

NEURAL NETWORKS

ANN Project Report

Adam Swart

September 17, 2016

1 Introduction

Neural Networks are constructed and trained for each of the following data sets, each of which was obtained from <https://archive.ics.uci.edu/ml/datasets.html>. The same approach is used for each neural network: import and normalise the data; partition the data into a training set, a test set, and a validation set; and find the optimum training algorithm through a process of trial and error. Each neural network's degree of fit is assessed, and performance is analysed using the MATLAB postregression function.

2 Wine

The wine dataset is the result of a chemical analysis of wines taken from three different cultivars from the same region in Italy. There are 13 attributes for each instance, and each attribute is a real value. The attributes are: alcohol, malic acid, ash, alkalinity of ash, magnesium, total phenols, flavanoids, nonflavanoid phenols, proanthocyanins, colour intensity, hue, OD280/OD315 of diluted wines, and proline.

Our first step is to import the data from the text file. No normalisation or encoding is necessary since all the input patterns are real numbers and the targets are integers. We construct a feed-forward network with 2 layers, each layer with 5 neurons. We find that with the Levenburg-Marquardt backpropagation with 1000 epochs and the maximum number of failed validations set to 20, the minimum gradient is found quickly (after 49 iterations). We partition the data set into a training set, test set, and validation set, and begin training.

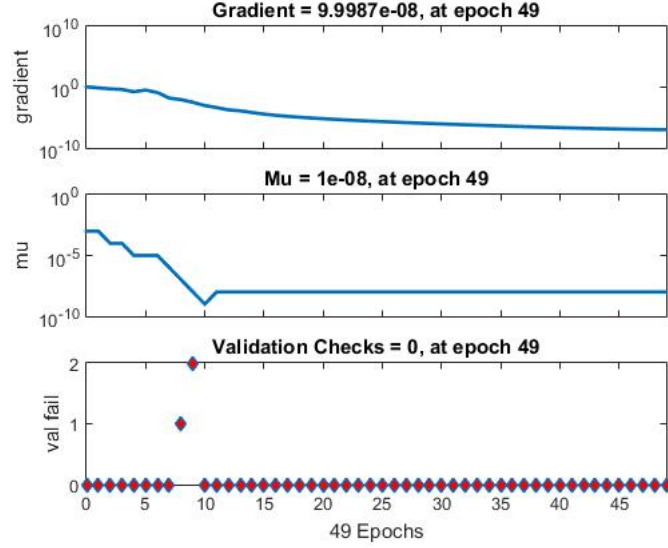


Figure 1: Wine train state.

Once training is finished, we save all variables. We are now ready to test the network. On the training set, the correlation coefficient is 0.999995 and the r^2 value is 0.999989. On the test set, the correlation coefficient is 0.960454 and the r^2 value is 0.913144. On the entire data set, the correlation coefficient is 0.98618 and the r^2 value is 0.971409. This indicates a high degree of accuracy. The postregression values were 0.999854, 1.00324, and 1.00132 for the training set, test set, and entire set respectively.

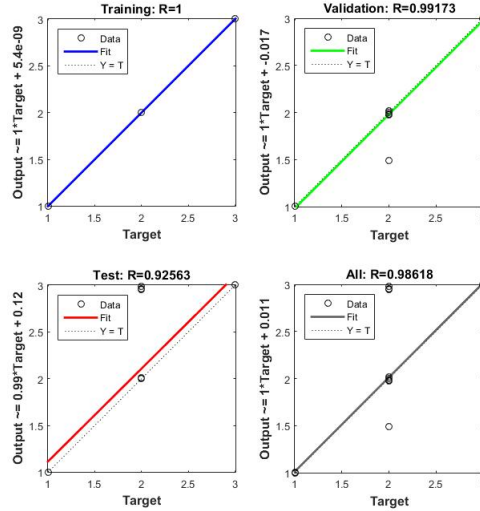


Figure 2: Wine neural network training regression.

The degree of fit of the network is highly accurate. The network trains well on the data and reaches a minimum error quickly.

3 Mushrooms

The mushroom dataset includes descriptions of hypothetical samples of 23 species of gilled mushroom in the Agaricus and Lepiota family. Each sample is classified as either edible or poisonous, and has 22 attributes. Each attribute can be one of several letters indicating the value of the attribute.

We import and encode the data. Since each attribute is a lower case letter, we convert the attribute to its ASCII value and subtract 97. Now if any attribute is less than 0 we know that this instance has a missing entry.

We construct a 2-layer feed-forward network. The first layer has 20 neurons and the second layer has 10 neurons. Since this is a relatively large dataset we choose the scaled conjugate gradient backpropagation algorithm with 1000 epochs and the maximum number of failed validations set to 20. We partition the data into a training set, a test set, and a validation set. We train the network.

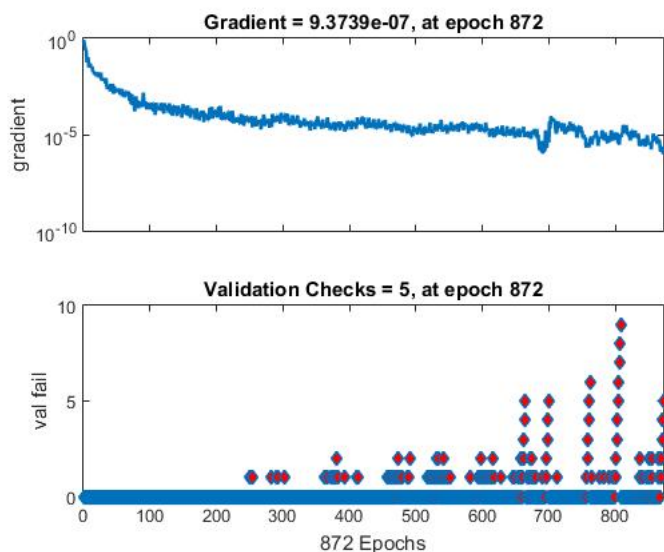


Figure 3: Mushrooms train state.

We find that the network converges to the minimum gradient close to the maximum number of epochs. Once training is complete we save all variables and begin testing. We find that on the training set, the test set, and the entire set, the correlation coefficient is 1 and the r^2 value is 1. The postregression value is 1, 0.999996, and 1 for the training set, test set, and whole set respectively. This indicates an optimal degree of fit.

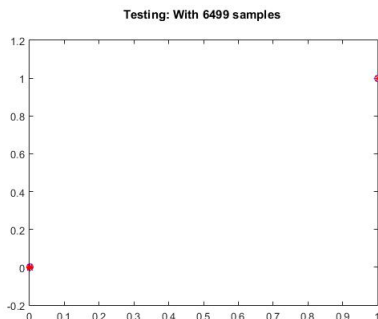


Figure 4: Mushrooms testing. Activations and targets match.

4 Iris

The iris dataset contains 3 classes, each of 50 instances. Each instance has 4 attributes: sepal length in centimetres, sepal width in centimetres, petal length in centimetres, and petal width in centimetres. The classes are Iris-Setosa, Iris-Versicolour, and Iris-Virginica.

We import the data and encode the targets: we set Iris-Setosa, Iris-Versicolour, and Iris-Virginica to 0, 1, and 2 respectively. We construct a 2-layer feed forward neural network, each layer with 10 neurons. We use the scaled conjugate gradient backpropagation algorithm with 1000 epochs and the maximum number of failed validations set to 20. We partition the data into a training set, a test set, and a validation set. We train the network.

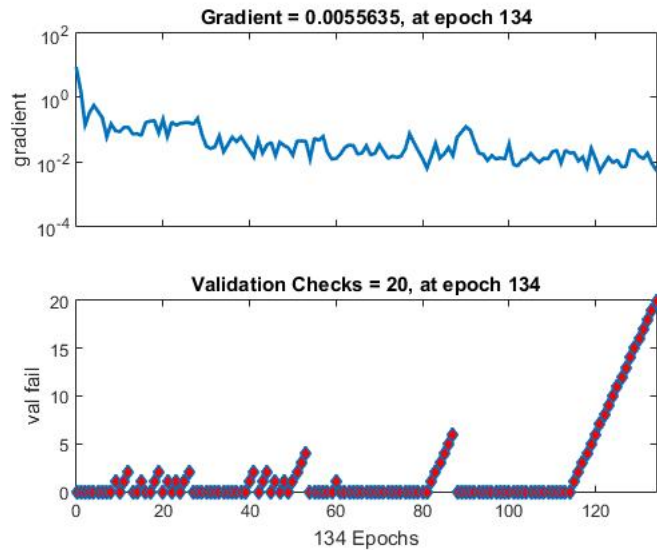


Figure 5: Iris train state.

The maximum number of failed validations is exceeded and training ceases. We save all variables and begin testing. On the training set, the correlation coefficient is 0.983743 and the r^2 value is 0.966971. On the test set, the correlation coefficient is 0.995552 and the r^2 value is 0.990767. On the entire data set, the correlation coefficient is 0.984217 and the r^2 value is 0.968362. The postregression values were 0.950495, 0.986778, and 0.958892 for the training set, test set, and entire set respectively.

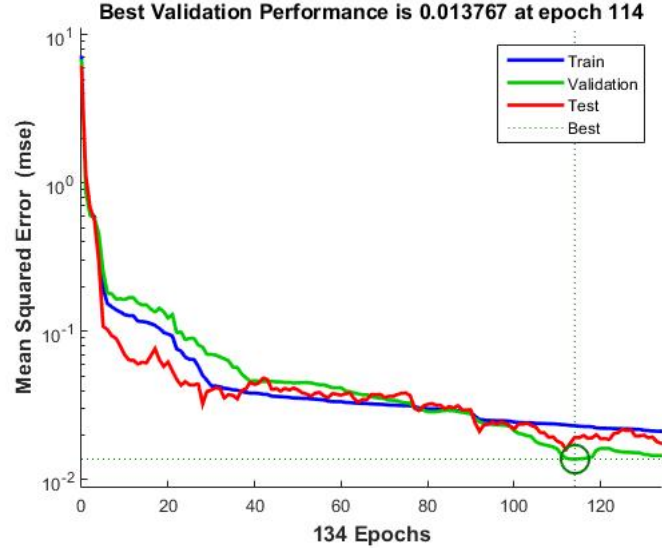


Figure 6: Iris performance.

While the postregression values, r^2 values, and correlation coefficients are not entirely optimal, the degree of fit is highly accurate.

5 Abalone

The abalone dataset predicts the age of abalone from 8 attributes. The dataset has 4177 instances. We import and normalise the data. We assign a value of 0 for the Male attribute, 1 for the Female attribute, and 2 for the Infant attribute. We construct a 2-layer feed forward neural network, each layer with 15 neurons. We use the Polak-Ribiere conjugate gradient backpropagation algorithm with 1000 epochs and the maximum number of failed validations set to 20. We partition the data into a training set, a test set, and a validation set, and train the network.

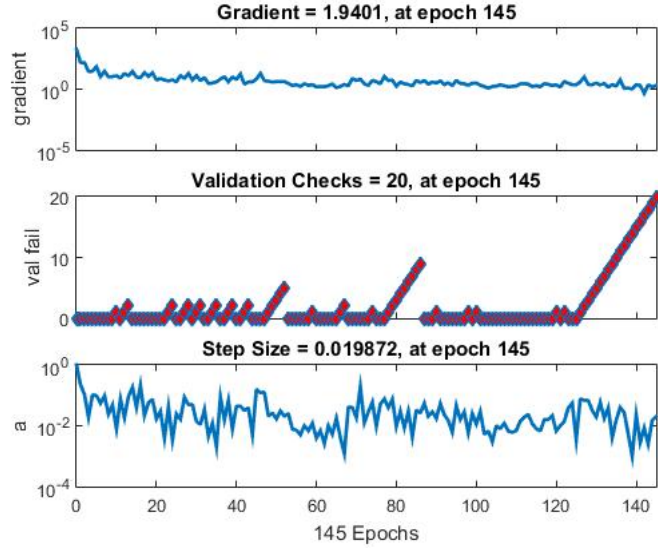


Figure 7: Abalone training state.

The maximum number of failed validations is exceeded and training ceases. We save all variables and begin testing. On the training set, the correlation coefficient is 0.758545 and the r^2 value is 0.574383. On the test set, the correlation coefficient is 0.758359 and the r^2 value is 0.574525. On the entire data set, the correlation coefficient is 0.759644 and the r^2 value is 0.57656. The postregression values were 0.597163, 0.586786, and 0.594014 for the training set, test set, and entire set respectively. This indicates a poor degree of fit.

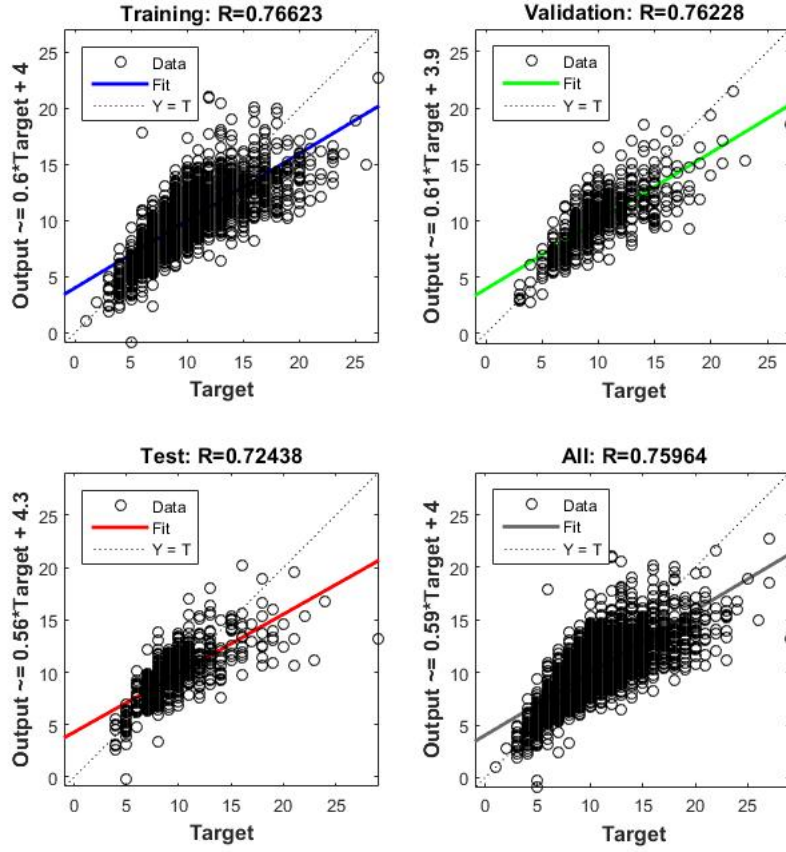


Figure 8: Abalone training regression.

The distribution of data makes it difficult to find an accurate degree of fit.

6 Concrete

The concrete benchmark dataset predicts the compressive strength of concrete given 8 quantitative inputs. We import the data. No normalisation or encoding is needed. We construct a 3-layer feed forward neural network, each layer with 20 neurons. We use the scaled conjugate gradient backpropagation algorithm with 1000 epochs and the maximum number of failed validations set to 20. We partition the data into a training set, a test set, and a validation set, and train the network.

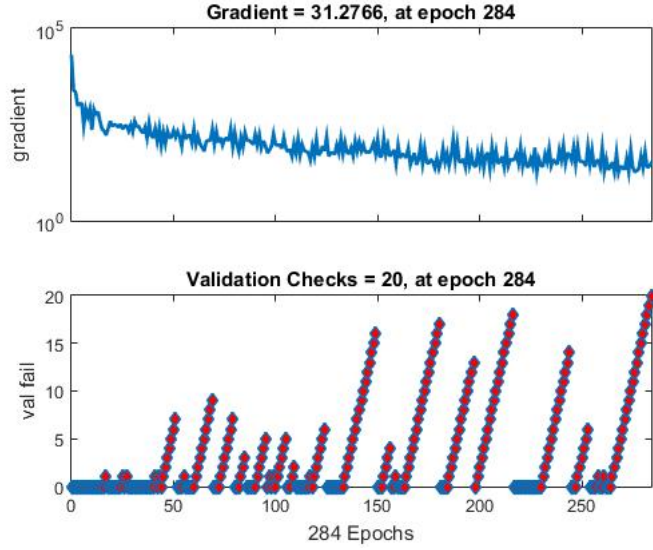


Figure 9: Concrete training state.

The maximum number of failed validations is exceeded and training ceases. We save all variables and begin testing. On the training set, the correlation coefficient is 0.960284 and the r^2 value is 0.921712. On the test set, the correlation coefficient is 0.950959 and the r^2 value is 0.904218. On the entire data set, the correlation coefficient is 0.958799 and the r^2 value is 0.918972. The postregression values were 0.942134, 0.911618, and 0.935799 for the training set, test set, and entire set respectively. This indicates an accurate degree of fit.

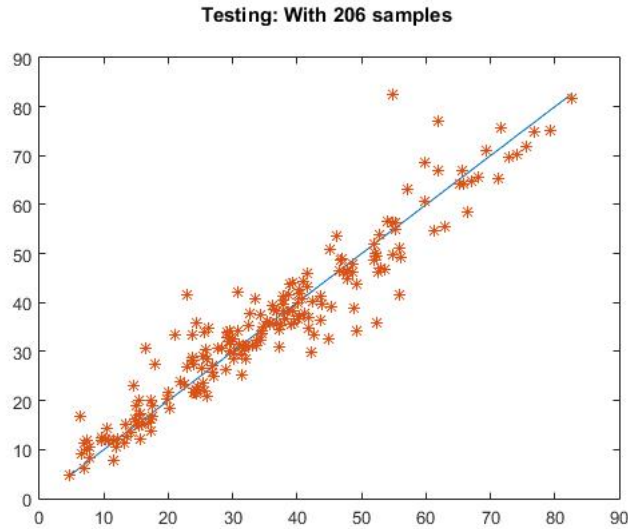


Figure 10: Concrete degree of fit.

The degree of fit, although not optimal, is accurate.