PERCEPTRON MODEL AND SUPERVISED LEARNING, SOLUTION

MAYUR MUDIGONDA

1. Pattern discrimination

From the plot below, it is obvious that there exists no linear setting of the weights (in the given space) for which the two classes are linearly separable. Thus, no weight setting of the McCulloch-Pitt's neuron will discriminate the two classes without error. It would seem that if we applied a non-linear transformation to the data we could get to a more separable space. Note, that typically projecting data from a lower dimension (two in this case) to a higher dimension (five or higher, say) usually leads to much better separability. This is commonly known as the kernel trick in machine learning. Some student solutions had some interesting plots to explain this as well.

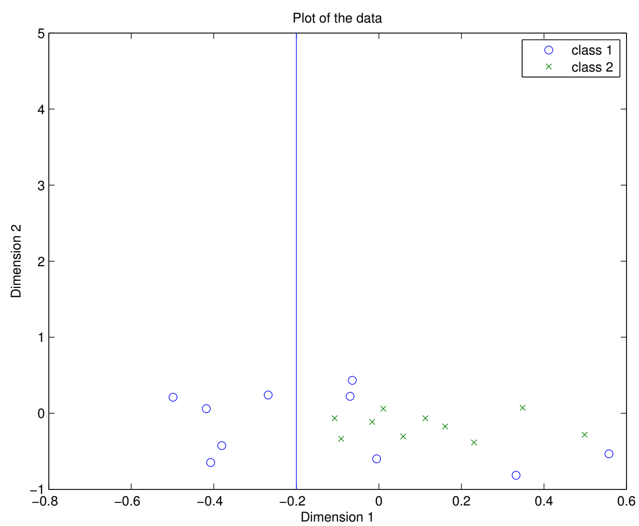

Figure 1. This is the plot of the data. Any number of solutions exist for the linear condition of the McCullogh-Pitt neuron. The above solution shows one of the many possible solutions a McCullogh Pitt neuron could arrive at. It is fine, if you got another solution as well.

There exist many different ways to solve this discrimination task. One obvious solution can be to apply some non-linearity to the input samples. For e.g., if we took each value of the the input and squared it

2. Linear neuron with sigmoidal output non-linearity

The objective function for a McCullogh-Pitt neuron (in general) is given by

$$E = \frac{1}{2}\sum_i [T_i - \sigma_i]^2 \tag{1}$$

where σ is defined as

$$\sigma(u) = \frac{1}{1+e^{-u}}$$

and $u = \sum_i w_i x_i$

Taking the partial derivative of $\sigma(u)$ with respect to w_i gives us the following equation using chain rule

$$\frac{\partial \sigma_u}{\partial w_i} = \frac{1}{1+e^{-u}}\frac{e^{-u}}{1+e^{-u}}x_i \tag{2}$$

The partial derivative of the Energy function with respect to $\sigma(u)$ is given by

$$\frac{\partial E}{\partial w_i} = -2\sum_{i=1}^{N}(y_i - \sigma(u))\frac{\partial \sigma_u}{\partial w_i} \tag{3}$$

Substituting the partial derivative of σ with respect to w we get the update rule as the follows:

$$\delta w_i = \eta \sum_i (T_i - \sigma_i(u))\sigma_i(1-\sigma_i)x_i \tag{4}$$

where η is the learning rate of the system.

3. Apples vs Oranges - Linear vs sigmoidal neurons

In this case, both solutions seemingly separate the two classes. You can observe that the error for the non-linear neuron is more erratic for the first few iterations. A few notes, this is obviously "a" solution. Modifying the η values can easily change the convergence time and error values. Further, if you modulate the value of η over time, you can often avoid minima in the error energy function.

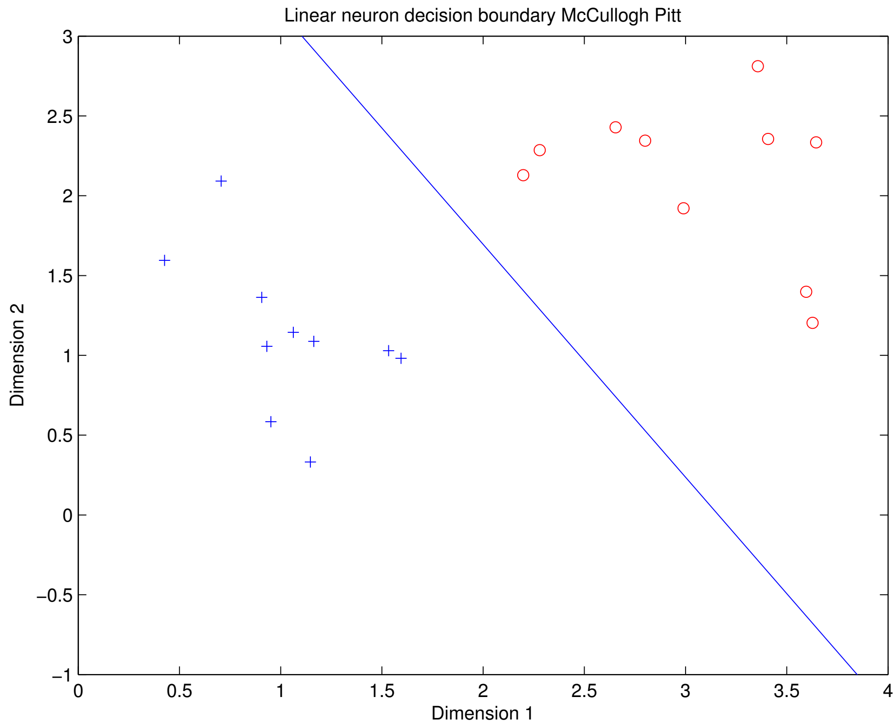

FIGURE 2. The solution of a linear McCullogh-Pitt neuron

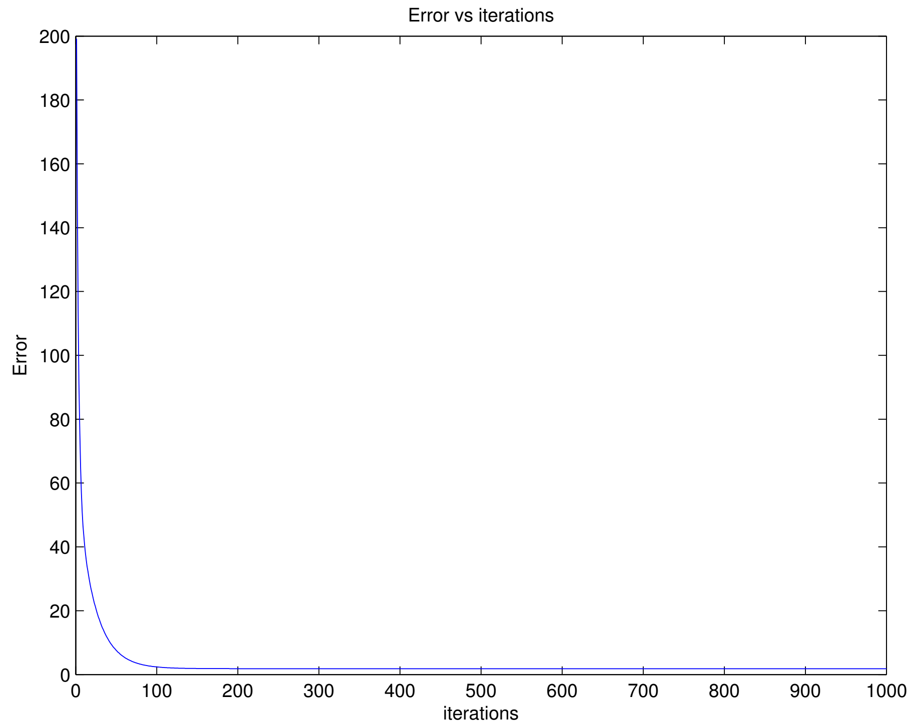

Figure 3. The plot of error with respect to iterations in the linear case

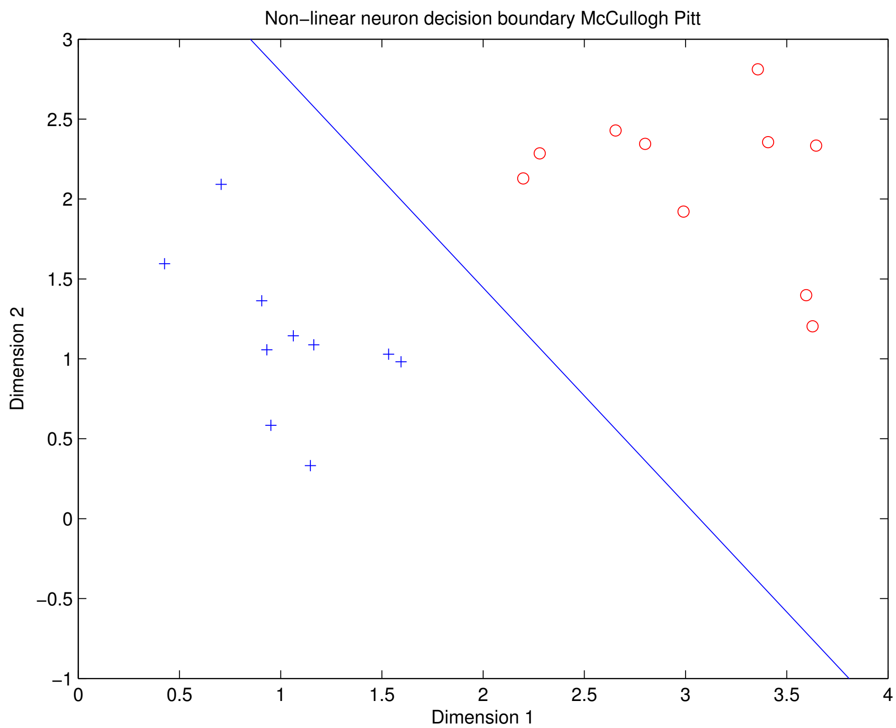

FIGURE 4. The solution of a non-linear McCullogh-Pitt neuron

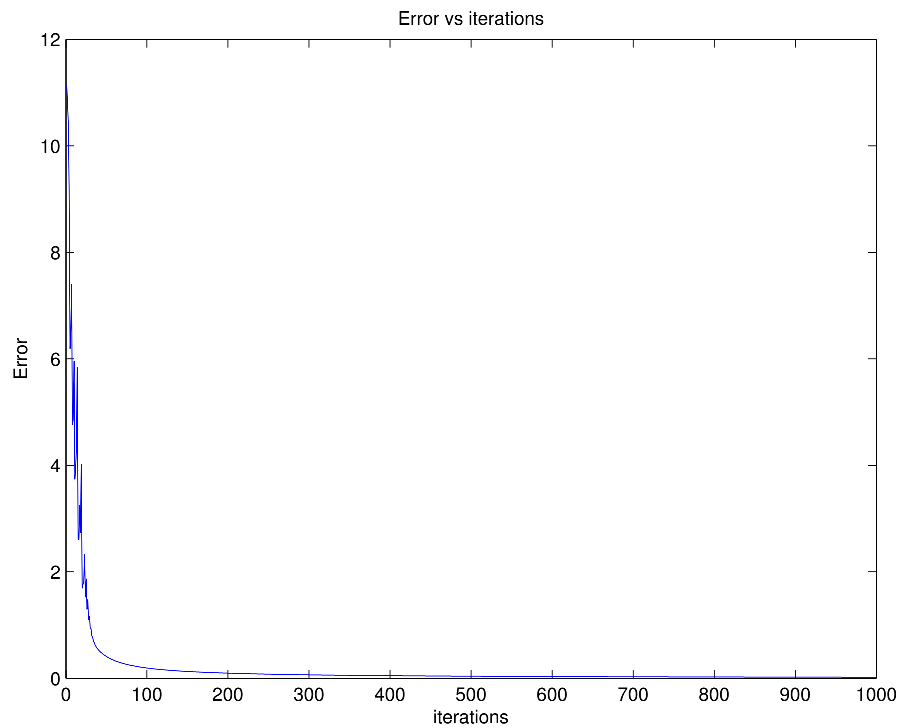

FIGURE 5. The plot of error with respect to iterations in the non-linear case