classmate
Date
Page

Assignment - 2

1.1 Gradient descent :

The gradient descent training rule decreases the error function to minimum by changing the weight values.

Let D be the set of training examples

The squared error is given

$$E_d = \frac{1}{2} \sum_{d \in D} (t_d - o_d)^2$$

To minimize Ed we must differentiate it with respect to the weights and equate

$$\therefore \frac{\partial E_d}{\partial w_i} = \frac{\partial}{\partial w_i} \frac{1}{2} \sum_{d \in D} (t_d - o_d)^2$$

$$\frac{\partial E}{\partial w_i} = \frac{1}{2} \sum_{d \in D} \frac{\partial}{\partial w_i} (t_d - o_d)^2$$

$$= \frac{1}{2} \cdot 2 \sum_{d \in D} (t_d - o_d) \frac{d}{\partial w_i} (o_d)$$

$$= \sum_{d \in D} (t_d - o_d) \frac{\partial}{\partial w_i} (w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_n x_n + w_1 x_1^2 + w_2 x_2^2 + \cdots + w_n x_n^2)$$

$$\frac{\partial E}{\partial w_i} = \sum_{d \in D} (t_d - o_d)(-x_{id} - x_{id}^2)$$

we have the following variables,

t_d = target output
O_d = perceptron output
D = set of training examples
x_{id} = value of ith attribute and dth training example.

The updated weight can be determined using,

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

$$= -\eta \sum_{d \in D} (t_d - O_d) - (x_{id} + x_{id}^2)$$

$$\Delta w_i = \eta \sum_{d \in D} (t_d - O_d)(x_{id} + x_{id}^2)$$

Also,

$$w_i^{new} = w_i^{old} + \Delta w_i$$

1.2 Comparing activation function

a. Since the input function $f(x) = x$ we can determine the inputs to the hidden layer neurons by adding the inputs to 3 and 4.

$$In_3 = x_1 w_{31} + x_2 w_{32}$$

$$= w_{31} x_1 + w_{32} x_2$$

$$In_4 = x_1 w_{41} + x_2 w_{42}$$

$$= w_{41} x_1 + w_{42} x_2$$

The hidden function uses an activation function of $h(x)$, hence the output from the neurons 3 and 4 will be

$$Out_3 = h(In_3) = h(w_{31}x_1 + w_{32}x_2)$$

$$Out_4 = h(In_4) = h(w_{41}x_1 + w_{42}x_2)$$

These outputs are given as input to the output along with the weights.

$$\therefore y_5 = h(w_{31}x_1 + w_{32}x_2)w_{53} + h(w_{41}x_1 + w_{42}x_2)w_{54}$$

b) $X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ $W^{(1)} = \begin{pmatrix} w_{31} & w_{32} \\ w_{41} & w_{42} \end{pmatrix}$ $W^{(2)} = \begin{pmatrix} w_{53} & w_{54} \end{pmatrix}$

classmate

Date

Page

$$y_5 = h\left(W^{(2)} * h\left(W^{(1)} * X\right)\right)$$

$$W^{(1)} * X = \begin{pmatrix} w_{31}x_1 + w_{32}x_2 \\ w_{41}x_1 + w_{42}x_2 \end{pmatrix}$$

$$W^{(2)} * h\left(W^{(1)} * X\right) = w_{53}h(w_{31}x_1 + w_{32}x_2) + w_{54}h(w_{41}x_1 + w_{42}x_2)$$

$$\therefore y_5 = h\left(w_{53}h(w_{31}x_1 + w_{32}x_2) + w_{54}h(w_{41}x_1 + w_{42}x_2)\right)$$

$\therefore$ vector format is given by

$$y_5 = h\left(W^{(2)} * h\left(W^{(1)} * X\right)\right)$$

c) $h_s(x) = \dfrac{1}{1+e^{-x}}$ $\qquad h_t(x) = \dfrac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$

We must compute the relationship between hs(x) and ht(x) and show that the parameters differs by a linear transformation.

Adding and subtracting e^{-x} in ht(x) numerator,

$$h_t(x) = \frac{e^{x} - e^{-x} - e^{-x} + e^{x}}{e^{x} + e^{-x}} = \frac{e^{x} + e^{-x}}{e^{x} + e^{-x}} - \frac{2e^{-x}}{e^{x} + e^{-x}}$$

$$= 1 - \frac{2e^{-x}}{e^{x} + e^{-x}}$$

multiplying numerator and denominator by e^{x}

$$1 - \frac{2e^{0}}{e^{2x} + e^{0}} = 1 - \frac{2}{e^{2x} + 1} \quad - (1)$$

$$\sigma(x) = h_s(x) = \frac{1}{1+e^{-x}}$$

$$\therefore h_s(-2x) = \frac{1}{1+e^{2x}} \quad - (2)$$

substituting (2) in (1) we get

$$h_t(x) = 1 - 2h_s(-2x) \quad - (3)$$

$$1 - h_s(x) = 1 - \frac{1}{1+e^{-x}} = \frac{e^{-x}}{1+e^{-x}}$$

multiply numerator and denominator with e^x

$$1 - h_s(x) = \frac{e^0}{e^x + e^0} = \frac{1}{e^x+1} = h_s(-x)$$

$$\therefore \quad 1 - h_s(x) = h_s(-x) \quad — \quad (4)$$

substituting (4) in (3) we get.

$$h_t(x) = 1 - 2h_s(-2x)$$

$$= 1 - 2[1 - h_s(2x)]$$

$$= 1 - 2 + 2h_s(2x)$$

$$\therefore \quad h_t(x) = 2h_s(2x) - 1$$

It is evident from the above equation that $h_t(x)$ differs linearly from $h_s(x)$, and is a rescaled version

Hence a neural net created using $\tanh(x)$ or $\sigma(x)$ can generate the same function.