

ECS 171 Homework 2 Report

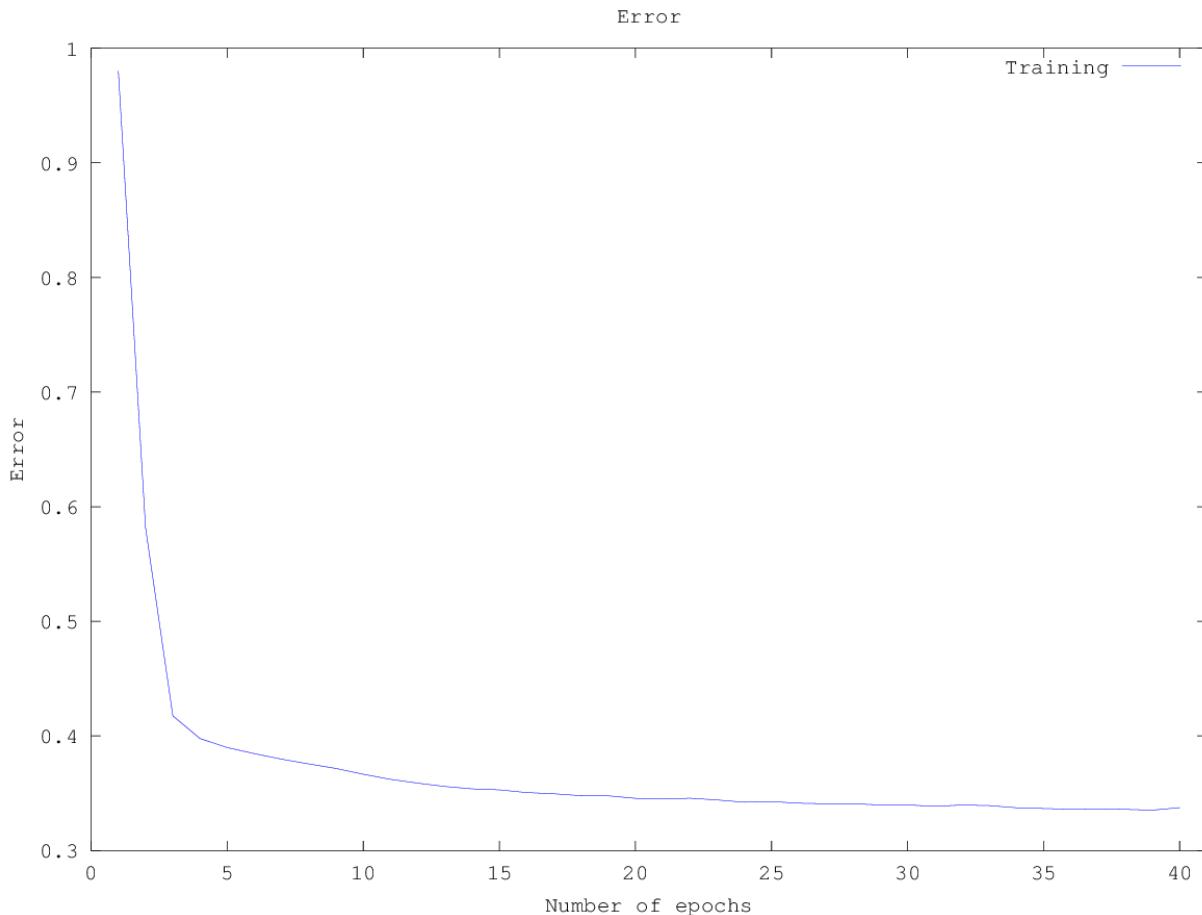
Parker Temple, 997223534, Section 1

November 2, 2014

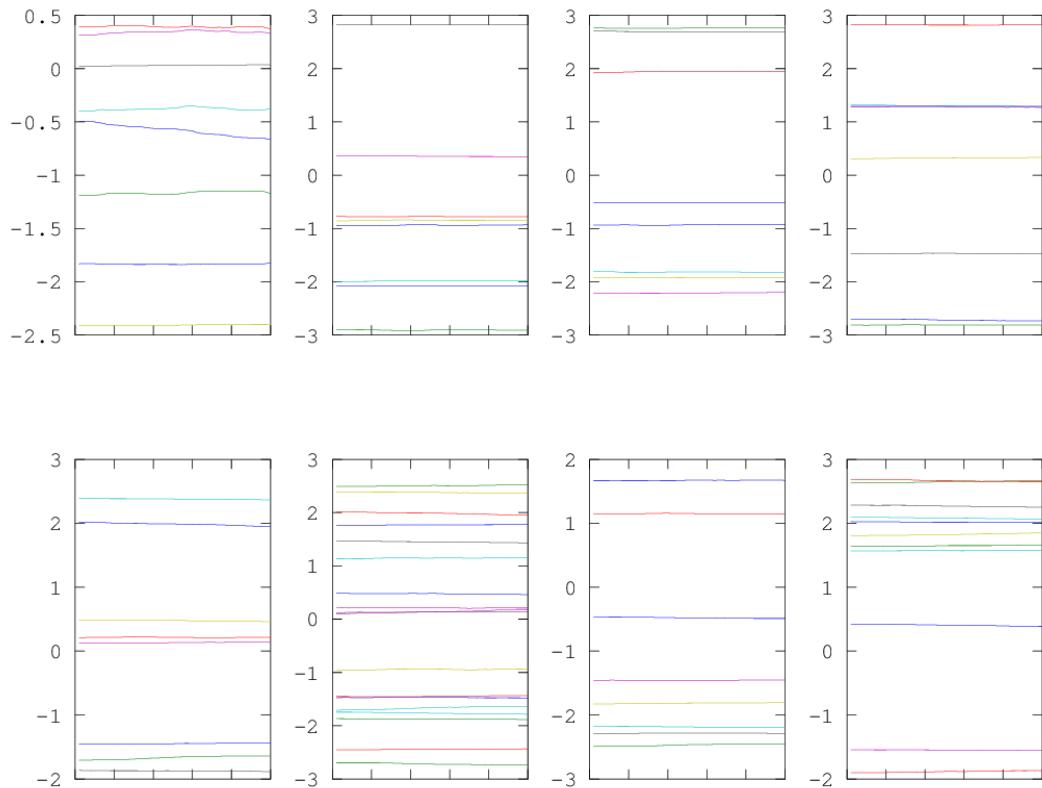
1 3 layer ANN. Data Split into random 75% training set, 25% testing set.

Testing Error: 0.522911

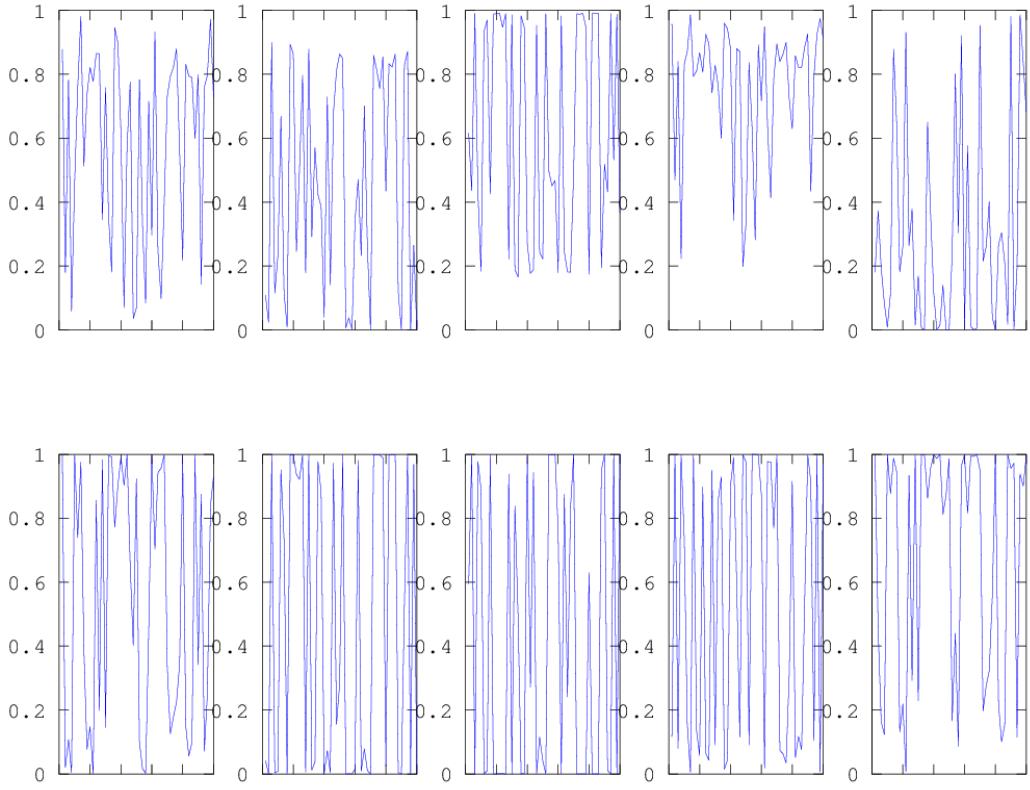
Error change over all batches



Change of 67 weights for first 50 iterations. 8 weights per graph, the bottom right graph, which has 11 weights on it.



Change of the 10 output activations for first 50 iterations. 1 activation per graph.



2 Training error, α_i equations

Training Error: 0.449461

Activation function equations

$$\begin{aligned}
 a_1^{(2)} &= g(2.164202 \cdot x_0 + 0.176168 \cdot x_1 + 0.213492 \cdot x_2 + 1.503652 \cdot x_3 + 0.215892 \cdot x_4 + -1.486713 \cdot x_5 + 1.820079 \cdot x_6 + -0.042223 \cdot x_7 + 0.035219 \cdot x_8) \\
 a_2^{(2)} &= g(0.543049 \cdot x_0 + 0.230263 \cdot x_1 + 0.790381 \cdot x_2 + -0.202021 \cdot x_3 + 1.330253 \cdot x_4 + 2.516649 \cdot x_5 + 2.473374 \cdot x_6 + 0.390521 \cdot x_7 + -2.431607 \cdot x_8) \\
 a_3^{(2)} &= g(-2.162728 \cdot x_0 + -0.260760 \cdot x_1 + 0.081877 \cdot x_2 + -0.012670 \cdot x_3 + 0.361112 \cdot x_4 + 3.017692 \cdot x_5 + -0.953964 \cdot x_6 + 0.203506 \cdot x_7 + -0.100807 \cdot x_8) \\
 a_1^{(3)} &= g(-0.905029 \cdot x_0 + 0.878869 \cdot a_1^{(2)} + -0.109502 \cdot a_2^{(2)} + 0.665166 \cdot a_3^{(2)}) \\
 a_2^{(3)} &= g(0.370766 \cdot x_0 + 1.329797 \cdot a_1^{(2)} + -0.954598 \cdot a_2^{(2)} + 2.167791 \cdot a_3^{(2)}) \\
 a_3^{(3)} &= g(0.983497 \cdot x_0 + 0.774112 \cdot a_1^{(2)} + 1.339974 \cdot a_2^{(2)} + 3.257448 \cdot a_3^{(2)}) \\
 a_4^{(3)} &= g(0.785000 \cdot x_0 + -2.476845 \cdot a_1^{(2)} + 0.237973 \cdot a_2^{(2)} + 0.008546 \cdot a_3^{(2)}) \\
 a_5^{(3)} &= g(0.120132 \cdot x_0 + -1.080554 \cdot a_1^{(2)} + -0.227997 \cdot a_2^{(2)} + 2.504238 \cdot a_3^{(2)}) \\
 a_6^{(3)} &= g(-2.905498 \cdot x_0 + 1.017531 \cdot a_1^{(2)} + -1.459890 \cdot a_2^{(2)} + 3.087088 \cdot a_3^{(2)}) \\
 a_7^{(3)} &= g(0.169951 \cdot x_0 + -1.243626 \cdot a_1^{(2)} + -0.254528 \cdot a_2^{(2)} + 2.601597 \cdot a_3^{(2)}) \\
 a_8^{(3)} &= g(-2.129364 \cdot x_0 + -0.451775 \cdot a_1^{(2)} + -0.510962 \cdot a_2^{(2)} + 0.661315 \cdot a_3^{(2)}) \\
 a_9^{(3)} &= g(-3.955992 \cdot x_0 + 0.606784 \cdot a_1^{(2)} + 0.561681 \cdot a_2^{(2)} + 0.857540 \cdot a_3^{(2)}) \\
 a_{10}^{(3)} &= g(-0.370148 \cdot x_0 + -1.814781 \cdot a_1^{(2)} + -0.578687 \cdot a_2^{(2)} + 2.764005 \cdot a_3^{(2)})
 \end{aligned}$$

3 Hand calculations

First sample:

X Values

0.91254 1.04527 0.42585 -1.07846 0.55858 -1.65363 0.47010 -0.68026

Y Value

0 0 1 0 0 0 0 0 0 0

Initial weights

[1,1] =

1 1 1 1 1 1 1 1 1

1 1 1 1 1 1 1 1 1

1 1 1 1 1 1 1 1 1

[1,2] =

1 1 1 1

1 1 1 1

1 1 1 1

1 1 1 1

1 1 1 1

1 1 1 1

1 1 1 1

1 1 1 1

1 1 1 1

1 1 1 1

1 1 1 1

Updated weights

[1,1] =

0.99882 0.99893 0.99877 0.99950 1.00127 0.99934 1.00195 0.99945 1.00080

0.99882 0.99893 0.99877 0.99950 1.00127 0.99934 1.00195 0.99945 1.00080

0.99882 0.99893 0.99877 0.99950 1.00127 0.99934 1.00195 0.99945 1.00080

[1,2] =

0.99983 0.99983 0.99983 0.99983

0.99983 0.99983 0.99983 0.99983

1.00000 1.00000 1.00000 1.00000

0.99983 0.99983 0.99983 0.99983

0.99983 0.99983 0.99983 0.99983

0.99983 0.99983 0.99983 0.99983

0.99983 0.99983 0.99983 0.99983

0.99983 0.99983 0.99983 0.99983

0.99983 0.99983 0.99983 0.99983

0.99983 0.99983 0.99983 0.99983

0.99983 0.99983 0.99983 0.99983

$$= a_i = y_1$$

$$\alpha_1^{(1)} = g \left([1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1] \begin{bmatrix} 0.71254 \\ 1.04527 \\ 0.42585 \\ -1.07846 \\ 0.55858 \\ -1.65363 \\ 0.47010 \\ -0.68026 \end{bmatrix} \right) = g(0.00001) = 0.499998$$

$$\alpha_2^{(1)} = 0.499998$$

$$\alpha_3^{(1)} = 0.499998$$

$$\alpha_1^{(2)} = g \left([1 \ 1 \ 1 \ 1] \begin{bmatrix} 0.71254 \\ 0.499998 \\ 0.499998 \\ 0.499998 \end{bmatrix} \right) = g(2.412534) = 0.917778$$

$$\alpha_2^{(2)} = 0.917778$$

$$\alpha_3^{(2)} = 0.917778$$

$$\alpha_4^{(2)} = 0.917778$$

$$\alpha_5^{(2)} = 0.917778$$

$$\alpha_6^{(2)} = 0.917778$$

$$\alpha_7^{(2)} = 0.917778$$

$$\alpha_8^{(2)} = 0.917778$$

$$\alpha_9^{(2)} = 0.917778$$

$$\alpha_{10}^{(2)} = 0.917778$$

$$\delta_1^{(3)} = \alpha_1^{(2)} - y_1 = 0.917778 - 0 = 0.917778$$

$$\delta_6^{(2)} = 0.917778$$

$$\delta_2^{(3)} = 0.917778$$

$$\delta_7^{(3)} = 0.917778$$

$$\delta_3^{(3)} = -0.082222$$

$$\delta_8^{(3)} = 0.917778$$

$$\delta_4^{(3)} = 0.917778$$

$$\delta_9^{(3)} = 0.917778$$

$$\delta_5^{(3)} = 0.917778$$

$$\delta_{10}^{(3)} = 0.917778$$

FREE CHECK REQUEST

$$S_1^{(2)} = d_0^{(2)} (1 - d_1^{(2)}) \sum_{i=1}^6 w_i \cdot s_i$$

$$= (0.49998)(1 - 0.49998)(8.12228)$$

$$= 2.044445$$

$$S_2^{(2)} = 2.044445$$

$$S_3^{(2)} = 2.044445$$

$$w_{10}^{(1)} = 0.92956, w_{20}^{(1)} = 0.97176, w_{30}^{(1)} = 0.92956$$

$$w_{11}^{(2)} = 1 - 0.01(d_0^{(1)} + S_1^{(2)})$$

$$= 1 - 0.01(0.91254 + 2.044445)$$

$$= 0.98134$$

$$w_{21}^{(2)} = 1 - 0.01(d_0^{(1)} + S_2^{(2)})$$

$$= 1 - 0.01(0.91254 + 2.044445)$$

$$= 0.98134$$

$$w_{31}^{(2)} = 0.98134$$

$$w_{12}^{(1)} = 0.97863$$

$$w_{22}^{(1)} = 0.97863$$

$$w_{32}^{(1)} = 0.97863$$

$$w_{13}^{(1)} = 0.99129$$

$$w_{23}^{(1)} = 0.99129$$

$$w_{33}^{(1)} = 0.99129$$

$$w_{14}^{(1)} = 1.0220$$

$$w_{24}^{(1)} = 1.0220$$

$$w_{34}^{(1)} = 1.0220$$

$$w_{15}^{(1)} = 0.98858$$

$$w_{25}^{(1)} = 0.98858$$

$$w_{35}^{(1)} = 0.98858$$

$$w_{16}^{(1)} = 1.0338$$

$$w_{26}^{(1)} = 1.0338$$

$$w_{36}^{(1)} = 1.0338$$

$$w_{17}^{(1)} = 0.99039$$

$$w_{27}^{(1)} = 0.99039$$

$$w_{37}^{(1)} = 0.99039$$

$$w_{18}^{(1)} = 1.0139$$

$$w_{28}^{(1)} = 1.0139$$

$$w_{38}^{(1)} = 1.0139$$

$$w_{10}^{(2)} = 1.01(1 - 0.91254)$$

$$= 0.99082$$

$$w_{20}^{(2)} = 0.99082$$

$$w_{30}^{(2)} = 1.0008$$

$$w_{40}^{(2)} = 0.99082$$

$$w_{50}^{(2)} = 0.99082$$

$$w_{60}^{(2)} = 0.99082$$

$$w_{70}^{(2)} = 0.99082$$

$$w_{80}^{(2)} = 0.99082$$

$$w_{90}^{(2)} = 0.99082$$

$$w_{11}^{(2)} = 0.99541$$

$$w_{21}^{(2)} = 0.99541$$

$$w_{31}^{(2)} = 1.0004$$

$$w_{41}^{(2)} = 0.99541$$

$$w_{51}^{(2)} = 0.99541$$

$$w_{61}^{(2)} = 0.99541$$

$$w_{71}^{(2)} = 0.99541$$

$$w_{81}^{(2)} = 0.99541$$

$$w_{91}^{(2)} = 0.99541$$

$$w_{12}^{(2)} = 0.99541$$

$$w_{22}^{(2)} = 0.99541$$

$$w_{32}^{(2)} = 1.0004$$

$$w_{42}^{(2)} = 0.99541$$

$$w_{52}^{(2)} = 0.99541$$

$$w_{62}^{(2)} = 0.99541$$

$$w_{72}^{(2)} = 0.99541$$

$$w_{82}^{(2)} = 0.99541$$

$$w_{92}^{(2)} = 0.99541$$

$$w_{13}^{(2)} = 0.99541$$

$$w_{23}^{(2)} = 0.99541$$

$$w_{33}^{(2)} = 1.0004$$

$$w_{43}^{(2)} = 0.99541$$

$$w_{53}^{(2)} = 0.99541$$

$$w_{63}^{(2)} = 0.99541$$

$$w_{73}^{(2)} = 0.99541$$

$$w_{83}^{(2)} = 0.99541$$

$$w_{93}^{(2)} = 0.99541$$

$$w_{14}^{(2)} = 0.99541$$

$$w_{24}^{(2)} = 0.99541$$

$$w_{34}^{(2)} = 0.99541$$

$$w_{44}^{(2)} = 0.99541$$

$$w_{54}^{(2)} = 0.99541$$

$$w_{64}^{(2)} = 0.99541$$

$$w_{74}^{(2)} = 0.99541$$

$$w_{84}^{(2)} = 0.99541$$

$$w_{94}^{(2)} = 0.99541$$

4 Change number of hidden layers and number of nodes per hidden layer

3 by 4 matrix of testing errors. Rows are 1 layer, 2 layers, 3 layers. Columns are 3 nodes per layer, 5 nodes, 8 nodes, 10 nodes.

0.54313 0.40970 0.41712 0.38544

0.50943 0.44609 0.42385 0.38544

0.62601 0.47978 0.51011 0.37736

The optimal configuration is 3 hidden layers with 10 nodes per layer.

The relationship between these attributes and the generalization error is that as nodes per hidden layer increase, error decreases, but as the number of hidden layers increases, the error tends to increase. This is probably due to the fact that more layers substantially increases intra-neuron interactions, and is similar to creating a higher order polynomial. This can lead to overfitting.

5 Sample prediction

Unknown sample prediction: MIT

6 Uncertainty Measures

If we average all the other percent likelihoods of the output nodes excluding the one that was the maximum, we can get an average percent uncertainty. The percent uncertainty of the unknown sample is 0.12315. This a simiplisitic model for predicting uncertainty.