Name : Pavani Rangineni

CWID: A20516359

DATA MINING HW7

1.1.
(a) Use forward propagation to compute the predicted output.

Given x = 4 w1 = 1 w2 = 1 w3 = -1 w4 = 0.5 w5 = 1 w6 = 2
Z1 = xw1 = 4
Z2 = xw2 = 4
Z3 = xw3 = -4
Output for hidden layers z1 = 4 z2 = 4 z3 = 0
Z => 6
a = σ(z)
σ(z=6) = 0.99752

(b) What is the loss or error value?

Loss = $(y – a)^2$
$= (0 – 0.99752)^2$
= 0.9950461

(c) Using backpropagation, compute the gradient of the weight vector, that is, compute the partial derivative of the error with respect to all of the weights.

$\partial L/\partial a$ = -2 (y – a)
= -2 (0 – 0.9975)
= 1.9950

$\partial a/\partial z$ = a (1 – a)
= 0.9975 (0 – 0.9975)
= 0.0024

$\partial L/\partial z = \partial L/\partial a * \partial a/\partial z$
= 1.9950 (0.0024)
= 0.00249

$\partial L/\partial w_4 = \partial L/\partial z * \partial L/\partial w_4$
= 0.0199

$\partial L/\partial w_5 = \partial L/\partial z * \partial L/\partial w_5$
= 0.0199

$\partial L/\partial w_6 = \partial L/\partial z * \partial L/\partial w_6$
= 0

$\partial L/\partial z_1 = \partial L/\partial z * \partial L/\partial z_1$
= 0.0024

$\partial L/\partial z_2 = \partial L/\partial z * \partial L/\partial z_2$
= 0.0049

$\partial L/\partial z_3 = \partial L/\partial z * \partial L/\partial z_3$
$= 0.0098$

$\partial L/\partial w_1 = 0.0046$

$\partial L/\partial w_2 = 0.0196$

$\partial L/\partial w_3 = 0.0392$

$w_1 = 1 - (0.1)(0.0096) = 0.99904$

$w_2 = 1 - (0.1)(0.0196) = 0.99804$
$w_3 = -1 - (0.1)(0.0392) = -1.00392$

$w_4 = 0.5 - (0.1)(0.0392) = 0.49801$

$w_5 = 1 - (0.1)(0.0199) = 0.99801$

$w_6 = 2 - (0.1)(0) = 2$

newly calculated weights are

$w_1 = 0.9904$ $w_2 = 0.99804$ $w_3 = -1.00392$ $w_4 = 0.49801$ $w_5 = 0.99801$ $w_6 = 2$

Forward computation
$z_1 = xw_1 = 4(0.9904)$
$= 0.399616$

$z_2 = xw_2 = 3.99216$

$z_3 = xw_3 = -4.01568$

$z = 1.99808 + 3.99216 + 0$
$z = 5.99024$

(d) Using a learning rate of 1.0, compute new weights from the gradient. With the new weights, use forward propagation to compute the new predicted output, and the loss (error).

$a = \sigma(z) = 0.9975031$
$L(y,a) = (y – a)^2 = 0.99501$

(e) Comment on the difference between the loss values you observe in (b) and (d).

First output = 0.99752
Output after update = 0.9975031
$Loss_1 = 0.99504$
$Loss_2 = 0.99501$
So $loss_2 < loss_1$ and output after update is closer to target (0).

1.2.14 For each of the Boolean functions givenbelow, state whether the problem is linearly separable.

(a) A AND B AND C

Linearly separable

(b) NOT A AND B

Linearly separable

(c) (A OR B) AND (A OR C)

Linearly separable

(d) (A XOR B) AND(A OR B)

Not Linearly separable

1.2.15 (a) Demonstrate how the perceptron Model can be used to represent the AND and OR functions between a pair of Boolean variables.

Let a and b be a pair of input variables. And let c be the output variable.

For AND function, a possible perceptron model is: c = sgn [a + b - 1.5]

For OR function, a possible perceptron model is: c = sgn [a + b - 0.5]

(b) Comment on the disadvantage of using linear functions as activation functions for multi-layer neural networks

Multilayer neural networks are helpful for simulating nonlinear interactions between the properties of the input and output. Linear functions can be employed as activation functions, and the output is still a linear combination of the input characteristics. This kind of network is as expressive to a perceptron.

1.3 Consider a dataset that has 8 predictors. You train a neural network with 3 hidden layers and an output layer that predicts a continuous value (a regression problem). The first hidden layer has 16 neurons, the second has 8 neurons, and the third has 4 neurons. In this network, how many total parameters will you have?

No of parameters = connection between layers + bias

= (8*16 + 16*8 + 8*4 + 4*1) + 16+8+4+1

= 321