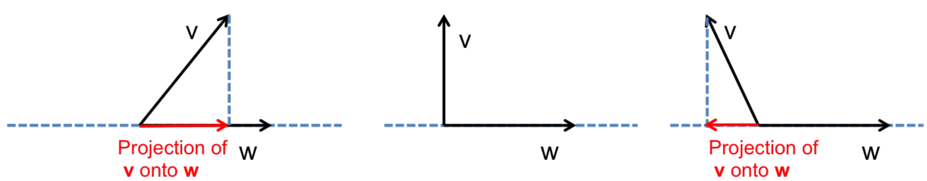
Seminar 2

1 Question 1

The sign of the dot product of two vectors $\mathbf{v}.\mathbf{w}$ can first be understood via the formula: $\mathbf{v}.\mathbf{w} = \|\mathbf{v}\|.\|\mathbf{w}\|.cos(\theta)$, where θ is the angle between the two vectors $\mathbf{v}$ and $\mathbf{w}$. This however is not the best method, because it is never good to solely rely on a formula. Instead we should be looking graphically at what the dot product is doing. We should understand the dot product as the projection of one vector onto the other. Here it is easier to look at the projection of $\mathbf{v}$ onto $\mathbf{w}$ rather than at the opposite because $\mathbf{w}$ is not changing. So draw in your head an imaginary line that is superimposed onto $\mathbf{w}$. Then project the vector $\mathbf{v}$ onto that imaginary line. If the two vectors, $\mathbf{w}$ and the projection of $\mathbf{v}$ onto the imaginary line, point toward the same direction then the dot product is positive. Otherwise the dot product $\mathbf{v}.\mathbf{w}$ is negative. If the two vectors $\mathbf{v}$ and $\mathbf{w}$ are perpendicular, then it is not possible to extract any direction because the projection is of length 0. Therefore on the left figure $\mathbf{v}.\mathbf{w} > 0$, on the middle one $\mathbf{v}.\mathbf{w} = 0$ and on the right figure $\mathbf{v}.\mathbf{w} < 0$.



2 Question 2

1) We consider $y(x, w) = w \times x$. The function y is made of two parameters: x and w. When we look at the partial derivative $\frac{\partial y}{\partial x}$, then we only differentiate the function according to x. During the differentiation and only during the differentiation we consider that w is a constant. On the contrary, if we look at $\frac{\partial y}{\partial w}$, then during the differentiation (and only during the differentiation) we consider that x is a constant. Therefore:

$$\frac{\partial y}{\partial x} = w \quad \text{and} \quad \frac{\partial y}{\partial w} = x$$

2) a) We now consider $y(x_1, x_2, x_3, w_1, w_2, w_3) = w_1 \times x_1 + w_2 \times x_2 + w_3 \times x_3$. Just as before, when we differentiate according to one parameter, we consider during the differentiation (and only during the differentiation) that all the other parameters are constant, therefore:

$$\frac{\partial y}{\partial x_1} = w_1 \qquad \frac{\partial y}{\partial w_2} = x_2 \quad \text{and} \quad \frac{\partial y}{\partial w_3} = x_3$$

$$\frac{\partial y}{\partial w_1} = x_1 \qquad \frac{\partial y}{\partial x_2} = w_2 \quad \text{and} \quad \frac{\partial y}{\partial x_3} = w_3$$

b) The gradient is the vector where each coordinate is a partial derivative. The gradient of $y(x_1, x_2, x_3, w_1, w_2, w_3)$ according to $\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$ is therefore:

$$\text{grad}(y)_{\mathbf{w}} = \begin{pmatrix} \frac{\partial y}{\partial w_1} \\ \frac{\partial y}{\partial w_2} \\ \frac{\partial y}{\partial w_3} \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \mathbf{x}$$

3) a) You should have realised that writing:

$$y(x_1, x_2, x_3, w_1, w_2, w_3) = w_1 \times x_1 + w_2 \times x_2 + w_3 \times x_3$$

is equivalent to writing:

$$y(x_1, x_2, x_3, w_1, w_2, w_3) = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} . \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \mathbf{w}.\mathbf{x}$$

Therefore to simplify the notations, we can write:

$$y(\mathbf{x}, \mathbf{w}) = \mathbf{w}.\mathbf{x}$$

b) Exactly as in 2) b), we can simplify the notations and write that the gradient of y according to $\mathbf{w}$ is:

$$\text{grad}(y)_{\mathbf{w}} = \text{grad}(\mathbf{x}.\mathbf{w})_{\mathbf{w}} = \mathbf{x}$$

and that the gradient of y according to $\mathbf{x}$ is:

$$\text{grad}(y)_{\mathbf{x}} = \text{grad}(\mathbf{x}.\mathbf{w})_{\mathbf{x}} = \mathbf{w}$$

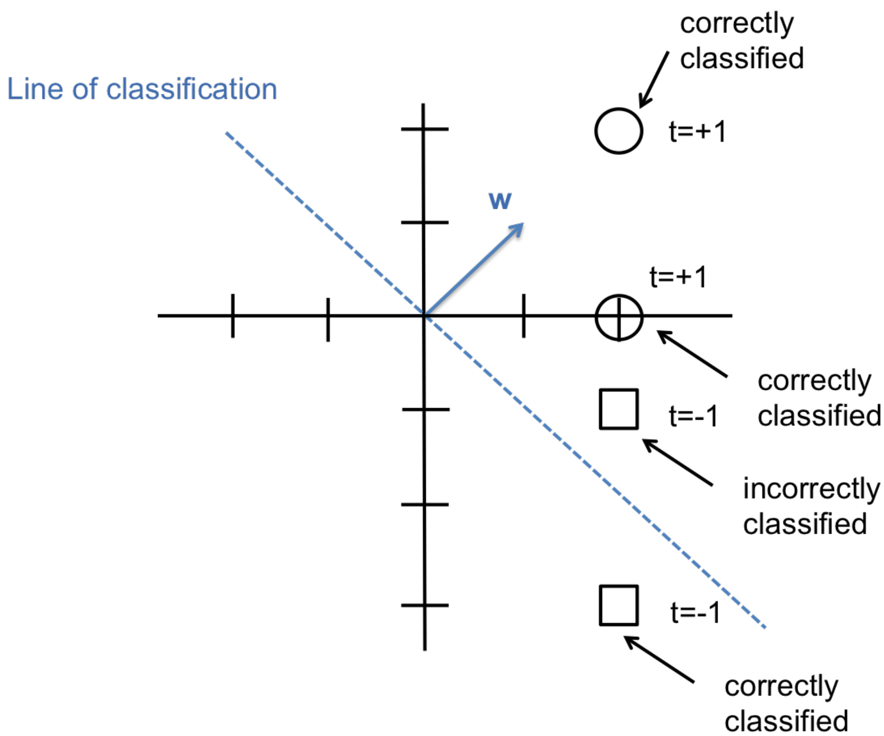
4) With $E(\mathbf{w}) = -(\mathbf{w}.\mathbf{x})t$, then exactly as in 3):

$$\text{grad}(E(\mathbf{w}))_{\mathbf{w}} = \text{grad}(-(\mathbf{w}.\mathbf{x})t)_{\mathbf{w}} = -\mathbf{x} \times t$$

3 Question 3

3.1 Computation of the error

In the following problem, we consider the problem of classifying circles ($t = +1$) and squares ($t = -1$) on the correct side of the classification line. Let's first



understand the error. It is defined as $E(\mathbf{w}) = -(\mathbf{w}.\mathbf{v})t$ where t is the target, $\mathbf{w}$ is the vector that is perpendicular to the classification line and $\mathbf{v}$ is the coordinate of each element we want to classify. For instance we have for the top circle: $\mathbf{v}_1 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$, for the bottom circle $\mathbf{v}_2 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$, for the top square $\mathbf{v}_3 = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$ and for the bottom square $\mathbf{v}_4 = \begin{pmatrix} 2 \\ -3 \end{pmatrix}$. We besides have $\mathbf{w} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. The error of the first square is :

$$\begin{aligned} E_3(\mathbf{w}) &= -(\mathbf{w}.\mathbf{v}_3)t_3 \\ &= -\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}.\begin{pmatrix} 2 \\ -1 \end{pmatrix}\right) \times (-1) \\ &= \begin{pmatrix} 1 \times 2 - 1 \times 1 \end{pmatrix} \\ &= 1 > 0 \end{aligned} \tag{1}$$

The error of the first square is therefore positive and we should also be happy because it is not correctly classified. Please note that in the perceptron, the error is only made of the terms that are incorrectly classified. Indeed if we were to compute the error of the first circle:

$$\begin{aligned} E_1(\mathbf{w}) &= -(\mathbf{w}.\mathbf{v}_1)t_1 \\ &= -\left(\begin{pmatrix}1\\1\end{pmatrix}.\begin{pmatrix}2\\2\end{pmatrix}\right) \times (1) \\ &= -\Big(1 \times 2 + 1 \times 2\Big) \\ &= -4 < 0 \end{aligned} \tag{2}$$

We would see that the error is negative and this does not make much sense. It was also possible to understand the sign of the error thanks to question 1). If you manage to understand the following you will earn a lot of time. We indeed immediately obtain from question 1) that:

$$\mathbf{w}.\mathbf{v}_1 > 0 \quad \mathbf{w}.\mathbf{v}_2 > 0 \quad \mathbf{w}.\mathbf{v}_3 > 0 \quad \text{and} \quad \mathbf{w}.\mathbf{v}_4 < 0$$

The sign of the error can be deducted now thanks to the formula of the error ($E(\mathbf{w}) = -(\mathbf{w}.\mathbf{v})t$) and we see that:

$$\begin{cases} \text{sign}(E_1) = \text{sign}\big(-(\mathbf{w}.\mathbf{v}_1) \times t_1\big) < 0 \\ \text{sign}(E_2) = \text{sign}\big(-(\mathbf{w}.\mathbf{v}_2) \times t_2\big) < 0 \\ \text{sign}(E_3) = \text{sign}\big(-(\mathbf{w}.\mathbf{v}_3) \times t_3\big) > 0 \\ \text{sign}(E_4) = \text{sign}\big(-(\mathbf{w}.\mathbf{v}_4) \times t_4\big) < 0 \end{cases}$$

3.2 Computation of the update rule

Now that we understand why we have such an error term, we need to understand how the update rule is working. The idea is to update the parameter $\mathbf{w} = \begin{pmatrix}w_1\\w_2\end{pmatrix}$ as a function of the gradient of the error. The gradient measures the slope of the error at the location $\mathbf{w}$. We want to go according to the direction that decreases the error and we therefore use the following formula:

$$\mathbf{w} = \mathbf{w} - \eta\frac{\partial E}{\partial w}$$

Which is equivalent to writing:

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

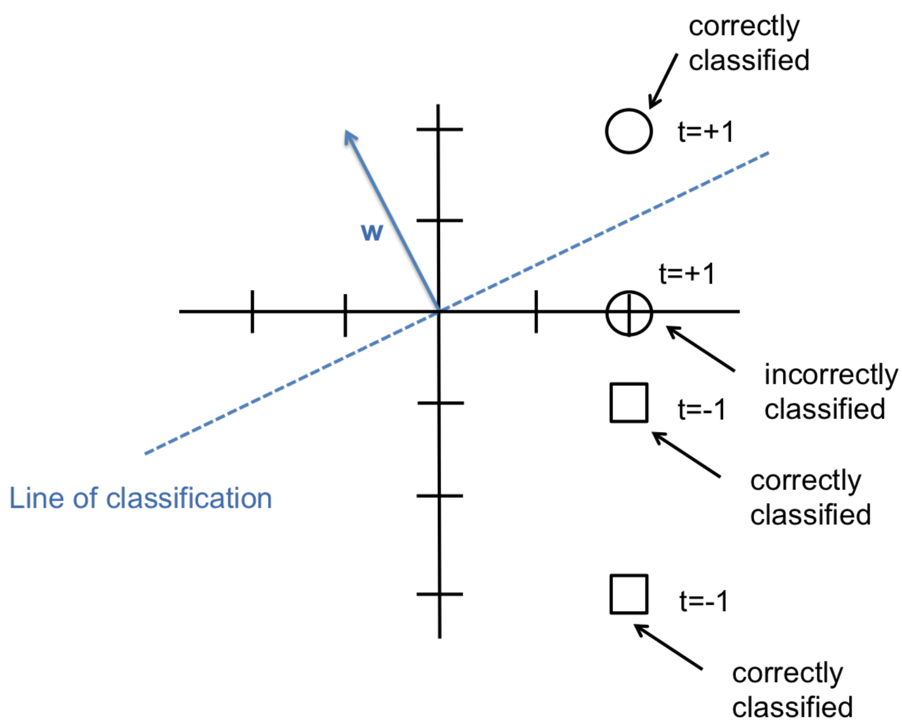
Where η is a parameter that should be chosen so that the update is not too fast. In the following we will update the gradient sequentially, that it to say that we will update the gradient according to each of the circle and the square iteratively. For instance, let's first update the error according to the first square. We choose $\eta = 1$ because this problem is linearly separable there is only one minimum and therefore the choice of learning rate is irrelevant. We first need to compute the partial derivative of the error:

$$\begin{pmatrix} \frac{\partial E_3}{\partial w_1} \\ \frac{\partial E_3}{\partial w_2} \end{pmatrix} = -\begin{pmatrix} v_3[1] \\ v_3[2] \end{pmatrix} t_3 = +\begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

We finally ask the computer to update $\mathbf{w}$ according to:

$$\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - 1 \times \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix} \tag{3}$$

We can now update the vector $\mathbf{w}$ and the classification line. We observe that the line has moved so that the first square is now well classified but it moved so much so that the second circle is now incorrectly classified.



The next step is now to update the parameter according to the incorrectly classified parameter, the second circle. We first need to compute the partial derivative of the error:

$$\begin{pmatrix} \frac{\partial E_2}{\partial w_1} \\ \frac{\partial E_2}{\partial w_2} \end{pmatrix} = -\begin{pmatrix} v_2[1] \\ v_2[2] \end{pmatrix} t_2 = \begin{pmatrix} -2 \\ 0 \end{pmatrix}$$

We finally ask the computer to update $\mathbf{w}$ according to:

$$\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix} - 1 \times \begin{pmatrix} -2 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \tag{4}$$

This time we observe that the second circle is now well classified but that the first square is now on the classification line. We shall now update according to the other parameters until we are happy with the final state. The process can continue...