Homework 1

Karan Sarkar
sarkak2@rpi.edu

September 9, 2019

Exercise 1.3.

(a) Suppose $\mathbf{x}(t)$ is misclassified by $\mathbf{w}(t)$. This yields two cases. First, suppose that we have a false positive. This means that $h(\mathbf{x}(t)) = \text{sign}(\mathbf{w}^T(t)\mathbf{x}(t)) > 0$ while $y(t) < 0$. Thus, $\mathbf{w}^T(t)\mathbf{x}(t) > 0$. Thus, $y(t)\mathbf{w}^T(t)\mathbf{x}(t) < 0$. Second, suppose that we have a false negative. This means that $h(\mathbf{x}(t)) = \text{sign}(\mathbf{w}^T(t)\mathbf{x}(t)) < 0$ while $y(t) > 0$. Thus, $\mathbf{w}^T(t)\mathbf{x}(t) < 0$. Thus, $y(t)\mathbf{w}(t)^T\mathbf{x}(t) < 0$. The result thus holds in both cases.

(b) Consider the quantity $y(t)\mathbf{w}(t+1)^T\mathbf{x}(t)$.

$$\begin{aligned}
y(t)\mathbf{w}(t+1)^T\mathbf{x}(t) &= y(t)[\mathbf{w}(t) + y(t)\mathbf{x}(t)]^T\mathbf{x}(t) \\
&= y(t)\mathbf{w}(t)^T\mathbf{x}(t) + y(t)[y(t)\mathbf{x}(t)]^T\mathbf{x}(t) \\
&= y(t)\mathbf{w}(t)^T\mathbf{x}(t) + y(t)^2\mathbf{x}(t)^T\mathbf{x}(t) \\
&= y(t)\mathbf{w}(t)^T\mathbf{x}(t) + y(t)^2||\mathbf{x}(t)||^2 \\
&\geq y(t)\mathbf{w}(t)^T\mathbf{x}(t)
\end{aligned}$$

Note that $y(t)^2||\mathbf{x}(t)||^2 \geq 0$, thus it follows that $y(t)\mathbf{w}(t+1)^T\mathbf{x}(t) \geq y(t)\mathbf{w}(t)^T\mathbf{x}(t)$.

(c) The quantity $y\mathbf{w}^T\mathbf{x}$ indicates how misclassified a data point is. As we proved in part (a), it is negative when a point $\mathbf{x}$ is misclassified. Thus, increasing $y\mathbf{w}^T\mathbf{x}$ reduces the degree to which we missclasify $\mathbf{x}$. In part (b), we prove that the update rule increases $y\mathbf{w}^T\mathbf{x}$. Therefore, the update from $\mathbf{w}(t)$ to $\mathbf{w}(t+1)$ is a move in the right direction.

Exercise 1.5. I believe c,d,e fit the learning paradigm whereas a,b fit the design model.

Exercise 1.6.

(a) I think supervised learning could be used to predict which book someone will buy given their previous history.

(b) Reinforcement learning can be used to predict optimal tic-tac-toe moves.

(c) Unsupervised learning could be used to cluster the movies.

(d) Learning to play music could be supervised. Try to predict music based on previous experience.

(e) Credit limits is supervised learning. Use credit history to predict default probability.

Exercise 1.7.

(a) We find that g is the function that always returns '•'. 1 function agrees with g on all points. 3 functions agree on 2 points. 3 functions agree on 1 point. 1 function agrees on 0 points.

(b) We find that g is the function that always returns '$\circ$'. 1 function agrees with g on all points. 3 functions agree on 2 points. 3 functions agree on 1 point. 1 function agrees on 0 points.

(c) Let g be the XOR function. 1 function agrees with g on all points. 3 functions agree on 2 points. 3 functions agree on 1 point. 1 function agrees on 0 points

(d) 1 function agrees with g on all points. 3 functions agree on 2 points. 3 functions agree on 1 point. 1 function agrees on 0 points.

Problem 1.1. Let A be the event that the first ball is black. Let B be the event that the second ball is black. Thus, we are looking for the value $P[B|A]$. By Bayes Theorem, this equals $\frac{P[A \cap B]}{P[A]}$. Note that

$$P[A] = \frac{1}{2} \cdot 1 + \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{4}.$$

This is because there are two equally probable cases i.e. which bag you select. One case yields a probability of 1 of drawing a black ball. The other yields $\frac{1}{2}$. Note that $P[A \cap B] = \frac{1}{2}$ as this only happens when you select the bag with two black balls. Therefore, we see that

$$P[B|A] = \frac{P[A \cap B]}{P[A]} = \frac{\frac{1}{2}}{\frac{3}{4}} = \frac{1}{2} \cdot \frac{4}{3} = \frac{2}{3}$$

There is a $\frac{2}{3}$ chance the second ball is black given that the first is black.

Problem 1.2.

(a) Note that $h(\mathbf{x}) = +1$ occurs when $\mathbf{w}^T\mathbf{x} > 0$. Similarly, $h(\mathbf{x}) = -1$ occurs when $\mathbf{w}^T\mathbf{x} < 0$. Thus the separation occurs when $\mathbf{w}^T\mathbf{x} = 0$ i.e. when $w_0 + w_1 x_1 + w_2 x_2 = 0$. This is the equation of a line. The slope is $a = -\frac{w_1}{w_2}$ and the intercept is $b = -\frac{w_0}{w_2}$.

(b) The two diagrams are identical.