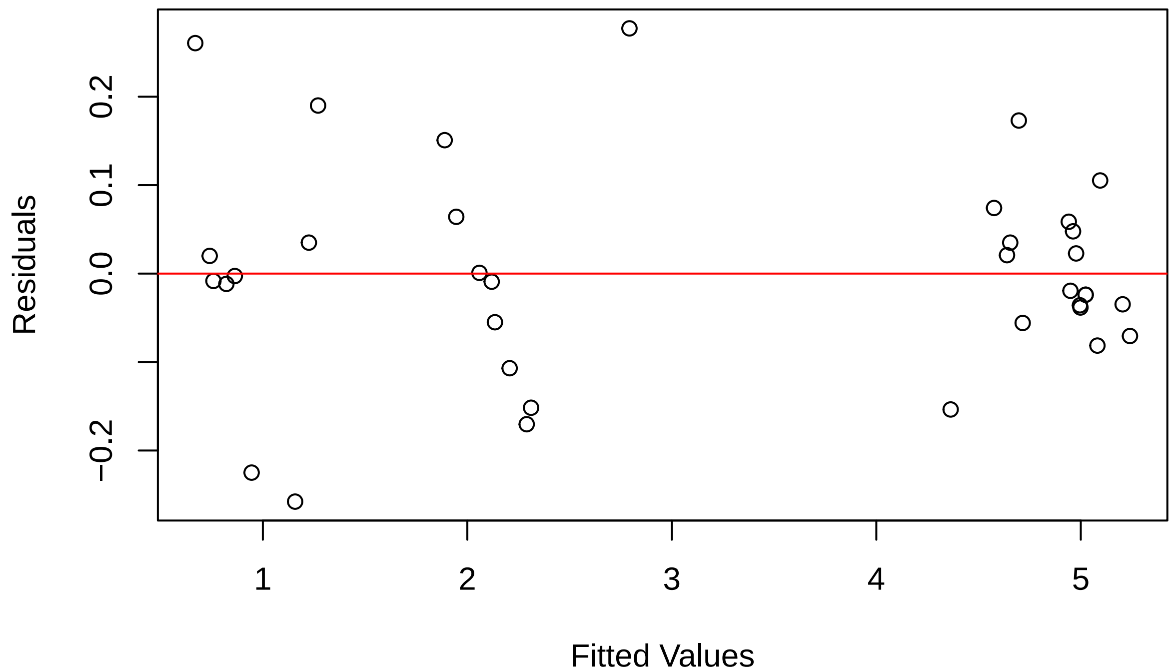


95%
log−Likelihood
−40
−50
−60
−70
−2
−1
0
1
2
λ

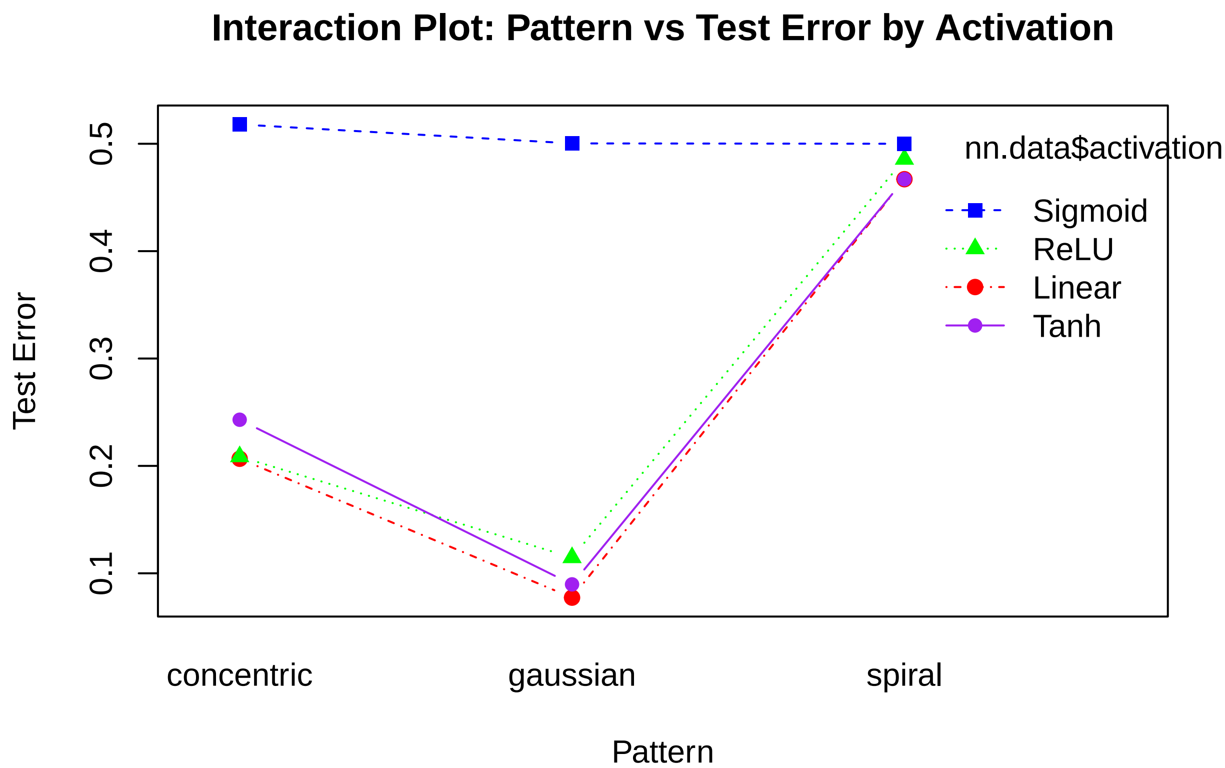

95%
log−Likelihood
−40
−20
0
20
40
−2
−1
0
1
2
λ



Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−0.2
0.0
0.1
0.2
−2
−1
0
1
2




```
##                           Df Sum Sq Mean Sq F value   Pr(>F)    
## pattern                    2  50.06  25.029 611.225 8.44e-13 ***
## activation                 3  40.36  13.452 328.509 8.84e-12 ***
## learning.rate              2   0.32   0.162   3.962   0.0477 *  
## pattern:activation         6  17.57   2.928  71.499 1.08e-08 ***
## activation:learning.rate   6   0.40   0.066   1.613   0.2262    
## pattern:learning.rate      4   0.10   0.025   0.599   0.6708    
## Residuals                 12   0.49   0.041                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

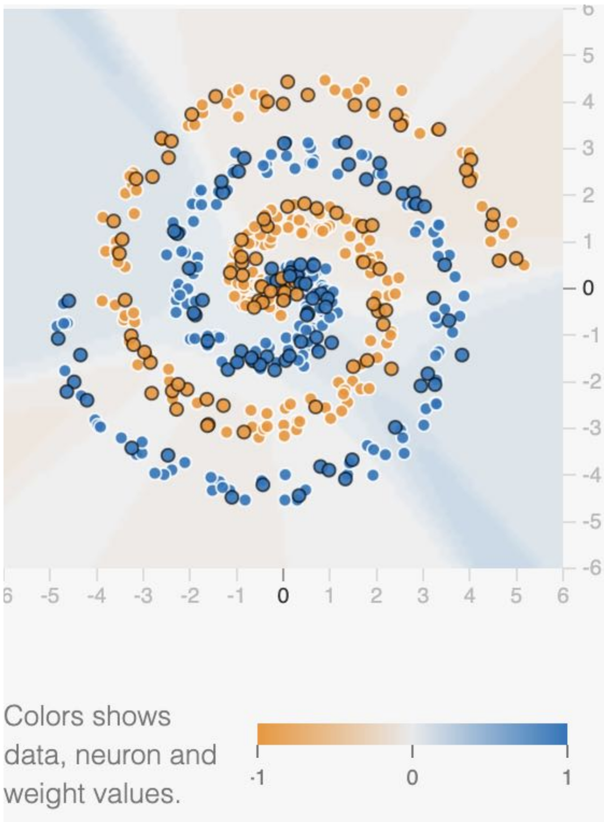


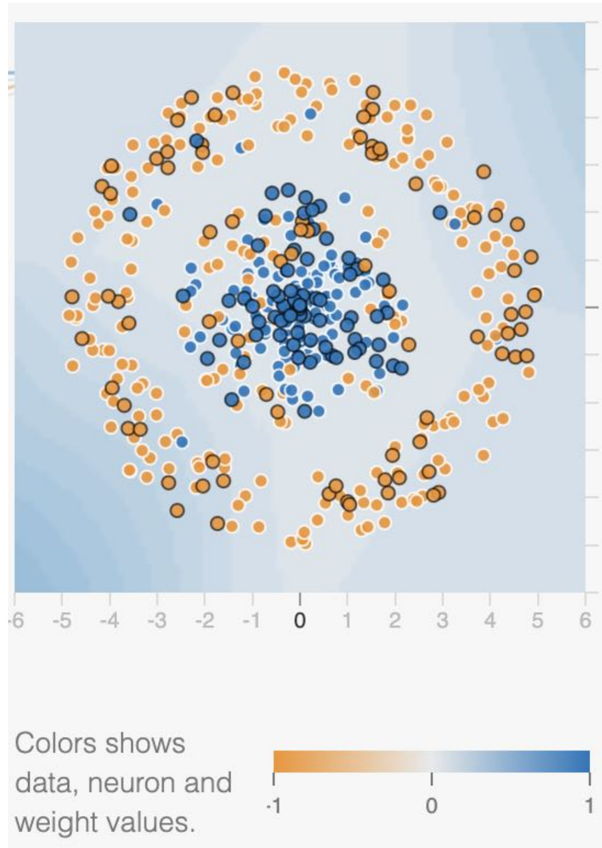
5 Results and Conclusions

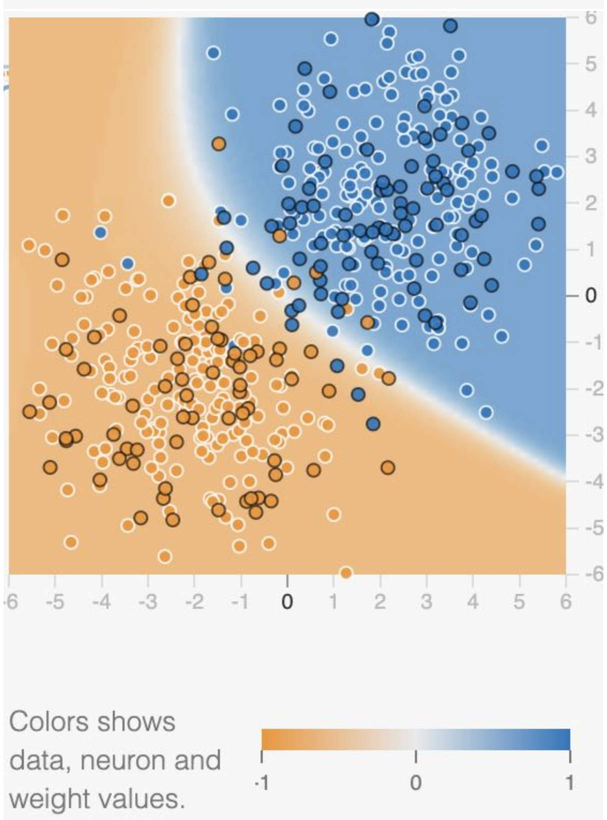
5.1 Descriptive Statistics

This experiment demonstrates the influence of activation functions and learning rates on neural network performance in classifying non-linear data patterns. We found that. . . Future work could explore the impact of additional network architectures, optimizers, and regularization techniques to improve model performance further.

6 Tables and Figures







7 Units

All values related to the model parameters (e.g., learning rates, test error) will be presented with their corresponding units. The learning rate is dimensionless, while the test error (misclassification rate) is given as a percentage.

8 Data Appendix

The dataset and the R code used for the analysis will be included in the appendix. The dataset contains the features `X_1`, `X_2`, `X_1^2`, `X_2^2`, and `X_1 * X_2`, with corresponding labels for each data point. The R code for generating the synthetic datasets, training the neural network, and conducting the statistical analysis will be provided.