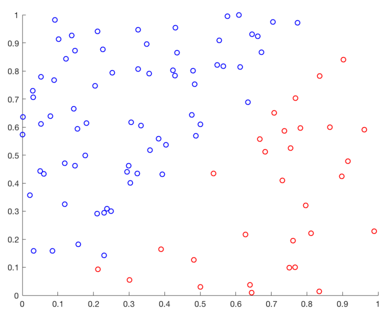
Seminar 3

1 Initialisation

We are given a matrix *data* made of three columns. The first column contains the x coordinates of the points, the second column contains the y coordinates of the points and the last column contains the target values. We want to find the classification line that best classifies the two kind of points. In blue we have the points that have a target value of -1 and in red, the points that have a target value of +1.



The decision boundary is defined by a vector $\mathbf{w} = \begin{pmatrix} w_0 \\ w_1 \\ w_2 \end{pmatrix}$. It can be at first a bit surprising to see that we define a 3D vector in a 2D plan, but there is a good reason for that. The decision boundary is defined by all the points $(x; y)$ that verify:

$$\begin{pmatrix} 1 \\ x \\ y \end{pmatrix} . \begin{pmatrix} w_0 \\ w_1 \\ w_2 \end{pmatrix} = 0$$

which is equivalent to:

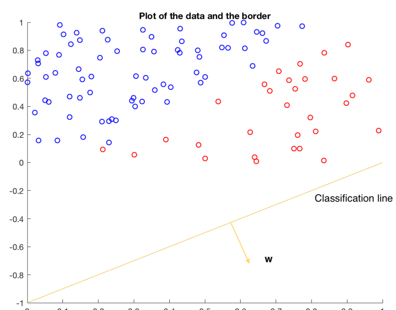
$$w_0 + w_1 x + w_2 y = 0.$$

This is a general formulation for a straight line. To convince yourself, consider the case when $w_2 \neq 0$, then the equation becomes

$$y = -\frac{(w_0 + w_1 x)}{w_2} = -\frac{w_1}{w_2}x - \frac{w_0}{w_2}$$

which has the form you are most familiar with ($y = ax+b$). However the general formulation also accounts for unusual cases, such as when $w_2 = 0$ in which case the decision boundary is a vertical line with x-intercept $x = -\frac{w_0}{w_1}$.

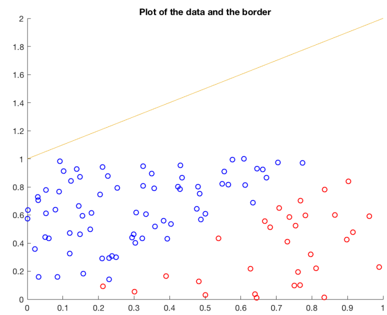
Now, if I set $\mathbf{w} = \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix}$, I obtain the following classification line:



Now the goal is to update correctly the weights $\begin{pmatrix} w_0 \\ w_1 \\ w_2 \end{pmatrix}$ so that the error is decreased. So we first need to define the error. Remember from last seminar that it is possible to tell the category assigned by the classifier thanks to the dot product. We know from the definition of the decision boundary that $\mathbf{w}$ is the vector that is perpendicular to the decision boundary. All the points $\mathbf{x} = \begin{pmatrix} 1 \\ x \\ y \end{pmatrix}$ that belong to the decision boundary verify $\mathbf{x}.\mathbf{w} = 0$. Therefore in the case of the preceding figure, we can tell what the classification is going to be by looking at the sign of the dot product. Since all points (red and blue) are on the side of the decision boundary opposite to the direction of $\mathbf{w}$, then:

$$\mathbf{p}.\mathbf{w} = \begin{pmatrix} 1 \\ p_1 \\ p_2 \end{pmatrix} . \begin{pmatrix} w_0 \\ w_1 \\ w_2 \end{pmatrix} < 0$$

In the next figure we put the classification line so that all points are on the same side of the decision boundary as $\mathbf{w}$ and therefore the dot product is always positive $\mathbf{p}.\mathbf{w} > 0$. You can observe that within the Matlab file. If you run the third subsection, all classifications are negative, while they are positive in the fourth subsection.



2 Implementation of the gradient descent

The goal is to move the decision boundary so that it reduces the error (the number of misclassified points). We have just seen that we know the classification of a point thanks to the sign of the dot product $\mathbf{p}.\mathbf{w}$. We want to have the blue points classified as -1 and the red ones as +1. Therefore we say that a point $\mathbf{p}$ is correctly classified if the sign of its dot product $\mathbf{p}.\mathbf{w}$ is of the sign of the target value (-1 for blue and +1 for red). Therefore a point is correctly classified if: $t - \text{sign}(\mathbf{p}.\mathbf{w}) = 0$, where t is the target value (-1 for blue and +1 for red) and `sign` is the function that gives 1 if the input is positive and -1 otherwise. Remember from last seminar that we update the weight if and only if a point is not correctly classified. We can therefore define the error of the point $\mathbf{p}$ of target t for the parameter $\mathbf{w}$ as:

$$E(\mathbf{p}, \mathbf{w}, t) = \begin{cases} |\mathbf{p}.\mathbf{w}| & \text{if } \text{sign}(\mathbf{p}.\mathbf{w}) - t \neq 0 \\ 0 & \text{if } \text{sign}(\mathbf{p}.\mathbf{w}) - t = 0 \end{cases}$$

We want the error to be always positive if the point is not correctly classified (absolute value $|.|$) and to be 0 if the point is correctly classified, i.e. when the sign of $\mathbf{p}.\mathbf{w}$ is the same as that of t. Please note that the absolute value of $\mathbf{p}.\mathbf{w}$ indicates how far from the classification line the $\mathbf{p}$ point is. We can separate the cases of the preceding equation as follows:

$$E(\mathbf{p}, \mathbf{w}, t) = \begin{cases} -\mathbf{p}.\mathbf{w} & \text{if } \text{sign}(\mathbf{p}.\mathbf{w}) - t \neq 0 \text{ and } \mathbf{p}.\mathbf{w} < 0 \\ \mathbf{p}.\mathbf{w} & \text{if } \text{sign}(\mathbf{p}.\mathbf{w}) - t \neq 0 \text{ and } \mathbf{p}.\mathbf{w} > 0 \\ 0 & \text{if } \text{sign}(\mathbf{p}.\mathbf{w}) - t = 0 \end{cases}$$

This can be neatly simplified by using the sign of t:

$$E(\mathbf{p}, \mathbf{w}, t) = \begin{cases} -t \times \mathbf{p}.\mathbf{w} & \text{if } \text{sign}(\mathbf{p}.\mathbf{w}) - t \neq 0 \\ 0 & \text{if } \text{sign}(\mathbf{p}.\mathbf{w}) - t = 0 \end{cases}$$

which is the Perceptron Criterion given in the lecture notes.

Therefore as we update the parameters $\mathbf{w}$, if there is an error according to:

$$\begin{pmatrix} w_0 \\ w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} w_0 \\ w_1 \\ w_2 \end{pmatrix} - \eta \begin{pmatrix} \frac{\partial E}{\partial w_0} \\ \frac{\partial E}{\partial w_1} \\ \frac{\partial E}{\partial w_2} \end{pmatrix}$$

From the last definition of the error, it is possible to compute the gradient of the error:

$$\begin{pmatrix} \frac{\partial E}{\partial w_0} \\ \frac{\partial E}{\partial w_1} \\ \frac{\partial E}{\partial w_2} \end{pmatrix} = \begin{cases} -t \times \begin{pmatrix} 1 \\ p_1 \\ p_2 \end{pmatrix} & \text{if } \mathrm{sign}(\mathbf{p}.\mathbf{w}) - t \neq 0 \\ 0 & \text{if } \mathrm{sign}(\mathbf{p}.\mathbf{w}) - t = 0 \end{cases}$$

such that the update of the parameters is:

$$\begin{pmatrix} w_0 \\ w_1 \\ w_2 \end{pmatrix} = \begin{cases} \begin{pmatrix} w_0 \\ w_1 \\ w_2 \end{pmatrix} + \eta t \times \begin{pmatrix} 1 \\ p_1 \\ p_2 \end{pmatrix} & \text{if } \mathrm{sign}(\mathbf{p}.\mathbf{w}) - t \neq 0 \\ \begin{pmatrix} w_0 \\ w_1 \\ w_2 \end{pmatrix} & \text{if } \mathrm{sign}(\mathbf{p}.\mathbf{w}) - t = 0 \end{cases}$$