Problem 1):

The empirical risk for linear perceptron is given by

$$R(\theta)=\frac{1}{N}\sum_{i=1}^{N} step(-y_i\theta^T x_i)$$

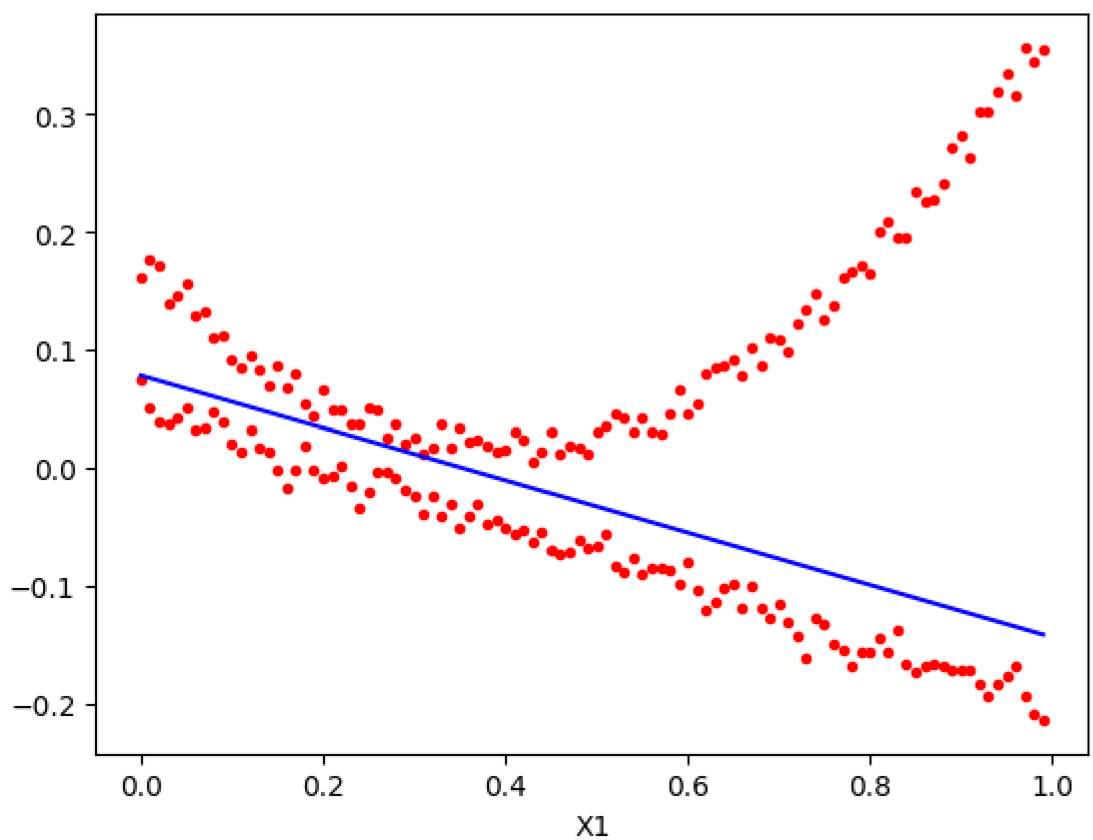
The binary error is defined by

$$E(\theta)=\frac{Total\,Misclassification}{Total}$$

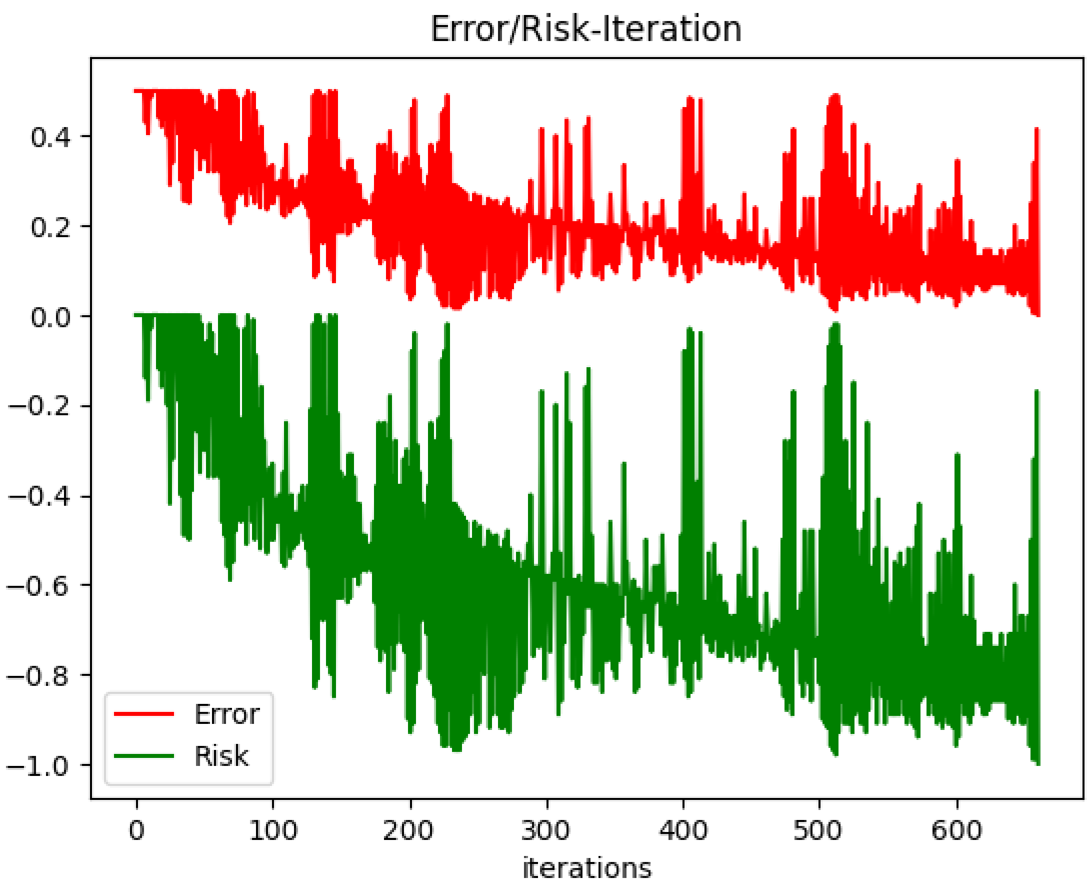
In SGD, if theres a misclassification, you apply the perceptron optimization model

$$\theta^{t+1}=\theta^t-\eta\nabla_\theta R^{per}|\theta_t=\theta^t+y_i x_i$$

The plot of decision boundry



The Risk, Error plot



Problem 2):

Solution 1):

The cross-entropy error is given by

$$E = -\sum_i \left(t_i \log(x_i) + (1 - t_i)\log(1 - x_i)\right)$$

The x_i element here is a sigmoid function, $x_i = \frac{1}{1 + e^{-s_i}}$ where, $s_i = \sum_j y_i w_{ji}$

now, performing backwards propagation on w_{ji} ,

$$\frac{\partial E}{\partial w_{ji}}=\frac{\partial L(x_i)}{\partial x_i}\frac{\partial x_i}{\partial s_i}\frac{\partial s_i}{\partial w_{ji}}$$

$$= \quad \frac{-\partial}{\partial x_i}(-\sum_i (t_i \log(x_i)+(1-t_i)\log(1-x_i)))\frac{\partial x_i}{\partial s_i}\frac{\partial s_i}{\partial w_{ji}}$$

$$= \quad (\frac{1-t_i}{1-x_i}-\frac{t_i}{x_i})\frac{\partial x_i}{\partial s_i}\frac{\partial s_i}{\partial w_{ji}}$$

$$= \quad (\frac{1-t_i}{1-x_i}-\frac{t_i}{x_i})\frac{\partial}{\partial s_i}\sigma(s_i)\frac{\partial s_i}{\partial w_{ji}}$$

$$= \quad (\frac{1-t_i}{1-x_i}-\frac{t_i}{x_i})x_i(1-x_i)\frac{\partial s_i}{\partial w_{ji}}$$

$$= \quad (\frac{1-t_i}{1-x_i}-\frac{t_i}{x_i})x_i(1-x_i)\frac{\partial}{\partial w_{ji}}\sum y_i w_{ji}$$

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

Now, we can denote the term x_i-t_i as δ

$$s_j=\sum_j z_k w_{jk} \quad , \quad y_j=\frac{1}{1+e^{-s_j}}$$

Now performing backwards propagation wrt

$$\frac{\partial E}{\partial w_{ji}}=\frac{\partial E}{\partial s_j}\frac{\partial s_j}{\partial w_{kj}}$$

$$= \quad \sum_i \frac{\partial E}{\partial s_i}\frac{\partial s_i}{\partial s_j}\frac{\partial s_j}{\partial w_{kj}}$$

$$= \quad \sum_i \delta_i \frac{\partial s_i}{\partial s_j}\frac{\partial s_j}{\partial w_{kj}}$$

$$= \quad \sum_i \delta_i \frac{\partial}{\partial s_j}(\sum_{(j)} y_j w_{ji})\frac{\partial s_j}{\partial w_{kj}}$$

$$= \quad \sum_i \delta_i w_{ji} y_j (1-y_j)\frac{\partial s_j}{\partial w_{kj}}$$

$$= \quad \sum_i \delta_i w_{ji} y_j (1-y_j)\frac{\partial}{\partial w_{kj}}\sum_j w_{kj} z_k$$

$$= \quad \sum_{i} \delta_i w_{ji} y_i (1-y_i) z_k$$

$$= \quad \sum_{(i)} (x_i - t_i) w_{ji} y_j (1-y_j) z_k$$

Solution 2:

The modified cross-entropy error is given by

$$E = -\sum_{i} t_i \log(x_i)$$

The softmax activation is given by,

$$x_i = \frac{e^{s_i}}{\sum_{c=1}^{m} e^{s_c}}$$

The sigmoid function is given by

$$\sigma = \frac{1}{1-e^{-x}}$$

performing backwards propagation on w_{ji}

$$\frac{\partial E}{\partial w_{ji}} = \frac{\partial L(x_i)}{\partial x_i} \frac{\partial x_i}{\partial s_i} \frac{\partial s_i}{\partial w_{ji}}$$

$$= \quad \frac{\partial}{\partial x_i} \left(-\sum_{i} t_i \log(x_i)\right) \frac{\partial x_i}{\partial s_i} \frac{\partial s_i}{\partial w_{ji}}$$

$$= \quad \frac{-t_i}{x_i} \frac{\partial x_i}{\partial s_i} \frac{\partial s_i}{\partial w_{ji}}$$

$$= \quad \frac{-t_i}{x_i} \frac{\partial}{\partial s_i} \left(\frac{e^{s_i}}{\sum_{i} e^{s_i}}\right) \frac{\partial}{\partial w_{ji}}$$

$$= \quad \frac{-t_i}{x_i} \left(\frac{(e^{s_i})(\sum_{i} e^{s_i})}{(\sum_{i} e^{s_i})^2} - \frac{e^{x_i} e^{x_i}}{(\sum_{i} e^{s_i})^2}\right) \frac{\partial s_i}{\partial w_{ji}}$$

$$= \quad \frac{-t_i}{x_i} \left(\left(\frac{e^{s_i}}{\sum_{i} e^{s_c}}\right) - \left(\frac{e^{s_i}}{\sum_{i} e^{s_c}} \frac{e^{s_i}}{\sum_{i} e^{s_c}}\right)\right) \frac{\partial s_i}{\partial w_{ji}}$$

$= \quad \frac{-t_i}{x_i}(x_i - x_i^2)\frac{\partial}{\partial w_{ji}}\sum_j y_j w_{ji}$

$= \quad -t_i(1-x_i)\, y_j$

Now, consider the term $\quad -t_i(1-x_i)\, y_j\, as\, \delta$

Now performing backwards propagation wrt $\quad w_{jk}$

$$\frac{\partial E}{\partial w_{ji}} = \frac{\partial E}{\partial s_j}\frac{\partial s_j}{\partial w_{kj}}$$

$= \quad \sum_i \left(\frac{\partial E}{\partial s_i}\frac{\partial s_i}{\partial s_j}\right)\frac{\partial s_j}{\partial w_{kj}}$

$= \quad \sum_i \left(\delta_i \frac{\partial}{\partial s_j}\left(\sum_j y_j w_{ij}\right)\frac{\partial}{\partial w_{kj}}\sum_j w_{jk} z_k\right)$

$= \quad \sum_i \delta_i w_{ji}\, y_i (1-y_i)\, z_k$

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

Problem 3:

The entropy of the distribution is given by

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

for the discrete distribution,

$$\{p_k | k = 1,2,\ldots N\}$$

The equity condition is

$$\sum_{k=1}^{N} p_k - 1 = 0 \qquad \text{--------- (1)}$$

L(X) = $-\sum_{k=1}^{N} p_k \log p_k - \lambda(\sum_{k=1}^{M} p_k - 1)$

$$\frac{\partial L(X)}{\partial p} = \frac{-\partial}{\partial p} \sum_{p_k}^{N} p_k \log p_k - \lambda(\sum_{k=1}^{N} p_k - 1) = 0$$

$$-(1 + \log p) - \lambda = 0$$

$$\lambda = -(1 + \log p)$$

$$\log p = -\lambda - 1 \qquad \text{-------- (2)}$$

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

Substituting the above value in (1)

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

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

$$-\lambda - 1 = \log \frac{1}{N}$$

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

Substituting the above equation in (2)

$$\log p = -\log N + 1 - 1$$

$$\log p = \log \frac{1}{N}$$

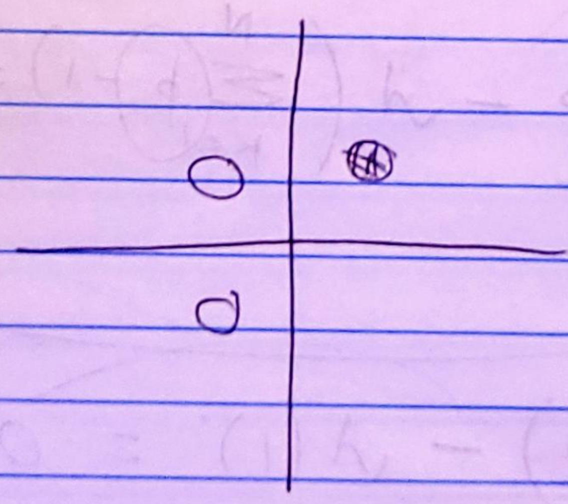
therefore,

$$p = \frac{1}{N}$$

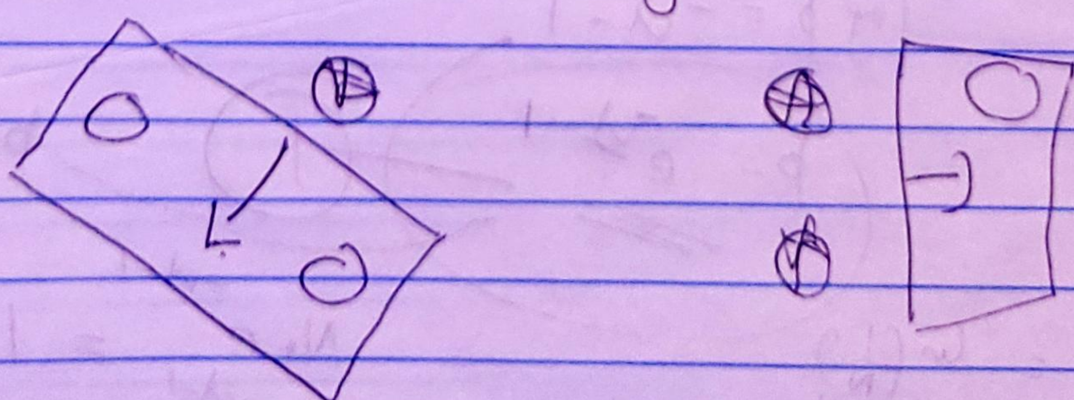
④

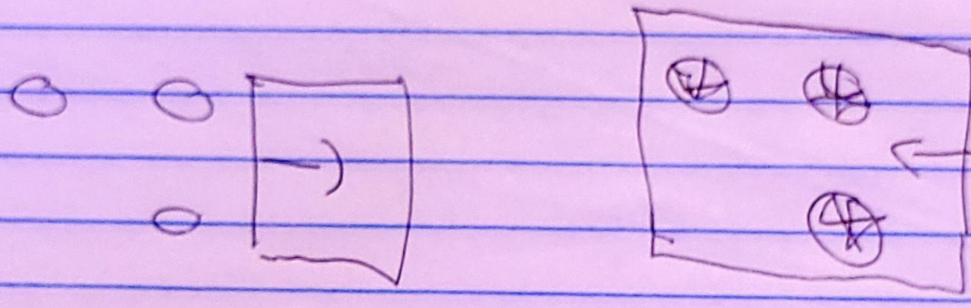
⇒ The VC Dimension of an axis-aligned Square in a 2d plane is 3

→ Consider a x,y coordinate system in which there are 3 dots in first 3 quadrants

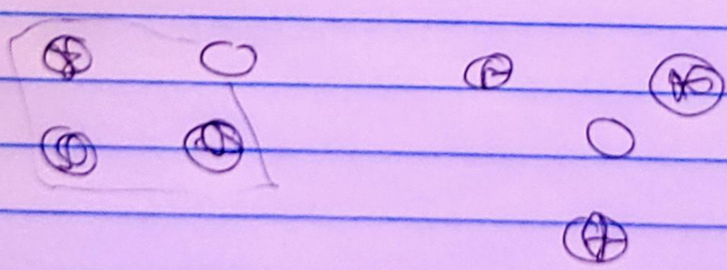


Now, for a given function f(x, θ), you can correctly classify the following distribution





But when theres a 4th point, It cannot be labelled correctly in most of the cases



Thus the VC-dimensions of axis aligned square in the Plane is 3 as No sets of 4 points can be fully Shattered