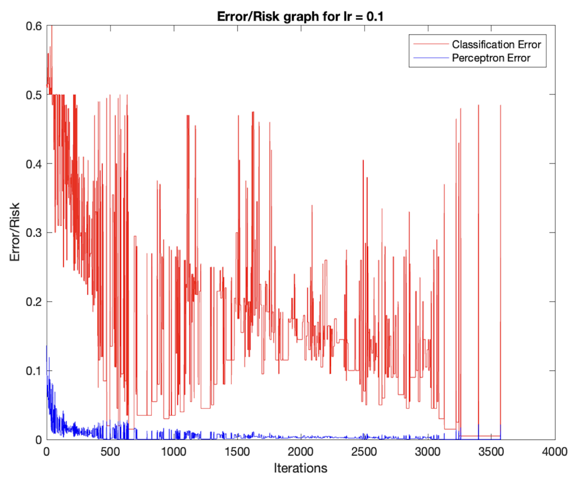
ML Assignment 2

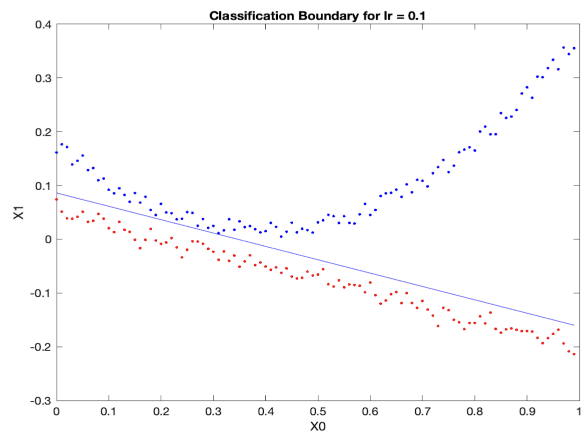
Simran Kucheria (sk11645@nyu.edu)

Problem 1

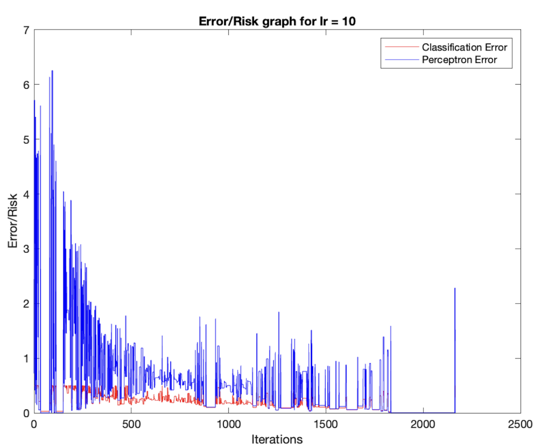
Implemented a linear perceptron with stochastic gradient descent.
Experimented with different values of learning rate and random inititialisation of weights. The error and risk was calculated for each iteration until convergence.

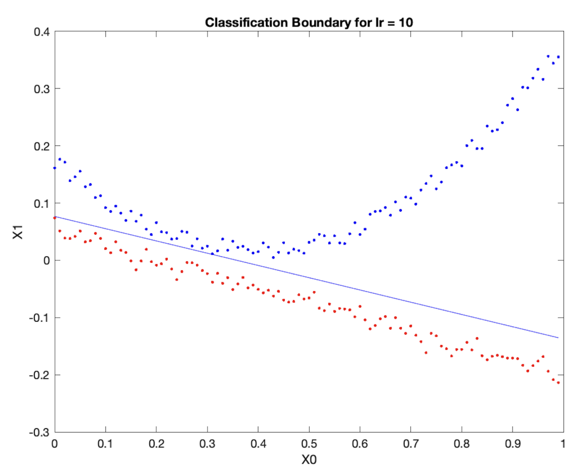
LR = 0.1 Convergence = 3574



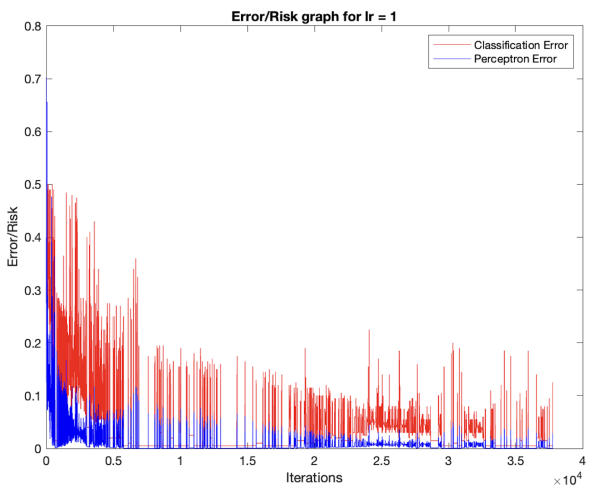


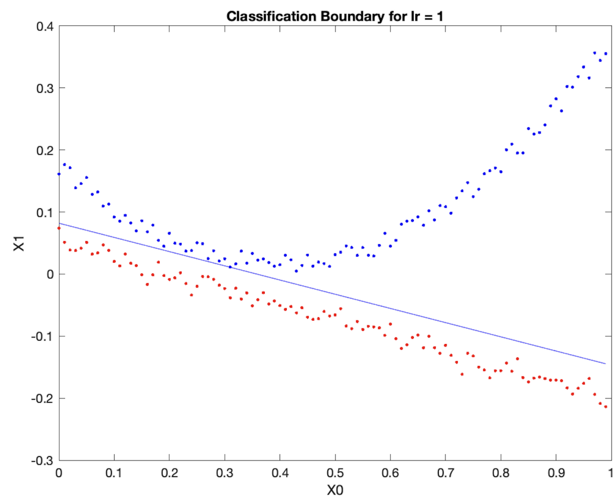
LR = 10 Convergence = 2166





LR = 1 Convergence = 37769





Problem 2

Q2

a. $E = L(x) = -\sum_i t_i \log(x_i) + (1-t_i)\log(1-x_i))$

logistic activation =

output $= x_i = \dfrac{1}{1+e^{-s_i}}$ $\quad \therefore S_i = \sum_j y_j w_{ji}$

hidden $y_j = \dfrac{1}{1+e^{-s_j}}$ $\quad S_j = \sum z_k w_{kj}$.

which is the sigmoid fn.

I. Back Propagation b/w output & hidden layer.

$$\frac{dE}{dw_{ji}} = \frac{dL(x_i)}{d(x_i)} \cdot \frac{d(x_i)}{d(S_i)} \cdot \frac{d(s_i)}{d(w_{ji})}.$$

$$= \frac{d}{d(x_i)}\left(-\sum_i t_i \log(x_i) + (1-t_i)\log(1-x_i)\right) \cdot \frac{d(x_i)}{d(s_i)} \cdot \frac{d(s_i)}{d(w_{ji})}$$

$$= \left(\frac{-t_i}{x_i} + \frac{1-t_i}{1-x_i}\right) \cdot \frac{d(x_i)}{d(s_i)} \cdot \frac{d(s_i)}{d(w_{ji})}. \longrightarrow (1).$$

$\dfrac{d(x_i)}{d(s_i} \rightarrow x_i \Rightarrow$ sigmoid.

$$\frac{d\,\sigma(s_i)}{d(s_i)} = \sigma(s_i)(1-\sigma(s_i)) \rightarrow (2).$$

Sub ② in ①.

$$\left(\frac{1-t_i}{1-x_i} - \frac{t_i}{x_i}\right)\left(x_i \cdot (1-x_i)\right) \cdot \frac{ds_i}{dw_{ji}}$$

$$\left(\left(\frac{1-t_i}{1-x_i}\right) - \frac{t_i}{x_i}\right)\left(x_i \cdot (1-x_i)\right) \cdot y_j.$$

$$\left((1-t_i)x_i - (1-x_i)t_i\right) \cdot y_j.$$

$$(x_i - t_i)\, y_j = \delta_i y_j$$

II Backpropagation b/w hidden layer & input.

$$\frac{dE}{dw_{kj}} = \sum_i \frac{dL(y_j)}{ds_i} \cdot \frac{ds_i}{dy_j} \cdot \frac{dy_j}{ds_j} \cdot \frac{ds_j}{dw_{kj}}.$$

$$= \sum_i \delta_i \cdot w_{ji} \cdot (y_j(1-y_j)) \cdot z_k.$$

$$= \sum_i (x_i - t_i) \cdot w_{ji} \cdot (y_j(1-y_j) \cdot z_k)$$

Q2b. $E = -\sum_i t_i \log(x_i)$.

$$x_i = \frac{e^{s_i}}{\sum_{c=1}^{m} e^{s_c}} \qquad s_i = \sum_j y_j w_{ji}$$

$$y_j = \frac{1}{1+e^{-s_j}} \qquad s_j = \sum z_k w_{kj}$$

I Back Propagation b/w output & hidden layer.

$$\frac{dE}{dw_{ji}} = \frac{dL(x_i)}{d(x_i)} \cdot \frac{d(x_i)}{d(s_i)} \cdot \frac{d(s_i)}{d w_{ji}}$$

$$= \frac{-t_i}{x_i} \cdot \frac{d(x_i)}{d(s_i)} \cdot y_j \longrightarrow (1)$$

$$\frac{d(x_i)}{d(s_i)} = \frac{d\left(\frac{e^{s_i}}{\sum_m e^{s_m}}\right)}{d(s_i)} = \frac{(e^{s_i})' \cdot \sum_m e^{s_m} - \left(\sum_m e^{s_m}\right)' \cdot e^{s_i}}{\left(\sum_m e^{s_m}\right)^2}$$

division rule

$$= \frac{e^{s_i} \cdot \sum_m e^{s_m} - e^{s_i} \cdot e^{s_i}}{\left(\sum_m e^{s_m}\right)^2} = \frac{e^{s_i}}{\sum_m e^{s_m}} \left(1 - \frac{e^{s_i}}{\sum_m e^{s_m}}\right)$$

$$= x_i(1-x_i) \longrightarrow (2)$$

[illegible] = [illegible]$_{kj}$

Sub (2) in (1).

$$\frac{dE}{dw_{ji}} = \frac{-t_i}{\cancel{x_i}} \cdot [\cancel{x_i}(1-x_i)] \cdot y_j \cdot = -t_i(1-x_i) \cdot y_j \cdot$$

$$= \delta_i y_j$$

II Backpropagating b/w hidden layer & output layer

$$\frac{dE}{dw_{kj}} = \sum_i \frac{dL(y_j)}{ds_i} \cdot \frac{ds_i}{dy_j} \cdot \frac{dy_j}{ds_j} \cdot \frac{ds_j}{dw_{kj}} \cdot$$

$$= \delta_i \cdot w_{ji} \cdot [y_j(1-y_j)] \cdot z_K \cdot$$

$$= -\sum_i t_i(1-x_i) \cdot w_{ji} \cdot y_j \cdot (1-y_j) \cdot z_K$$

Problem 3

Q3.

$H = -\sum_{k=1}^{N} p_k \log p_k$ $\Rightarrow$ maximising.

$\Leftrightarrow \sum_{k=1}^{N} p_k \log p_k \rightarrow$ minimise

constraint = probability = 1

$\sum_{k=1}^{N} p_k = 1$ $\qquad \sum_{k=1}^{N} p_k - 1 = 0$.

$\min_{p_k} \max_{\lambda} \sum_{k=1}^{N} p_k \log p_k - \lambda \left(\sum_{k=1}^{N} p_k - 1\right)$

Taking $\frac{\partial}{\partial p}$.

$= \log p_k + 1 - \lambda = 0$. $\quad \therefore \log p_k = \lambda - 1$

$p_k = e^{\lambda - 1}$

~~$\lambda = \log p_k - 1$~~

Setting it in the constraint fn.

$\sum_{k=1}^{N} e^{\lambda - 1} - 1 = 0$

$N e^{\lambda - 1} = 1$.

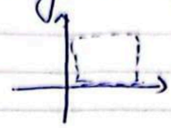
$\log N + \lambda - 1 = 0$

$\lambda = 1 - \log N$

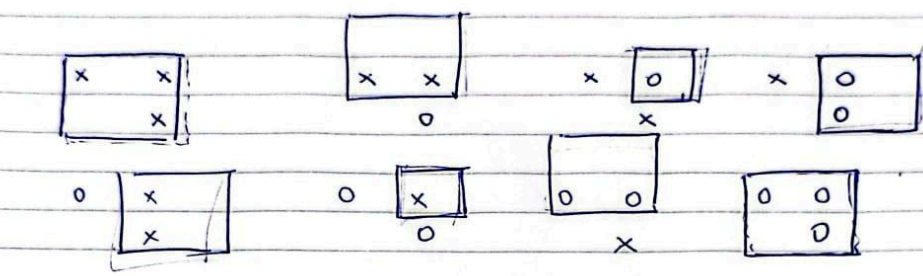
Problem 4

Q4.

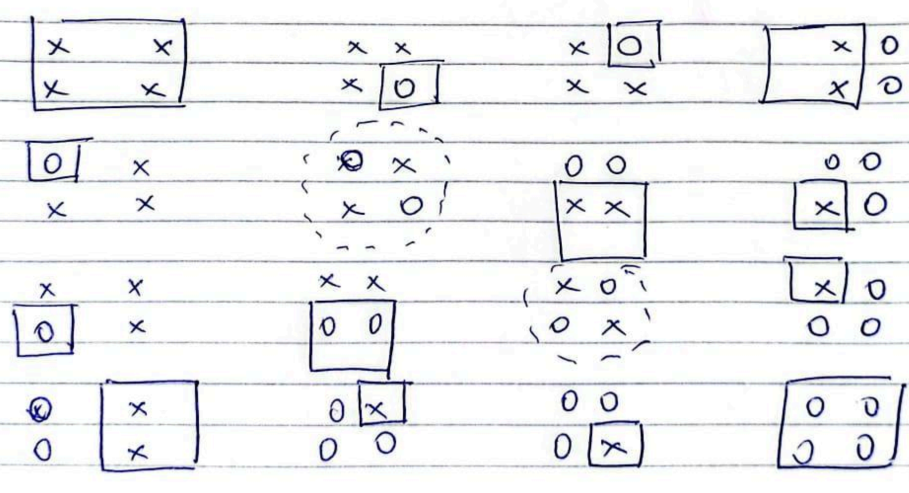
VC dimension of axis aligned squares.

Axis aligned square = 

$n = 3$.



It is possible to shatter 3 points.



Not possible to shatter 4 pts. Hence VC dim $= 3$