# Second Chapter Exercises Assignment 3

**Zahra Oruji[1]**

[1]Department of Mathematics, Shahid Beheshti University

1 - Given $S = ((xi; yi))_{i=1}^{m}$, define the multivariate polynomial

$$PS(x) = -\prod_{i\in[m]:y_i=1} \|x - x_i\|^2$$

Then, for every i s.t. $y_i = 1$ we have $PS(x_i) = 0$, while for every other $x$ we have $PS(x) < 0$.

2 - By the linearity of expectation,

$$\begin{aligned}
\mathbb{E}_{S|_{x\sim D^m}}[L_S(h)] &= \mathbb{E}_{S|_{x\sim D^m}}\left[\frac{1}{m}\sum_{i=1}^{m}\mathbb{1}_{[h(x_i)\neq f(x_i)]}\right] \\
&= \frac{1}{m}\sum_{i=1}^{m}\mathbb{E}_{S|_{x_i\sim D}}\left[\mathbb{1}_{[h(x_i)\neq f(x_i)]}\right] \\
&= \frac{1}{m}\sum_{i=1}^{m}\mathbb{P}_{S|_{x_i\sim D}}\left[h(x_i)\neq f(x_i)\right] \\
&= \frac{1}{m}\cdot m\cdot L_{(D,f)}(h) \\
&= L_{(D,f)}(h)
\end{aligned}$$

3 - (a) First, observe that by definition, $A$ labels positively all the positive instances in the training set. Second, as we assume realizability, and since the tightest rectangle enclosing all positive examples is returned, all the negative instances are labeled correctly by $A$ as well. We conclude that $A$ is an ERM.

(b) Fix some distribution $D$ over $\chi$, and define $R^\star$ as in the hint. Let $f$ be the hypothesis associated with $R^\star$ a training set $S$, denote by $R(S)$ the rectangle returned by the proposed algorithm and by $A(S)$ the corresponding hypothesis. The definition of the algorithm $A$ implies that $R(S) \subseteq R^\star$ for every $S$. Thus,

$$L_{(D,f)}(R(S)) = D(R^\star \setminus R(S))$$

Fix some $\epsilon \in (0, 1)$. Define $R_1$; $R_2$; $R_3$ and $R_4$ as in the hint. For each $i \in [4]$, define the event

$$F_i = S|_x : S|_x \cap R_i = \emptyset$$

Applying the union bound, we obtain

$$D^m(S : L_{(D,f)}(A(S)) > \epsilon) \leqslant D^m\left(\bigcup_{i=1}^{4} F_I\right) \leqslant \sum_{i=1}^{4} D^m(F_i)$$

Thus, it suffices to ensure that $D^m(F_i) \leqslant \delta/4$ for every $i$. Fix some $i \in [4]$. Then, the probability that a sample is in $F_i$ is the probability that all of the instances don't fall in $R_i$, which is exactly $(1-\epsilon/4)^m$. Therefore,

$$D^m(F_i) = (1-\epsilon/4)^m \leqslant \exp(-mc/4)$$

and hence,

$$D^m(S : L_{(D,f)}(A(S)) > \epsilon) \leqslant 4\exp(-mc/4)$$

Plugging in the assumption on m, we conclude our proof.

(c) The hypothesis class of axis aligned rectangles in $\mathbb{R}^d$ is defined as follows. Given real numbers $a_1 \leqslant b_1$, $a_2 \leqslant b_2$,...,$a_d \leqslant b_d$, define the classifier $h_{(a_1,b_1,\ldots,a_d,b_d)}$ by

$$h_{(a_1,b_1,\ldots,a_d,b_d)}(x_1,\ldots,x_d) = \begin{cases} 1 & \forall i \in [d], a_i \leqslant x_i \leqslant b_i \\ 0 & otherwise. \end{cases}$$

The class of all axis-aligned rectangles in $\mathbb{R}^d$ is defined as

$$H^d_{rec} = h_{(a_1,b_1,\ldots,a_d,b_d)} : \forall i \in [d], a_i \leqslant b_i$$

It can be seen that the same algorithm proposed above is an ERM for this case as well. The sample complexity is analyzed similarly. The only difference is that instead of 4 strips, we have $2d$ strips (2 strips for each dimension). Thus, it suffices to draw a training set of size $\left\lceil \frac{2d\log(2d/\delta)}{c} \right\rceil$.

(d) For each dimension, the algorithm has to find the minimal and the maximal values among the positive instances in the training sequence. Therefore, its runtime is $O(md)$. Since we have shown that the required value of m is at most $\left\lceil \frac{2d\log(2d/\delta)}{c} \right\rceil$, it follows that the runtime of the algoritshm is indeed polynomial in $d$,$1/\epsilon$, and $\log(1/\epsilon)$.