Machine Learning Assignment

Part I

1.

$x^1 \{ x_1 = -5 \ \& \ x_2 = 3, \quad r = 1$

$x^2 \{ x_1 = 2 \ \& \ x_2 = 3 \quad r = 0$

$w_0 = w_1 = w_2 = 0.01 \qquad \eta = 0.005$

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

$$W = \begin{bmatrix} w_2 \\ w_1 \\ w_0 \end{bmatrix} \qquad x = \begin{bmatrix} x_2 \\ x_1 \\ 1 \end{bmatrix}$$

$$W^T x = \begin{bmatrix} w2 & w1 & w0 \end{bmatrix} \begin{bmatrix} x_2 \\ x_1 \\ 1 \end{bmatrix}$$

for first training example

$$w^T x = \begin{bmatrix} 0.01 & 0.01 & 0.01 \end{bmatrix} \begin{bmatrix} -5 \\ 3 \\ 1 \end{bmatrix}$$

$$= 0.01 \times -5 + 0.01 \times 3 + 0.01 \times 1 = -0.01$$

$$g(x) = P(C|x^1) = \frac{1}{1+e^{-0.01}} = 0.4975$$

$\therefore$

b) The predicted label is C_2 as the value of $g(x^1)$ is less than $\frac{1}{2}$.

It is inconsistent with the training set

c) $g(x^2) = 2 \times 0.01 + 3 \times 0.01 + 1 \times 0.01$

$= 0.06$

$P(C_1|x^2) = \frac{1}{1+e^{-0.06}} = 0.5149$

$\therefore$ Predicted value is C_1 since $P(C_1|x^2) > \frac{1}{2}$

Again, this is not consistent

d) Since both the labels are predicted wrong, then the error is 100%

(e) Cross Entropy Error $= -\sum_t \left(r^t \log y^t + (1-r^t) \log (1-y^t) \right)$

$= -[1 \cdot \ln(0.4975) + 0$
$+ 0 + \ln(1-0.5149)]$

$= -[\ln(0.4975) + \ln(0.4851)]$

$= 1.4215$

(f) $w_0 \leftarrow w_0 + \eta(r_0^0 - y_0^0) + \eta(r_0^0 - y_0^0)$

$= 0.01 + 0.005(1-0.498) + 0.005(0-0.515)$

$= 0.01 + 0.005 \times .502 - 0.005 \times 0.515$

$= 0.009935$

$w_1 \leftarrow w_1 + \eta(r_1^1 - y_1^1) + \eta(r_1^2 - y_1^2)x_4$

$= 0.01 + 0.005(1-.498) \cdot 5 + 0.005(0-0.515)\cancel{x}(2)$

$= 0.01 + 0.005 \times .502 - 0.005 \times 0.515$

$=$ ~~0.0099935~~

$= -0.0077$

$w_2 \leftarrow w_2 + \eta(r_1^2 - y^2) + \eta(r_1^2 - y_1^2)$

$= 0.01 + 0.005(1-.498) \cdot 3 + 0.005(0.515) \cdot 3$

$= 0.009805$

(g) Cross Entropy Error

$g(x^1) = 0.009805 \times 3 + 0.0077 \times 5 + 0.009935$

$= 0.07785$

$h(x^1) = 0.519$

$g(x^2) = 0.009805 \times 3 - 0.0077 \times 2 + 0.00935 = 0.02395$

$h(x^2) = 0.506$

Scanned with CamScanner

$\therefore Err = -\log 0.519 - \log(1 - 0.506)$

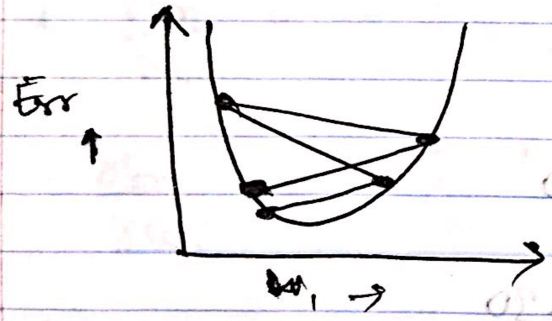
≈ 1.36

(h) Yes, the cross entropy error went down after first iteration.
This is expected since we are finding parameters which minimize the error.

(i) The training error is now 50%, since one of the examples is being predicted correctly.

Q2

(a) This will happen when η is large, since gradient decent will jump over the ~~res~~ minimum of the function.



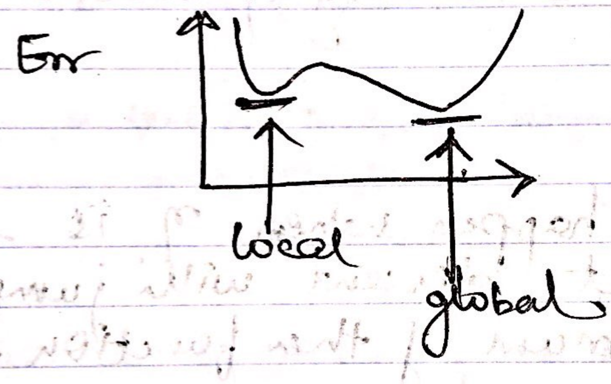
Diagramatically, this is happening:

To fix the problem, we should decrease the learning rate.

(b) This happens because η is too small and it will need many iterations to reduce. To fix it, we should choose a higher η

(c) In this case, the algorithm may have converged to a local minima instead of a global minimum.

To fix this, we can change the starting points for gradient descent.



Q3

Since $h(x) > 0.30$

$$h(x) = \frac{1}{1+e^{-g(x)}} > 0.30$$

$$1+e^{-g(x)} < \frac{1}{0.30}$$

$$\frac{1}{0.30} - 1 > e^{-g(x)}$$

$$-g(x) < \log\left(\frac{1}{.3} - 1\right)$$

$$g(x) > -0.847$$

$$g(x) + 0.847 > 0$$

$$\boxed{= w^T x + w_0 + 0.847 > 0}$$ Ans.

Q4

$$w_j \leftarrow w_j - \eta \frac{\partial Err_{reg}}{\partial w_j} \quad \text{for } j = 1, \ldots, d.$$

$$w_j \leftarrow w_j - \eta \frac{\partial Err_{reg}}{\partial w_j} \quad \text{for } j = 0$$

~~for j≠0~~ Now, we substitute the derivative of the regularised error

$$\frac{\partial Err}{\partial w_j} = -\sum_t (r^t - y^t) x_j + \lambda w_j \quad j = 1, \ldots, d$$

$$\frac{\partial Err}{\partial w_j} = -\sum_t (r^t - y^t) \quad j = 0$$

for $j \neq 0$

$$\frac{\partial Err_{regularized}}{\partial w_j} = \left[\frac{\partial Err}{\partial w_j} + \frac{\partial}{\partial w_j} \left(\frac{\lambda}{2} \sum_{i=1}^{d} w_j^2 \right) \right]$$

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

$$\left[= -\sum_t (r^t - y^t) x_j + \lambda w_j = \frac{\partial Err_{reg}}{\partial w_j} \right] \text{—} ①$$

for $j = 0$

$$\frac{\partial Err_{reg}}{\partial w_j} = \left[\frac{\partial Err}{w_j} + \frac{\lambda}{2} \frac{\partial}{\partial w_j} \sum_{j=1}^{d} w_j^2 \right]$$

$$= -\sum_t (r^t - y^t) + \frac{\lambda}{2} (0)$$

$$\boxed{= -\sum_t (r^t - y^t) = \frac{\partial Err_{reg}}{\partial w_0}} \text{—} ②$$

$\therefore$ Since we are using a learning rate η which is multiplied by $\frac{\partial Err_{reg}}{\partial w_j}$

We get

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

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

Scanned with CamScanner

$$w_j \leftarrow w_j - \eta \frac{\partial E_{rr_{reg}}}{\partial w_j} \quad j=0$$

$$w_j \leftarrow w_j + \eta \left(\sum_t (r^t - y^t) \right) \quad j=0$$

Hence proved.