CS5102: Solution Mid-1

Problem 1

(a): Use the relation:

$\Delta w_i = \eta (y - o) x_i$ for input layer update

~~$\Delta w_1 = (0.5)$~~

First perform forward pass for the data point:

$y = -0.2 + (0.6)(-2) + (0.9)(1)$

$= -0.9 \rightarrow 0$

so there will be an update [5 points]

$\Delta w_0 = 0.5(1)(1)$ [5 pts]

$\Delta w_1 = 0.5(1)(-2)$ [5 pts]

$\Delta w_2 = 0.5(1)(1)$ [5 pts]

(b): One perceptron needed to implement ~~$x_2 \leq x_1$~~, $x_2 \leq x_1$ and one for $x_2 \geq 2$, then joining by AND

Demonstrating that the student gets above part right [8 points]

P1: $x_2 \leq x_1$, $x_1 - x_2 \geq 0$ so weights $[0, 1, -1]$

P2: $x_2 \geq 2$, $x_2 - 2 \geq 0$ so weights $[-2, 0, 1]$

These can be joined by AND in P3 with weights $[-1.5, 1, 1]$ to implement equation

$x_1 + x_2 - 1.5 \geq 0$

Problem 2: (a)

1): Calculate δ_3

$$\delta_3 = o_3(1-o_3)(y-o_3)$$

and calculate * check this

$\Delta w_3 = \eta \delta_3 o_1$, change w_3 to $w_3' = w_3 + \Delta w_3$ [6 pts]

2):

Calculate

$$\delta_1 = o_1(1-o_1) \sum_k \delta_k w_{k1}$$

in this case, it will be: *

$$\delta_1 = o_1(1-o_1)[\delta_3 w_3]$$ check this [6 pts]

3) Calculate

$$\Delta w_1 = \eta \delta_1 x_1$$ [3 pts]

Doing in correct order -[5 pts]

(b):

Need to go back to backpropagation derivation.

$$\Delta w_i = -\eta \frac{\partial E}{\partial w_i}$$

Now we have, $E' = E + \Sigma w_i^2$, so

$$\frac{\partial E'}{\partial w_i} = \frac{\partial E}{\partial w_i} + 2w_i \quad \rightarrow (1)$$

For the output layer, the new update will become

$$\Delta w_3' = -\eta \frac{\partial E'}{\partial w_3} = -\eta\left(\frac{\partial E}{\partial w_3} + 2w_3\right)$$

$$\Delta w_3' = -\eta \frac{\partial E}{\partial w_3} - 2\eta w_3 = \Delta w_3 - 2\eta w_3 \rightarrow (2)$$

Similarly, Δw_1 will become

$$\Delta w_1' = \Delta w_1 - 2\eta w_1. \rightarrow (3)$$

Notice that this will have to be applied in backpropagation order i.e $\Delta w_3'$ depends on $\Delta w_1'$

Understanding (1) ⇒ [7 points]

Understanding (2) [7 points]

Understanding (3) [6 points]

[Students may have done this derivation correctly in other ways so check their Math].

Problem 3 :

(a) : Assuming n independent data points,
$n = d$ will result in overfit.

So to avoid overfit $n >> d$ (no. of free params)

(b):

Total number of free parameters in this case = m

So now the condition changes to $n >> m$