ECE/CS 559: Neural Networks Homework 3

Samuele Pasquale

September 17, 2024

These problems were discussed in collaboration with Simone Ughetto, Francesco Presta, Virginia Tasso

1. Question 1:

   (a) The randomly generated weight vector is:

   $$\mathbf{W}^* = \begin{bmatrix} w_0^* \\ w_1^* \\ w_2^* \end{bmatrix} = \begin{bmatrix} 0.12 \\ 0.44 \\ 0.45 \end{bmatrix}$$

   (b) In Figure 1, the plane is shown with the linear separator and the two classes labeled in red and blue. The equation of the line is:

   $$0.12 + 0.44 \cdot x_1 + 0.45 \cdot x_2 = 0$$

   (c) In Figure 1, it is shown how the vector $\begin{bmatrix} w_1^* \\ w_2^* \end{bmatrix}$ is perpendicular to the line $0.12+0.44 \cdot x_1+0.45 \cdot x_2 = 0$.
   To demonstrate that the distance between the origin and the line is equal to $\frac{|w_0^*|}{\sqrt{w_1^{*2}+w_2^{*2}}}$, the point-to-line distance formula is used.

   $$d(P) = d(x_1, x_2) = \frac{|w_0^* + w_1^* x_1 + w_2^* x_2|}{\sqrt{w_1^{*2} + w_2^{*2}}}$$

   Given the point P, the origin of the axes $(0, 0)$, and the line $0.12 + 0.44 \cdot x_1 + 0.45 \cdot x_2 = 0$, it is found that the distance is:

   $$d = \frac{|w_0^*|}{\sqrt{w_1^{*2} + w_2^{*2}}}$$

2. Question 2

   (a) As shown in Figure 2, the vectors w and w^* are not equal because. The two vectors differ because w^* is the linear separator around which the data set for supervised learning was constructed, while w is the vector resulting from the learning algorithm. As can be seen in Figure 2, there is not just one existing boundary for classifying the two different classes (red and blue); rather, there are multiple linear separators. The goal of the perceptron learning algorithm is to find a valid linear separator, and the search ends with the first result that meets this condition. Therefore, as the data set increases, w and w^* will converge to become the same line.

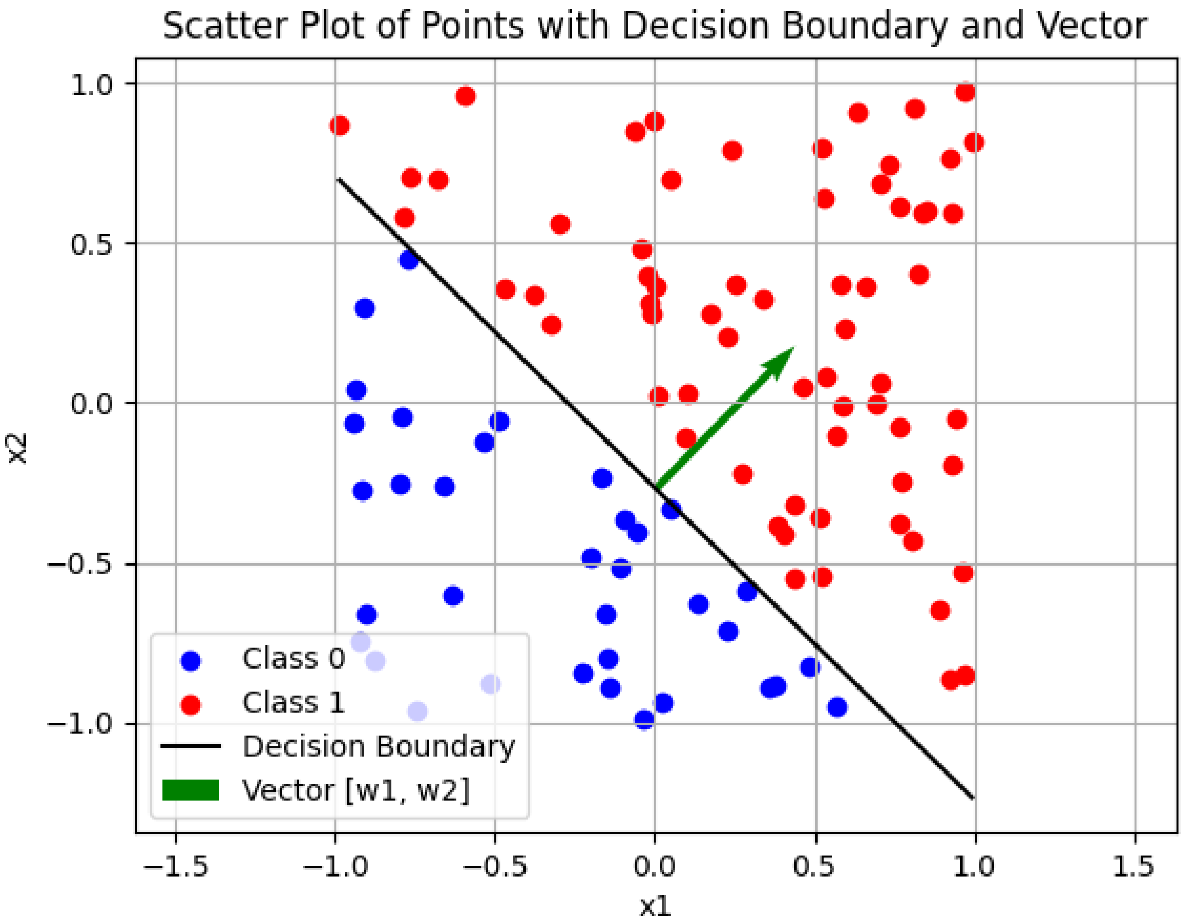


Figure 1: Scatter Plot of Points with Decision Boundary and Vector

(b) From Figure 3, it can be observed how η affects the convergence speed of the algorithm. Analyzing the error trend, it can be seen that with a higher η, the number of errors decreases more rapidly; on the other hand, the more the algorithm converges to the final solution, the more oscillations occur, requiring more epochs to reach the final solution. This behavior is due to the fact that with a higher η, the weights of the weight vector are changed more drastically compared to a lower η. In conclusion, although a higher η allows for a faster reduction of errors in the initial phase, the total number of epochs required to converge to the final solution will be greater.

(c) In Figure 4, the vectors w and w^* are shown with a dataset of 1000 points.

$$\mathbf{W}^* = \begin{bmatrix} w_0^* \\ w_1^* \\ w_2^* \end{bmatrix} = \begin{bmatrix} 0.12 \\ 0.44 \\ 0.45 \end{bmatrix} \qquad \mathbf{W} = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 10.98 \\ 10.96 \end{bmatrix}$$

Analyzing the results obtained in Q2(a) and Q2(c), it is observed that the number of points in the dataset is 10 times higher (from 100 to 1000), but the boundaries remain within $[-1, 1]^2$. Therefore, increasing the number of points also increases the density, given the same square size

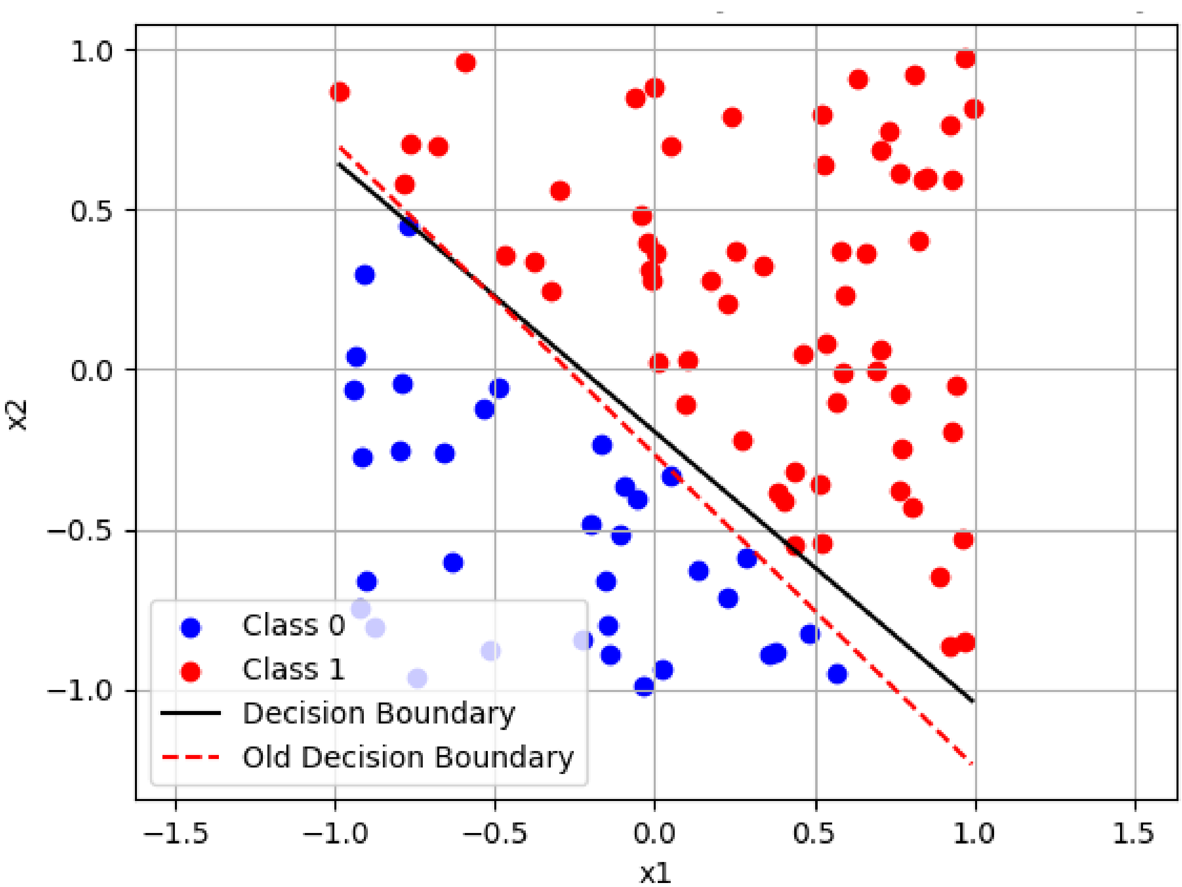


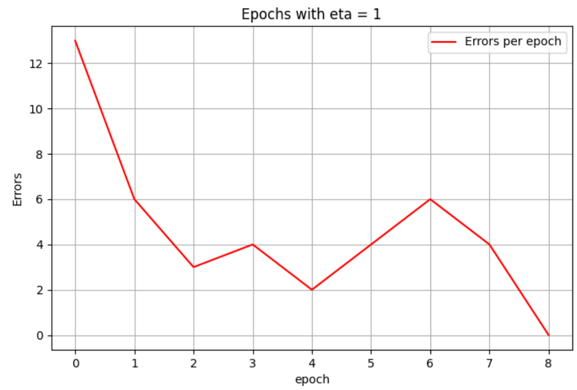
Figure 2: Comparison between the linear separator defined in Q1(a) in black and the boundary separator generated by the Perceptron Learning Algorithm, dashed in red.

of $[-1, 1]^2$. However, increasing the point density adds more constraints on the line W generated by the Learning Algorithm. As a result, the line W will converge more closely to W^* with 1000 points compared to just 100 points.

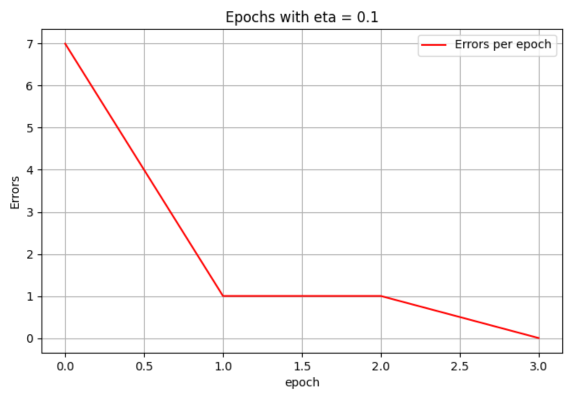
(d) Figure 5 shows the mean errors, as well as the 10th and 90th percentiles, for each epoch with different η values. As can be seen, the trend is similar for all three η values, with the mean error tending to decrease across the epochs. What differentiates the behavior of the different η values is the time to convergence and the oscillations.

As observed, the larger the η, the more epochs will be needed to converge to the final solution. This is because the changes made to the vector w will be larger compared to a smaller η.

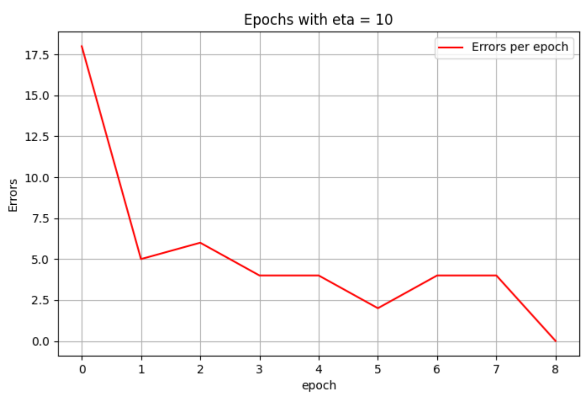
Another effect of the choice of η is reflected in the oscillations. The graph with $\eta = 0.1$ shows a smoother function with fewer oscillations. This is because the corrections made by the algorithm are more gradual, reducing abrupt variations in errors between one epoch and another.



(a) Plot of the number of errors vs. epochs with $\eta = 1$



(b) Plot of the number of errors vs. epochs with $\eta = 0.1$



(c) Plot of the number of errors vs. epochs with $\eta = 10$

Figure 3: Plot of the number of errors vs. epochs with different η's

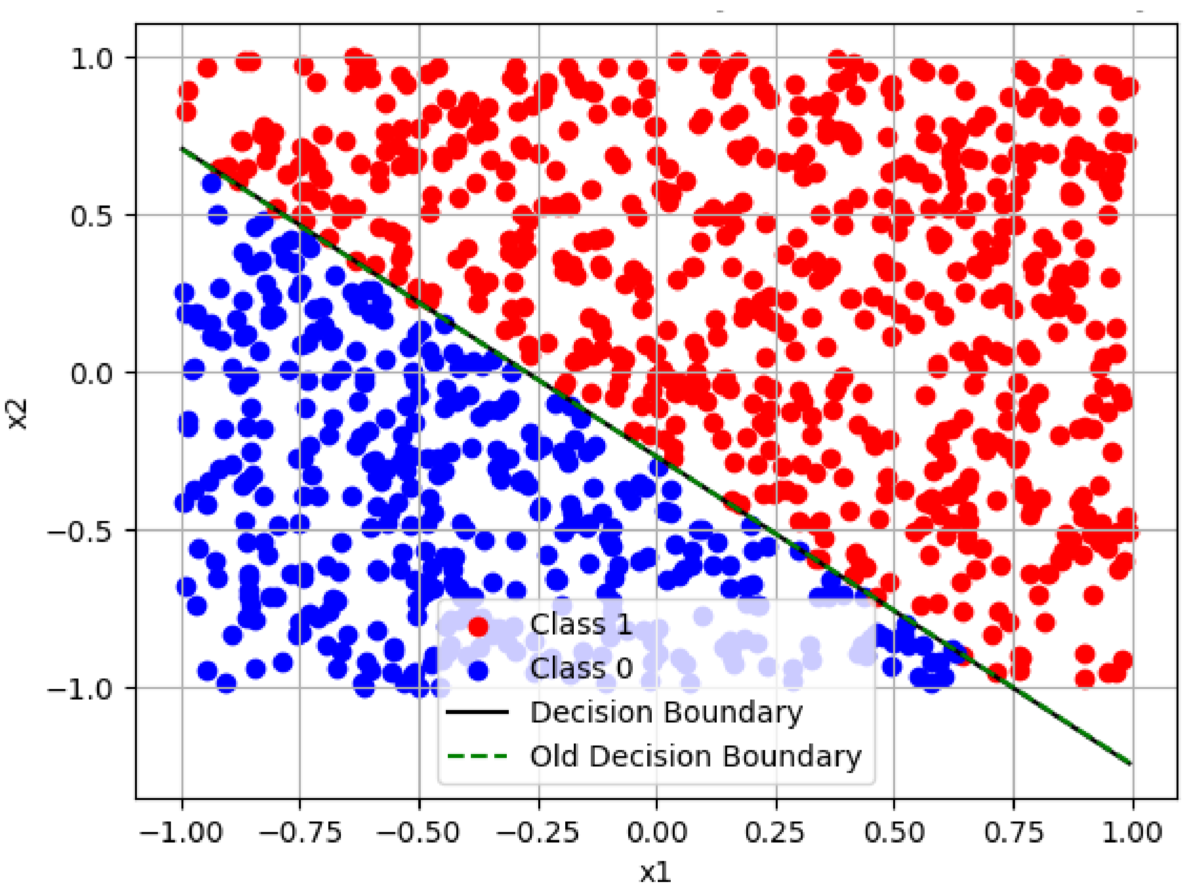
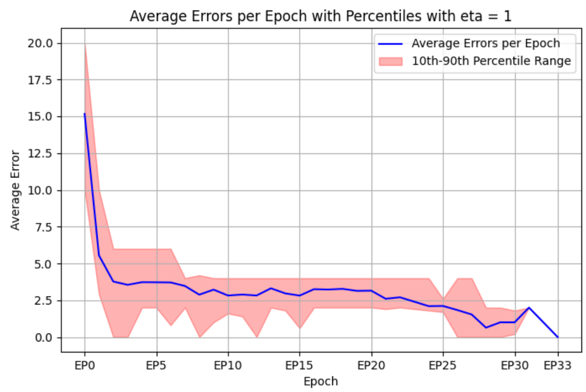
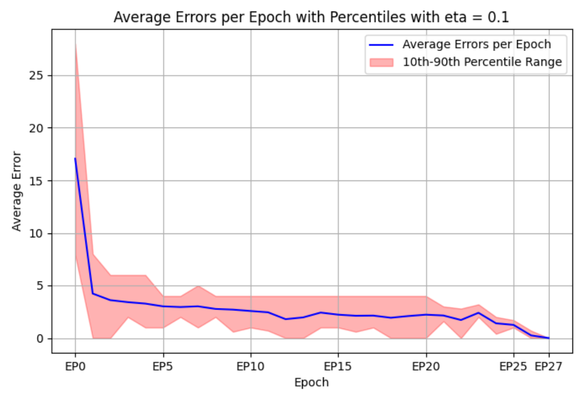


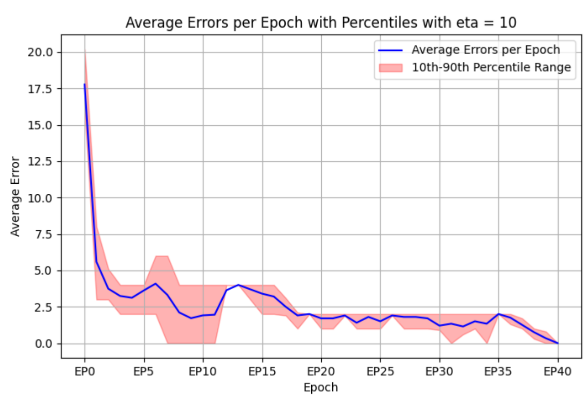
Figure 4: Comparison between the linear separator defined in Q1(a) in black and the boundary separator generated by the Perceptron Learning Algorithm with a larger data set, dashed in green.



(a) Plot of the number of errors vs. epochs with $\eta = 1$



(b) Plot of the number of errors vs. epochs with $\eta = 0.1$



(c) Plot of the number of errors vs. epochs with $\eta = 10$

Figure 5: Plot of the number of errors vs. epochs with different η's