Learning From Data
Exercise 1.3
Proof

Oleksii Symon

January 09, 2018

Task

Exercise 1.3

The weight update rule in (1.3) has the nice interpretation that it moves in the direction of classifying $\mathbf{x}(t)$ correctly.

(a) Show that $y(t)\mathbf{w}^{\mathsf{T}}(t)\mathbf{x}(t) < 0$. *[Hint: $\mathbf{x}(t)$ is misclassified by $\mathbf{w}(t)$.]*

(b) Show that $y(t)\mathbf{w}^{\mathsf{T}}(t+1)\mathbf{x}(t) > y(t)\mathbf{w}^{\mathsf{T}}(t)\mathbf{x}(t)$. *[Hint: Use (1.3).]*

(c) As far as classifying $\mathbf{x}(t)$ is concerned, argue that the move from $\mathbf{w}(t)$ to $\mathbf{w}(t+1)$ is a move 'in the right direction'.

Solution

Equation for perceptron:

$$h(x) = sign\left(\sum_{i=0}^{d} w_i x_i\right) = w^{\intercal}x$$

Equation for updating weights:

$$w(t+1) = w(t) + y(t)x(t)$$

0.1 (a)

In task **(a)** there are 2 possibilities:

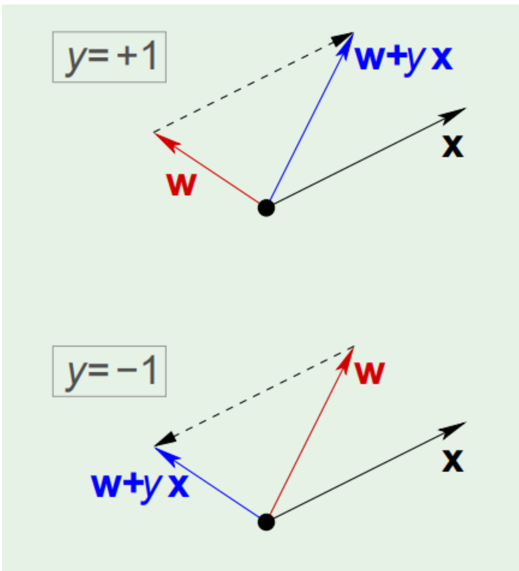
- $w(t)^\intercal x(t) = 1$, then $y(t) = -1$. So $1 * (-1) = -1 < 0$
- $w(t)^\intercal x(t) = -1$, then $y(t) = 1$. So $-1 * (1) = -1 < 0$

0.2 (b)

Since $w^\intercal(t)$ became $w^\intercal(t+1)$ we can say that $w(t)^\intercal x(t)$ classification was not correct. There are two possibilities:

- $w(t)^\intercal x(t) = 1$, then $y(t) = -1$. So $w(t+1) = w(t) - x(t)$. As result $w(t+1) < w(t)$ and as $y(t)w(t)^\intercal x(t) < 0$ we can say that $y(t)w(t+1)^\intercal x(t) > y(t)w(t)^\intercal x(t)$
- $w(t)^\intercal x(t) = -1$, then $y(t) = 1$. So $w(t+1) = w(t) + x(t)$. As result $w(t+1) > w(t)$ and as $y(t)w(t)^\intercal x(t) > 0$ we can say that $y(t)w(t+1)^\intercal x(t) > y(t)w(t)^\intercal x(t)$

0.3 (c)



At the first image dot product of $w \cdot x < 0$ and $y = 1$ so we increase w
At the second image dot product of $w \cdot x > 0$ and $y = -1$ so we decrease w