Homework 5/Practice final questions – Due Thu Dec 6 at beginning of class

1) If the error function for a neural network is

$$E = 1/2 \sum_p (y_p - t_p)^2$$

a) What is dE/dy ?

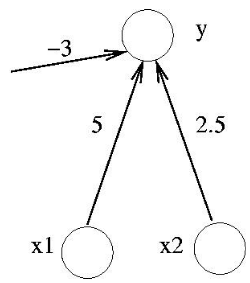


Figure 1: Figure for Question 1

In the figure above assume that w1 is the weight from x1 to y and w2 is the weight from x2 to y.

b) What is $\partial E/\partial w1$ for a linear activation function?

c) What is $\partial E/\partial w1$ for a sigmoid activation function?

d) What is $\partial E/\partial b$ for a sigmoid activation function? (b is the bias weight)

e) Write down the update rule for b

$$b(new) =$$

2)

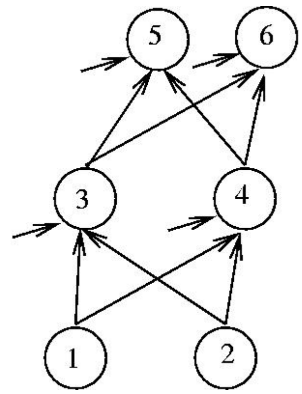


Figure 2: Figure for Question 2

a) How many output units/input units/hidden units are there in the network?

b) Complete the equation for 1/2 the square error in this network (the 1/2 is already entered for you). HINT– look at question 1 but note that there are two output units here.

$$E = 1/2$$

c) Using a LINEAR activation for the output units and the error function you derived in b) above, write down the equation for the derivative of the Error with respect to w_{53}. Use the standard notation that w_{ij} is the connection from unit j to unit i.

$$\partial E/\partial w_{53} =$$

d) Using a LINEAR activation for the output units and the error function you derived in b) above, write down the equation for the derivative of the Error with respect to y_3. HINT: note that y_3 will affect both y_5 and y_6. This is the hardest/longest question on this HW.

$$\partial E/\partial y_3 =$$

e) Using the answer above, write down the equation for $\partial E/\partial net3$ where net3 is the weighted summed input to unit 3. You may assume sigmoidal activation functions for the hidden units or you may assume a generic activation function $y_3 = f(net_3)$. HINT: Do not redo all the work you did above – use your completed answer from above.

$$\partial E/\partial net_3 =$$

3)

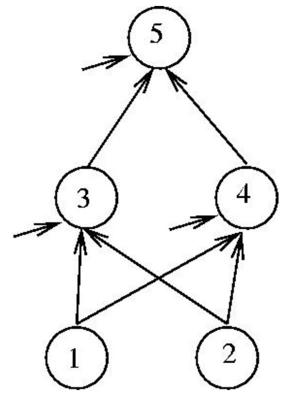


Figure 3: Figure for Question 3

In the figure above use standard notation that w_{ij} is the connection from unit j to unit i.

$$\partial E/\partial w_{31} = \sum_p (y_5 - t_5) f'(net_5) w_{53} f'(net_3) x_1 = -\delta_3 * x_1$$

a) Assume that this equation computes to the value "a". That is $\partial E/\partial w_{31} = a$ for some number a. Write down the update rule for w_{31} in terms of a.

$$w_{31}(new) =$$

b) Using δ_3 in the answer complete the equation for

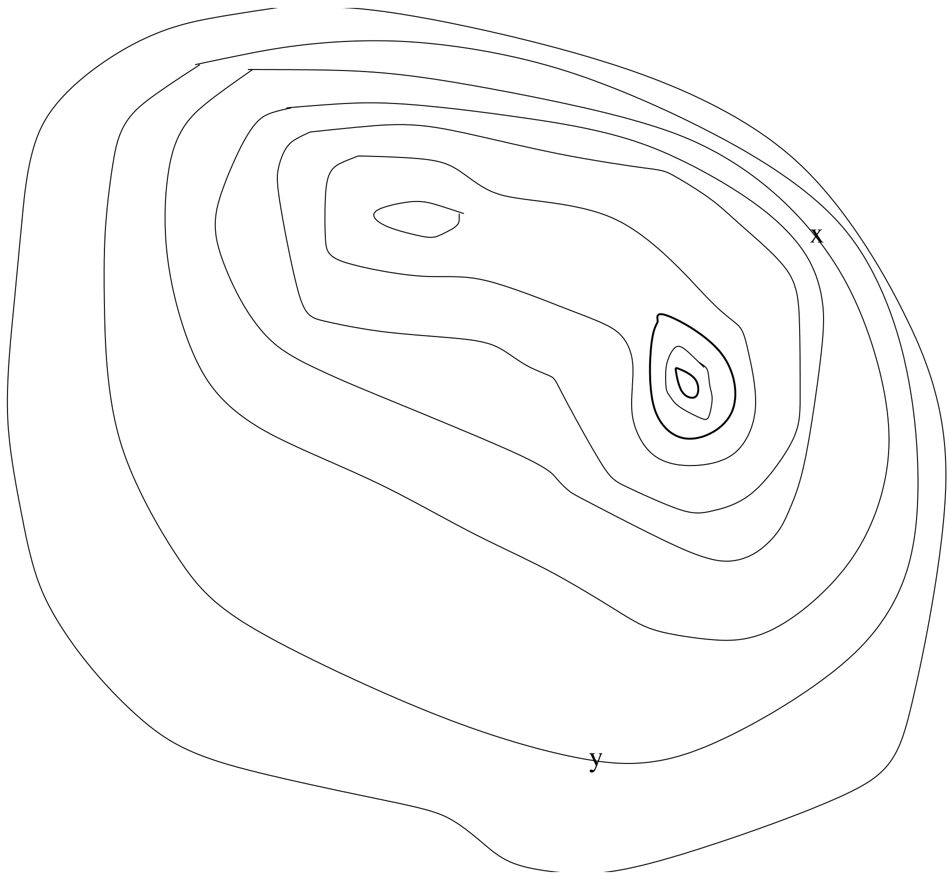
$$\partial E/\partial w_{32} =$$

c) Write down the equation for how δ_3 is computed from δ_5.

$$\delta_3 =$$

d) For the network in Figure 2, write down the equation for how δ_3 is computed from other δs.

$$\delta_3 =$$



4)

Consider the contour plot above (where contours are drawn at even spacing of the Error function). Assume that the contour plot is a good representative of the error surface (that the surface varies smoothly between the contours) and that the highest contour is the outside one.

1. (1) Mark the Global minimum with an A
2. (1) Mark the obvious other local minima with a B
3. (1) from point x draw a step that gradient descent would make (assume a reasonable step size). Mark the end point with a new x.
4. (2) from point y draw the result of performing a single (but complete) line search in the direction of the negative gradient. (mark the end point with a new y)
5. (1) Consider starting at point x or point y and performing regular (steepest) gradient descent learning. From where (x or y) could a larger learning rate (η) be reasonably used?