

Assignment-2

1.1 Gradient descent.

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

let D be the set of training examples

The equated error is given

$$E_{d} = \frac{1}{2} \leq (td - 0d)^{2}$$

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

$$\frac{\partial E \partial}{\partial \omega_{i}} = \frac{\partial}{\partial \omega_{i}} + \frac{1}{2} \underbrace{\sum_{i=1}^{2} (t \partial_{i} - \partial_{i})^{2}}_{2}$$

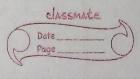
$$\frac{\partial E}{\partial \omega_i} = \frac{1}{2} \underbrace{\frac{\partial}{\partial \omega_i} (t \partial - \partial \partial)^2}_{\text{deg}}$$

$$=\frac{1}{2}.2 \leq (+d-0d)d (od)$$

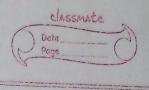
$$\frac{1}{2}ded \frac{1}{2}ded \frac{$$

$$= \underbrace{S(t_2 - 0_d)}_{\partial \omega_i} \underbrace{\partial (\omega_0 + \omega_0 \chi_1 + \omega_2 \chi_2 + \cdots + \omega_n \chi_n)}_{d \in \mathcal{D}}$$

 $\partial t = \frac{5}{360} (td-0d) (-xid-xid^2)$



we have the following variable, td = target output of = perception output D = set of training examples xid = value of ith attribute and ofth training example. The updated reeight can be determined DW; = - n dE = -7 5 (ts-0) - (xis+xis) DW; = y \(\tau (\taid+xid2) Also, W; new = wiold + sw;



1.2 Comparing activation function

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

 $In_3 = \chi_1 w_{31} + \chi_2 w_{32}$ = $w_{31} \chi_{11} + w_{32} \chi_2$

In4 = 21 W41 + 22 W42

= W41 x1 + W42 x2

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

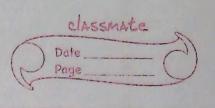
out = h (ln3) = h (W31x1+W32x2)

out = $h(ln_4) = h(w_4|x_1+w_{42}|x_2)$

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

 $\frac{1}{12} = h(w_{31}x_1 + w_{32}x_2)w_{53} + h(w_{4}x_1 + w_{42}x_2)w_{54}$

b) $\chi = \begin{pmatrix} \chi_1 \\ \chi_2 \end{pmatrix}$ $W^{(1)} = \begin{pmatrix} W_{31} & W_{32} \\ W_{41} & W_{42} \end{pmatrix}$ $W^{(2)} = \begin{pmatrix} W_{53} & W_{54} \\ W_{41} & W_{42} \end{pmatrix}$



$$W^{(1)} + X = \left(W_{31} \chi_1 + W_{32} \chi_2 \right)$$

$$W_{41} \chi_1 + W_{42} \chi_2$$

$$W^{(1)} * h (W^{(1)} * X) = W_{53}h(W_{51}x_{1}+W_{52}x_{2})+W_{54}h(W_{41}x_{1}+W_{52}x_{2})$$

$$W_{53}h(W_{31}x_{1}+W_{52}x_{2})+W_{54}h(W_{41}x_{1}+W_{42}x_{2})$$

$$W_{53}h(W_{31}x_{1}+W_{52}x_{2})+W_{54}h(W_{41}x_{1}+W_{42}x_{2})$$

vector format le given by



c)
$$h_S(x) = \frac{1}{1+e^{-x}}$$
 $h_t(x) = e^x - e^{-x}$ $e^x + e^{-x}$

we must compute the relationship between hs(n) and ht(n) and show that the parameters differ by a linear transformation.

Adding and subtracting e in ht(n) numerator,

 $h_{t}(x) = e^{x} - e^{-x} - e^{-x} + e^{x} = e^{x} + e^{-x} - 2e^{-x}$ $e^{x} + e^{-x} = e^{x} + e^{-x}$

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

multiplying numerator and deux nimator by

$$\frac{1 - 2e^{3}}{e^{2x} + e^{0}} = \frac{1 - 2}{e^{2x} + 1} = 0$$

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

substituting D in 1) we get

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



$$1 - hs(x) = 1 - 1 = e^{-x}$$

$$1 + e^{-x}$$

multiply numerator and demoninator with ex

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

$$\frac{1-h_s(x)=h_s(-x)}{-a}$$

$$ht(x) = 1 - 2hs(-2x)$$

$$= 1 - 2 \left[1 - hs(2x) \right]$$

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

$$h_{t}(x) = 2h_{s}(2x) - 1$$

equation that he(x) differs linearly from hs(x), and is a rescaled version

Hence a neural net vicated using tanh(x) or -(n) can generate the same function.