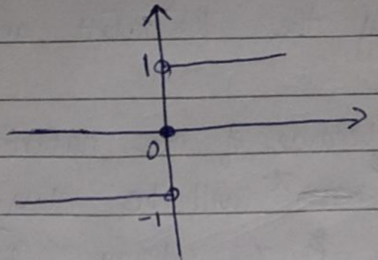
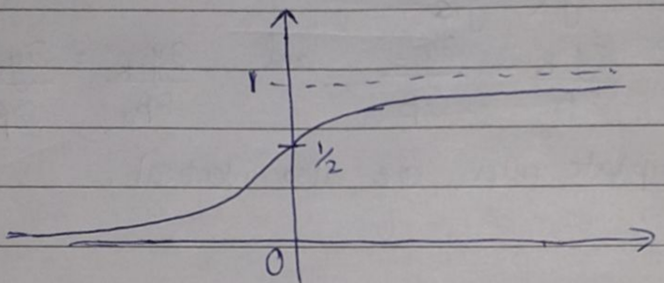
Q5 a) Sign function



Sigmoid function $f(x) = \frac{1}{1+e^{-x}}$



In sign function for all $x > 0$, the values is 1 and in sigmoid function for all $x > 0$, the value is $> \frac{1}{2}$. A similar nature is for $x < 0$.

let parameters in Rosenblatt perceptron be β and β_0 and for the modified perceptron, β and β_0

Update rule for original Rosenblatt perceptron

$$\beta_{new} = \beta_{old} - \eta\left(\frac{\partial y_{iR}}{\partial \beta}\right)$$

$$\beta_0\ new = \beta_0\ old - \eta\left(\frac{\partial y_{iR}}{\partial \beta_0}\right)$$

Update rule for modified perceptron

$$\beta_{new} = \beta_{old} - \eta\left(\frac{\partial y_{is}}{\partial \beta}\right)$$

$$\beta_0 new = \beta_0 old - \eta\left(\frac{\partial y_{is}}{\partial \beta_0}\right)$$

y_{iR} and y_{is} are corresponding to x_i's which have been misclassified by the Rosenblatt and Sigmoid perceptron respectively.

As seen before, because of the nature of 2 activation functions, all x_i will be classified into same class if those x_i satisfy $\beta^T x_i + \beta_0 > 0$ or < 0

Hence the x_i's misclassified by both perceptrons are exactly the same and therefore,

$$y_{iR} = y_{is}$$

$$\Rightarrow \frac{\partial y_{iR}}{\partial \beta} = \frac{\partial y_{is}}{\partial \beta} \quad \text{and} \quad \frac{\partial y_{iR}}{\partial \beta_0} = \frac{\partial y_{is}}{\partial \beta_0}$$

Thus the update rules are also identical.

b) $$\phi(\beta,\beta_0) = -\sum_{i=1}^{N} y_i(\beta^T x_i + \beta_0)$$

$$\frac{\partial(\phi(\beta,\beta_0))}{\partial\beta} = \frac{\partial\left(-\sum_{i=1}^{N} y_i(\beta^T x_i + \beta_0)\right)}{\partial\beta} = -\sum_{i=1}^{N} y_i x_i$$

$$\frac{\partial(\phi(\beta,\beta_0))}{\partial\beta_0} = -\sum_{i=1}^{N} y_i$$

Update rule

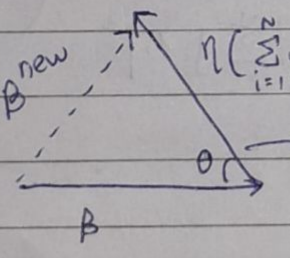
$$\beta^{new} = \beta^{old} - \eta\left(-\sum_{i=1}^{N} y_i x_i\right)$$

$$= \beta + \eta\left(\sum_{i=1}^{N} y_i x_i\right)$$

We have $\|\beta\|_2 = 1$ and we want $\|\beta^{new}\|_2 = 1$

As we cannot change β^{old}/β and $\sum_{i=1}^{N} y_i x_i$ in order to ensure that $\|\beta\|_2 = 1$, we change η only.

By triangle law of vector addition



$$\xrightarrow{} \quad \theta = \cos^{-1}\left(\frac{\langle\beta, \eta\sum y_i x_i\rangle}{\|\beta\| \cdot |\eta\sum y_i x_i|}\right) = \cos^{-1}\left(\frac{\langle\beta, \eta\sum y_i x_i\rangle}{\eta|\sum y_i x_i|}\right)$$

$$\cos\theta = \frac{|\eta(\sum_{i=1}^{N} y_i x_i)|^2 + |\beta|^2 - |\beta^{new}|^2}{2 \cdot |\beta| \cdot |\eta\sum_{i=1}^{N} y_i x_i|}$$

As $|\beta| = |\beta^{new}| = 1$ and η is a scalar

$$\cos\theta = \frac{\eta^{\cancel{2}}|\sum_{i=1}^{N} x_i y_i|^{\cancel{2}}}{2\cancel{\eta} \cdot |\sum_{i=1}^{N} \cancel{y_i x_i}|} = \frac{\eta}{2}\left|\sum_{i=1}^{N} x_i y_i\right|$$

~~$\cos\theta = \langle \beta$~~, so $\eta = \dfrac{2\cos\theta}{\left|\sum_{i=1}^{N} x_i y_i\right|}$

if $\theta > 90^\circ \Rightarrow \eta < 0$ which would mean going in the same direction as gradient and not gradient descent

And It would not be possible to have $\|\beta\|_2 = 1$ while also having $\eta > 1$

where $\cos\theta = \dfrac{\langle \beta, \sum_{i=1}^{N} y_i x_i \rangle}{\left|\sum_{i=1}^{N} x_i y_i\right|}$

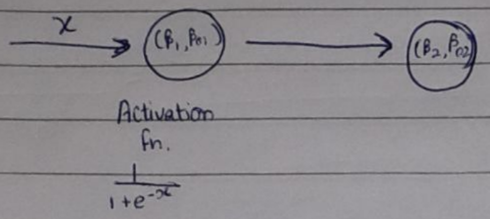
$$\boxed{\eta = \frac{2\langle \beta, \sum_{i=1}^{N} y_i x_i \rangle}{\left|\sum_{i=1}^{N} x_i y_i\right|^2}}$$

Update rule:

$$\beta^{new} = \beta^{old} - \frac{2\langle \beta, \sum_{i=1}^{N} y_i x_i \rangle}{\left|\sum_{i=1}^{N} y_i x_i\right|^2}\left(-\sum_{i=1}^{N} y_i x_i\right)$$

$$\beta_0^{new} = \beta_0^{old} - \frac{2\langle \beta, \sum_{i=1}^{N} y_i x_i \rangle}{\left|\sum_{i=1}^{N} y_i x_i\right|^2}\left(-\sum_{i=1}^{N} y_i\right)$$

Q6)



The output from the 1st perceptron would be ~~β~~ $\frac{1}{1+e^{-(\beta_1 x+\beta_{01})}}$ which would fed to the 2nd perceptron.

On applying the weights and bias ~~of~~ the 2nd perceptron we get,

$$\frac{\beta_2}{1+e^{-(\beta_1 x+\beta_{01})}} + \beta_{02}$$

If for some input x_i it is misclassified then

$$d_i = -y_i\left(\frac{\beta_2}{1+e^{-(\beta_1 x_i+\beta_{01})}} + \beta_{02}\right)$$

$$\Rightarrow \quad \frac{\partial d_i}{\partial \beta_1} = -y_i x_i\left(\frac{\beta_2}{(1+e^{-(\beta_1 x_i+\beta_{01})})^2} e^{-(\beta_1 x_i+\beta_{01})}\right)$$

$$\frac{\partial d_i}{\partial \beta_{01}} = -y_i\left(\frac{\beta_2 \, e^{-(\beta_1 x_i+\beta_{01})}}{(1+e^{-(\beta_1 x_i+\beta_{01})})^2}\right)$$

$$\frac{\partial d_i}{\partial \beta_2} = -y_i\left(\frac{1}{1+e^{-(\beta_1 x_i+\beta_{01})}}\right)$$

$$\frac{\partial d_i}{\partial \beta_{02}} = -y_i$$

Update rules: $(\eta = \text{learning rate})$

$$\beta_1^{new} = \beta_1^{old} + \eta y_i x_i\left(\frac{\beta_2 \, e^{-(\beta_1 x_i+\beta_{01})}}{(1+e^{-(\beta_1 x_i+\beta_{01})})^2}\right)$$

$$\beta_2^{new} = \beta_2^{old} + \eta y_i\left(\frac{1}{1+e^{-(\beta_1 x_i+\beta_{01})}}\right)$$

$$\beta_{02}^{new} = \beta_{02}^{old} + \eta y_i$$

$$\beta_{01}^{new} = \beta_{01}^{old} + \eta y_i\left(\frac{\beta_2 \, e^{-(\beta_1 x_i+\beta_{01})}}{(1+e^{-(\beta_1 x_i+\beta_{01})})^2}\right)$$