Faiyaz Sundrani fs1459
Ilyas Habeeb mih278

H.W 3
Part-I (Theory)

1. $h(x) = \dfrac{1}{1+e^{-(w^T x + w_0)}}$

if $h(x) > \frac{1}{2}$, then classify as C_1 (1)
else , as C_2 (0)
(or)
$w^T x + w_0 \geq 0$, classify as C_1 (1)
else , C_2 (0)

	x_1	x_2	r
1	-5	3	1
2	2	3	0

Predict $w_2 x_2 + w_1 x_1 + w_0$

At start, $w_0, w_1, w_2 = 0.01$
& $\eta = 0.005$

(a)

$$P(C_1|x^{(1)}) = \frac{1}{1+e^{-(3w_2+5w_1+w_0)}}$$

$\because w_0, w_1, w_2 = 0.01$

$$= \frac{1}{1+e^{-(0.03-0.05+0.01)}}$$

$$= \frac{1}{1+e^{0.01}} = \underline{\underline{0.497}}$$

$$\boxed{\therefore P(C_1|x^{(1)}) = 0.497}$$

(b)

Since $P(C_1|x^{(1)}) < \frac{1}{2}$, the predicted label of $x^{(1)}$ is $\boxed{C_2 \text{ (ie } r=0)}$

(c) $P(C_1|x^{(2)}) = \dfrac{1}{1+e^{-(3w_2+2w_1+w_0)}}$

$= \dfrac{1}{1+e^{-(0.03+0.02+0.01)}}$

$\therefore \boxed{P(C_1|x^{(2)}) = 0.514}$

Since $P(C_1|x^{(2)}) > \frac{1}{2}$, the predicted label of $x^{(2)}$ is $\boxed{C_1 \text{ (ie } r=1)}$

$\therefore$

	x_1	x_2	r (output)	predicted
1	-5	3	1	0
2	2	3	0	1

d) Training Error = % of labels predicted inaccurately

$= \dfrac{\text{\# predicted inaccurately}}{\text{\# total}} \times 100$

$= \dfrac{2}{2} \times 100 = \boxed{100\%}$

(e)

$$\underbrace{Err(w, w_0 | X)}_{\text{Cross Entropy Error}} = -\sum_t \left(r^t \log y^t + (1-r^t) \log (1-y^t) \right)$$

Since $y^{(1)} = 0.497$, $r^{(1)} = 1$

$y^{(2)} = 0.514$, $r^{(2)} = 0$

$$\therefore Err(w, w_0 | X) = -\left[\left(r^{(1)} \log y^{(1)} + (1-r^{(1)}) \log (1-y^{(1)}) \right) + r^{(2)} \log y^{(2)} + (1-r^{(2)}) \log (1-y^{(2)}) \right]$$

$$= -\left[(1 \cdot \log(0.497) + (1-1) \ldots) + (0(\ldots) + (1-0) \log(1-0.514)) \right]$$

$$= -\left[\log(0.497) + \log(0.486) \right]$$

$$= -[-1.4207] = 1.4207$$

$$\boxed{\therefore Err(w, w_0 | X) = 1.4207}$$

P) $w_0 \longleftarrow w_0 + \eta \sum_t (r^t - y^t)$

$w_1 \longleftarrow w_1 + \eta \sum_t (r^t - y^t) x_1$

$w_2 \longleftarrow w_2 + \eta \sum_t (r^t - y^t) x_2$

Since $\eta = 0.005$

initial $w_0, w_1, w_2 = 0.01$

$w_0 \longleftarrow 0.01 + 0.005 \left[(1 - 0.497) + (0 - 0.514) \right]$

$\Rightarrow \boxed{w_0 \longleftarrow 0.009945}$

$w_1 \longleftarrow 0.01 + 0.005 \left[(1 - 0.497) \overset{(-5)}{\wedge} + (0 - 0.514)(2) \right]$

$\Rightarrow \boxed{w_1 \longleftarrow -0.0077}$

$w_2 \longleftarrow 0.01 + 0.05 \left[(1 - 0.497)(3) + (0 - 0.514)(3) \right]$

$\Rightarrow \boxed{w_2 \longleftarrow 0.009805}$

g)

$$y^{(1)} = \frac{1}{1+e^{-(0.009805\cdot 3+0.0077\cdot 5+0.009935)}}$$

$$= \frac{1}{1+e^{-(0.07785)}}$$

$$\boxed{\therefore\ y^{(1)} = 0.519}$$

$$y^{(2)} = \frac{1}{1+e^{-(0.009805\cdot 3-0.0077\cdot 2+0.009935)}}$$

$$= \frac{1}{1+e^{-(0.02395)}}$$

$$\boxed{\therefore\ y^{(2)} = 0.506}$$

$$Err(\omega, \omega_0 | X) = -\log(0.519) - \log(1-0.506)$$

$\therefore$

$\boxed{\therefore \ Err(w, w_0 | \mathcal{X}) = 1.36}$

(h) The cross entropy error ~~the~~ went down after one iteration from 1.4207 to 1.36.

I expected this, since we are optimizing the weights in G.D to reduce our cross-entropy error.

(i) The training error of classifier has gone down from 100% to 50%. since it is classifying $y^{(1)}$ correctly (as $C_1, r=1$)

Q.

a)

Why it happens: Learning rate is too high

Therefore, it oscillates to and fro from the minimum and never converges.

How to fix : Reduce Learning Rate

b)

Why it happens : Learning rate too slow

Thus, will need a lot of iterations to converge

How to fix : Increase Learning Rate

c)

Why it happens : Converges to local minimum instead of global minimum

How to fix : Randomly choose starting points & ~~Ex~~ pick the one with lowest minimum

3.

Ans- Given :

$$\frac{1}{1+e^{-(\omega^T x+\omega_0)}} > 0.30$$

$$1 > 0.30 + 0.30\, e^{-(\omega^T x+\omega_0)}$$

$$0.70 > 0.30\, e^{-(\omega^T x+\omega_0)}$$

$$\frac{0.70}{0.30} > e^{-(\omega^T x+\omega_0)}$$

$$\frac{7}{3} > e^{-(\omega^T x+\omega_0)}$$

Apply log on both sides

$$\log e^{-(\omega^T x+\omega_0)} < \log(7) - \log(3)$$

$$-(\omega^T x+\omega_0) < 1.945910149 - 1.098612287$$

$$\omega^T x+\omega_0 > 0.847297860.$$

Scanned by CamScanner

$\therefore$ When $w^T x + w_0 > \cancel{0.85}\ 0.847$, classify as C_1

else, as C_2

4\. Ans -

Given :

$$Err_{reg}(w, w_0) = \left[-\sum_t \left(r^t \log y^t + (1-r^t) \log (1-y^t) \right) \right] + \frac{\lambda}{2} \sum_{j=1}^{d} w_j^2$$

When $j = 0$,

$$\frac{\partial Err_{reg}}{\partial w_j} = -\sum_t (r^t - y^t)$$

When $j > 0$

$$\frac{\partial Err_{reg}}{\partial w_j} = \cancel{\frac{\lambda}{2} \frac{\partial}{\partial w_j}} - \sum_t (r^t - y^t) x_j + \lambda w_j$$

w.K.T

$$w_j \leftarrow w_j - \eta \frac{\partial Err}{\partial w_j} \quad \text{for } j = 0, 1, \ldots d$$

$$\therefore w_j \leftarrow w_j - \eta \left(-\sum_t (r^t - y^t) \right)$$

$$\Rightarrow \boxed{w_j \leftarrow w_j + \eta \sum_t (r^t - y^t) \quad \text{for } j = 0}$$

$$\therefore w_j \leftarrow w_j - \eta \left(-\sum_t (r^t - y^t) x_j + \lambda w_j \right)$$

$$\Rightarrow \boxed{w_j \leftarrow w_j + \eta \left(\sum_t (r^t - y^t) x_j - \lambda w_j \right) \quad \text{for } j > 0}$$

— — — — — x — — x — — — x — —