Neural Networks

Barney Tannahill and Kyle Kastner

October 16, 2013

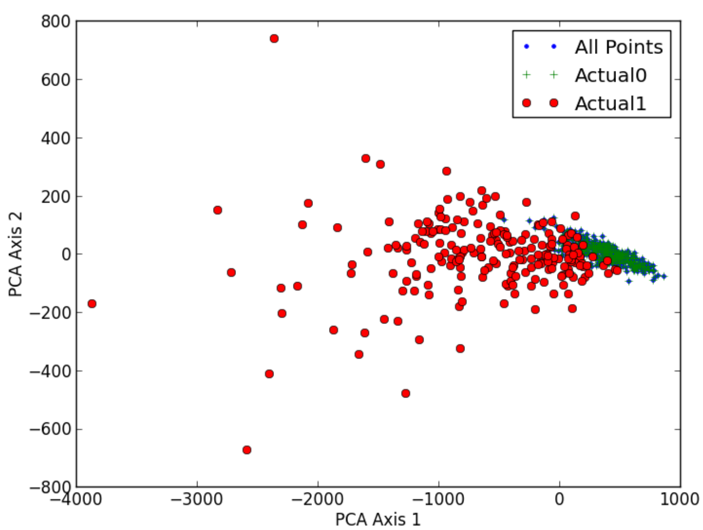
Cancer Dataset

The data set used for this exercise consists of 569 different breast cancer diagnoses. Each diagnosis contains 30 additional measurement parameters.

Feedforward Neural Network Classification

Figure 1 shows the distribution of samples between the two diagnoses with the dimension of the data reduced to two in order to better display the data. In an

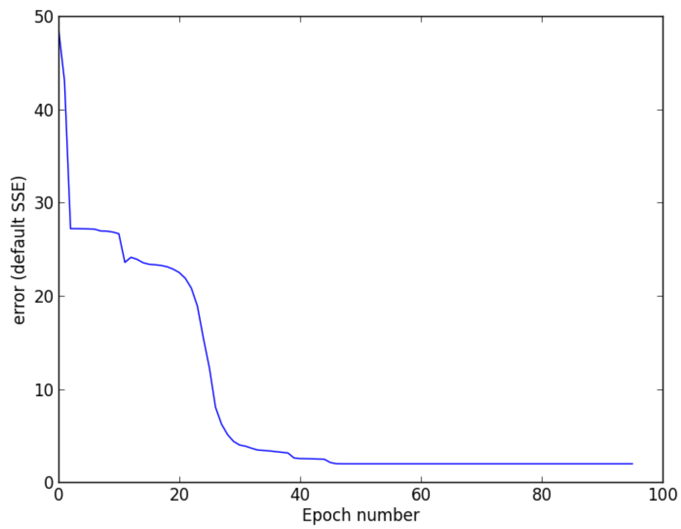
Figure 1: Distribution of samples



attempt to generate a classification method to identify benign and malignant diagnoses based solely on the gathered parameters, the Feedforward Neural Network algorithm available from the neurolab library was used. The network

was chosen to have a single hidden layer of perceptrons and a single output perceptron.
It was decided that 60% of the data would be used to train the neural network. Figure 2 shows the training profile, showing the Sum Squared Error (SSE) of the network decreasing over the training epochs.

Figure 2: Training Profile

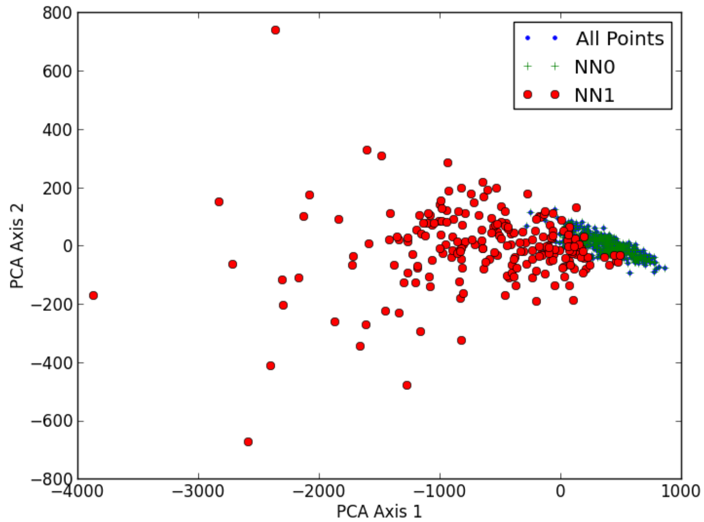


These additional statistics were calculated to help show the effectiveness of the method at classifying the diagnosis data as benign or malignant. For the statistics below, if the neural network output was greater than 0.5, the test result was counted as a malignant indication, and if the output was less than 0.5, the result was counted as a benign indication.

```
True Malignant Identification (%): 0.981
False Malignant Identification (%): 0.019
True Benign Identification (%): 0.983
False Benign Identification (%): 0.017
```

These statistics show that the categorization is accurate at classifying both malignant and benign diagnosis samples. Figure 3 shows the same data as Figure 1, but the data is separated based on the output of the neural network. It can be seen that only minimal differences are present, again demonstrating the accuracy of the neural network classification method.

Figure 3: Network Results



Object Recognition

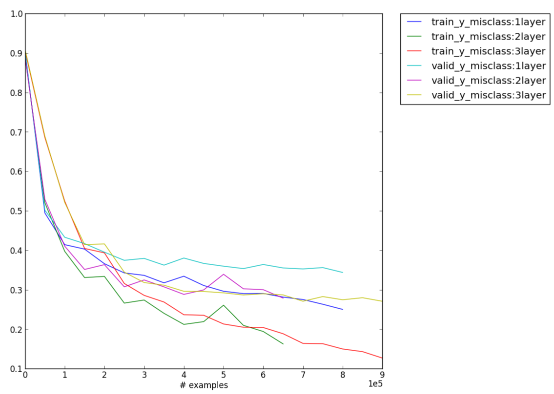
Convolutional neural networks were used to perform object recognition on two different datasets Each set contains 60000 images with 10 separate classes. The MNIST dataset is made handwritten digits (0 - 9) in greyscale, with each image size 28x28 pixels. The CIFAR10 dataset is made of tiny color images (RGB), with each image size 32x32 pixels. The classes for the CIFAR10 datset are shown below.

```
airplane
automobile
bird
cat
deer
dog
frog
horse
ship
truck
```

The current state of the art for CIFAR10 on unaugmented data is 18%. Our best result to date does not match this, but there are further optimizations to be made. Our MNIST unaugmented data results match the state of the art.

```
CIFAR10 1 Layer:0.31
CIFAR10 2 Layer:0.28
CIFAR10 3 Layer:0.25
MNIST 2 Layer:0.0069
```

Figure 4: CIFAR10 Training Schedule



References

[1] Dr. William H. Wolberg, W. Nick Street, Olvi L. Mangasarian *Breast Cancer Wisconsin (Diagnostic) Data Set* UCI Machine Learning Repository http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+i%28Diagnostic%29

[2] *Sci-kit Fuzzy Library* https://github.com/scikit-fuzzy

[3] *Comparison of LDA and PCA 2D projection of Iris Dataset* http://scikit-learn.org/stable/auto_examples/decomposition/plot_pca_vs_lda.html#example-decomposition-plot-pca-vs-lda-py

[4] https://code.google.com/p/neurolab

[5] A. Krizhevsky *Learning Multiple Layers of Features from Tiny Images, 2009* http://www.cs.toronto.edu/~kriz/cifar.html

[6] Y. LeCun, C. Cortes, C. Burges *The MNIST database of handwritten digits* http://yann.lecun.com/exdb/mnist/

Figure 5: CIFAR10 Learned Filters

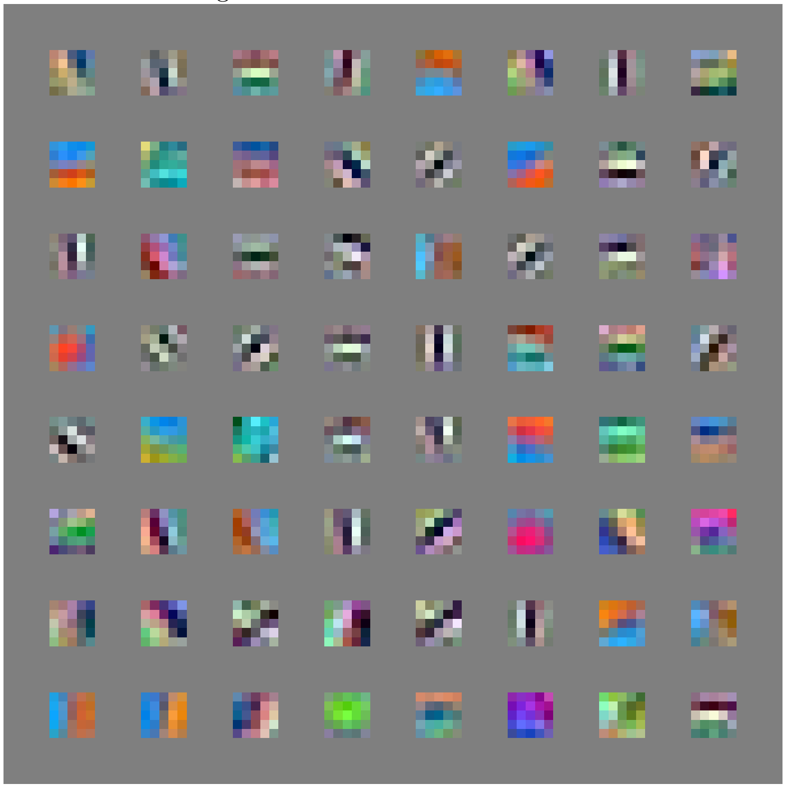


Figure 6: MNIST Training Schedule

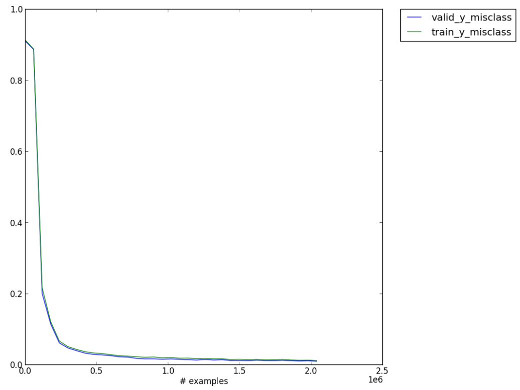


Figure 7: MNIST Learned Filters