Question 1

First calculate y:

$$\begin{aligned} y &= a(w_{13}z_3 + w_{14}z_4) \\ &= a(w_{13}(v_{31}z_1 + v_{32}z_2) + w_{14}(v_{41}z_1 + v_{42}z_2)) \\ &= a(w_{13}(v_{31}(u_{11}x_1 + u_{12}x_2) + v_{32}(u_{21}x_1 + u_{22}x_2)) + w_{14}(v_{41}(u_{11}x_1 + u_{12}x_2) + v_{42}(u_{21}x_1 + u_{22}x_2)) \end{aligned}$$

Then calculate the error:

$$\begin{aligned} E &= \frac{1}{2}(d-y)^2 + E_{prev} \\ \delta &= (d-y)a'(net_i) \\ &= (d-y)\alpha sech^2(\alpha w_{13}z_3 + \alpha w_{14}z_4) \end{aligned}$$

Step 4: propagate error signal backwards

$$\begin{aligned} \Delta w_{13} &= -\eta \nabla w_{13} E(w) \\ &= -\eta \frac{\delta E}{\delta y}\frac{\delta y}{\delta net_y}\frac{\delta net_y}{\delta w_{13}} \\ &= -\eta(d-y)a'(w_{13}z_3 + w_{14}z_4)z_3 \\ &= -\eta(d-y)\alpha sech^2(\alpha w_{13}z_3 + \alpha w_{14}z_4)z_3 \\ \Delta w_{14} &= -\eta(d-y)\alpha sech^2(\alpha w_{13}z_3 + \alpha w_{14}z_4)z_4 \\ \Delta v_{31} &= -\eta \nabla v_{31} E(w) \\ &= -\eta \frac{\delta E}{\delta y}\frac{\delta y}{\delta z_3}\frac{\delta z_3}{\delta net_{z_3}}\frac{\delta net_{z_3}}{\delta v_{31}} \\ &= -\eta \frac{\delta E}{\delta y}\frac{\delta y}{\delta z_3}\frac{\delta z_3}{\delta net_{z_3}} z_1 \\ &= -\eta \frac{\delta E}{\delta y}\frac{\delta y}{\delta z_3} a'(v_{31}z_1 + v_{32}z_2)z_1 \\ &= -\eta \frac{\delta E}{\delta y} a'(w_{13}z_3 + w_{14}z_4)(w_{13}v_{31})a'(v_{31}z_1 + v_{32}z_2)z_1 \\ &= -\eta(d-y)a'(y)a'(w_{13}z_3 + w_{14}z_4)(w_{13}v_{31})a'(v_{31}z_1 + v_{32}z_2)z_1 \\ \Delta v_{32} &= -\eta(d-y)a'(y)a'(w_{13}z_3 + w_{14}z_4)(w_{13}v_{32})a'(v_{43}z_1 + v_{42}z_2)z_1 \\ \Delta v_{41} &= -\eta(d-y)a'(y)a'(w_{13}z_3 + w_{14}z_4)(w_{14}v_{41})a'(v_{41}z_1 + v_{42}z_2)z_1 \\ \Delta v_{42} &= -\eta(d-y)a'(y)a'(w_{13}z_3 + w_{14}z_4)(w_{14}v_{42})a'(v_{41}z_1 + v_{42}z_2)z_2 \end{aligned}$$

Using the above equations propagate the error signal backwards and update the weights as you go.

Question 2

a)

First Equation:

number of samples	average error
10	0.545115141138
20	0.245958005301
30	0.286637438925
40	0.3207765774
50	0.192286288873
60	0.209239831276
70	0.243411233516
80	0.163824200894
90	0.3772575898
100	0.346558296457
110	0.249353776456
120	0.279108234307
130	0.287499210734
140	0.370061850401
150	0.442842574873
160	0.199451224358
170	0.218001441535
180	0.301192509787
190	0.217476980488

It appears that the average error tended to decrease as the number of samples increased until around 50 samples at which point the error starts to increase, probably as the result of overfitting.

Second equation:

number of samples	average error
10	0.0449312608146
20	0.109177596847
30	0.109312547095
40	0.183494780438
50	0.185883223375
60	0.0885110338373
70	0.171174321787
80	0.0986018793229
90	0.0928874629844
100	0.208938778366
110	0.109345307753
120	0.18107348994
130	0.231213524392
140	0.165746310569
150	0.153850856933
160	0.120771236064
170	0.135302449973
180	0.154132737489
190	0.119314927447

It appears that the average error tended to decrease as the number of samples increased until around

100 samples at which point the error starts to increase, probably as the result of overfitting.

b)

For the first equation, when evaluated over 100 data samples and within 1-100 hidden nodes the best error was found at 56 nodes. The data seems to cycle the error values seemed to cycle.

For the second equation, when evaluated over 100 data samples and within 1-100 hidden nodes the best error was found at 11 nodes. The data seems to cycle the error values seemed to cycle.

c)

It appears that the error values oscillate cyclically, there is a probably an optimal value to be found before the network starts to get too finely tuned.

Question 3

The number of nodes

2	0.154132737489
6	0.109345307753
8	0.0928874629844
12	0.183494780438
20	0.231213524392

The performance of the system seems to peak around 8 hidden neurons. This seems to imply that there the number of neurons can inversely effect the accuracy of the system if there are too many. This could have the same effect as overfitting.