

# HOMEWORK 7

>>Sean Sun<<  
>>9078202463<<

**Instructions:** Although this is a programming homework, you only need to hand in a pdf answer file. There is no need to submit the latex source or any code. You can choose any programming language, as long as you implement the algorithm from scratch.

Use this latex file as a template to develop your homework. Submit your homework on time as a single pdf file to Canvas. Please check Piazza for updates about the homework.

## 1 VC dimension (30 pts)

Let the input  $x \in X = \mathbb{R}$ . Consider  $F = \{f(x) = \text{sgn}(ax^2 + bx + c) : a, b, c \in \mathbb{R}\}$ , where  $\text{sgn}(z) = 1$  if  $z \geq 0$ , and 0 otherwise. What is  $VC(F)$ ? Prove it.

$VC(F) = 3$

Proof:

When there are 4 points,  $F$  can not satisfy the binary vectore of  $[0, 1, 0, 1]$  or  $[1, 0, 1, 0]$ . So,  $F$  can't shatter a set with 4 points.

When there are more than 4 points, for every contiguous four elements in the binary vector,  $F$  still can't satisfy the binary vectore of  $[0, 1, 0, 1]$  or  $[1, 0, 1, 0]$ . So,  $F$  can't shatter a set with more than 4 points.

So,  $VC(F) = 3$

## 2 Verify PAC Bound (30 pts)

Let  $h$  be the VC dimension of function family  $F$ . For any  $\delta > 0$ , with probability at least  $1 - \delta$  we have

$$R(\hat{f}_S) - \hat{R}_S(\hat{f}_S) \leq 2\sqrt{2\frac{h \log n + h \log \frac{2e}{\delta} + \log \frac{2}{\delta}}{n}},$$

where  $R(f) = \mathbb{E}1_{[f(x) \neq y]}$  is the risk of  $f$ ,  $S = (x_1, y_1), \dots, (x_n, y_n)$  is a training set of size  $n$ ,  $\hat{R}_S(f) = \frac{1}{n} \sum_{i=1}^n 1_{[f(x_i) \neq y_i]}$  is the empirical risk of  $f$  on  $S$ , and  $\hat{f}_S \in \arg \min_{f \in F} \hat{R}_S(f)$  is an empirical risk minimizer (ERM).

We now verify this bound on a simple classification task. Let  $p(x) = \text{uniform}([-1, 1])$ . Given  $x$ , the label is deterministic:  $y = \text{sgn}(x)$ . Recall  $\text{sgn}(x) = 1$  if  $x \geq 0$ , and 0 otherwise. This means the true decision boundary is at  $x = 0$ . Let  $f_\theta(x) := \text{sgn}(x - \theta)$  which has threshold at  $\theta$ . Let  $F = \{f_\theta : \theta \in [-1, 1]\}$ .

1. Given  $S$ , find the smallest positive item:  $a = \min_{(x_i, y_i) \in S: y_i = 1} x_i$  if one exists, otherwise let  $a = 1$ . Is

$\hat{f}_S := f_a$  an ERM? Justify your answer.

$\hat{f}_S := f_a$  is an ERM.

When there are positive  $x_i$ s, every  $x_j > \min_{(x_i, y_i) \in S: y_i = 1} x_i$  where  $((x_j, y_j) \in S)$  will be labeled as 1 and all other points will be labeled as 0. So, setting the threshold at  $\min_{(x_i, y_i) \in S: y_i = 1} x_i$  will lead to zero empirical risk.

When there are no positive  $x_i$ , all points will be labeled as 0, so setting the threshold at 1 will label all points to be 0 and thus leads to zero empirical risk.

So,  $\hat{f}_S := f_a$  is an ERM.

2. Derive  $R(f_a)$ .

$$R(f_a) = 1 \times p(a > x \geq 0) = \frac{a}{2}$$

3. Derive  $\hat{R}_S(f_a)$ .

$$\hat{R}_S(f_a) = 0$$

4. What is the VC dimension of  $F$ ? Note  $F$  contains threshold classifiers of the type “left negative, right positive.”

$$VC(F) = 1$$

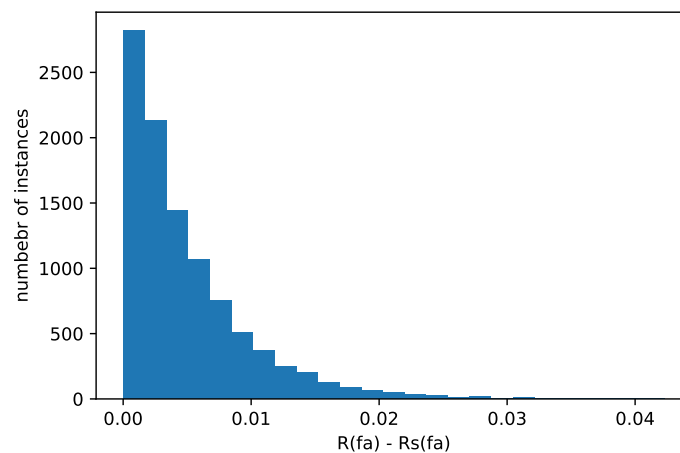
this one-sided threshold classifier can't satisfy the binary vector  $[1,0]$  because all points on the left of threshold will be labeled as 0 and all points on the right of threshold will be labeled as 1

5. Fix  $\delta = 0.05$  (95% confidence) and  $n = 200$ . Compute  $2\sqrt{2\frac{h \log n + h \log \frac{2\epsilon}{h} + \log \frac{2}{\delta}}{n}}$ . This is natural log.

$$2\sqrt{2\frac{h \log n + h \log \frac{2\epsilon}{h} + \log \frac{2}{\delta}}{n}} = 2\sqrt{2\frac{1 \log 200 + 1 \log \frac{2\epsilon}{1} + \log \frac{2}{0.05}}{200}} = 2\sqrt{\frac{\log 200 + \log 2 + 1 + \log 40}{100}} = 2\sqrt{\frac{\log 16000 + 1}{100}} = 0.654$$

6. Generate 10,000 random training sets from  $p(x)$  and the associated labels. Each training set  $S$  has  $n = 200$  points. On each  $S$  you will compute  $R(f_a) - \hat{R}_S(f_a)$ . Now you have 10,000 numbers. (1) Produce a histogram of them. (2) Find the 95% quantile of them. (3) Compare the 95% quantile to the bound in the previous question. Discuss your observations.

(1)



(2)

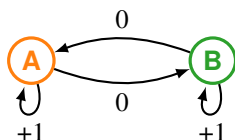
The 95% quantile is 0.01475

(3)

The bound with 95% confidence from question 5 was 0.654, which is a much broader bound than 0.01475. So the bound we got from question 5 is acceptable.

### 3 Q-learning (40 pts)

Consider the following Markov Decision Process. It has two states  $s$ . It has two actions  $a$ : move and stay. The state transition is deterministic: “move” moves to the other state, while “stay” stays at the current state. The reward  $r$  is 0 for move, 1 for stay. There is a discounting factor  $\gamma = 0.9$ .



The reinforcement learning agent performs Q-learning. Recall the  $Q$  table has entries  $Q(s, a)$ . The  $Q$  table is initialized with all zeros. The agent starts in state  $s_1 = A$ . In any state  $s_t$ , the agent chooses the action  $a_t$  according to a behavior policy  $a_t = \pi_B(s_t)$ . Upon experiencing the next state and reward  $s_{t+1}, r_t$  the update is:

$$Q(s_t, a_t) \leftarrow (1 - \alpha)Q(s_t, a_t) + \alpha \left( r_t + \gamma \max_{a'} Q(s_{t+1}, a') \right).$$

Let the step size parameter  $\alpha = 0.5$ .

1. Run Q-learning for 200 steps with a uniformly random behavior policy:  $\pi_B(s_t) = \text{move or stay with } 1/2$  probability for any  $s_t$ . Show the Q table at the end.

	stay	move
A	9.48078	7.99048
B	9.00271	8.29473

2. Reset and repeat the above, but with an  $\epsilon$ -greedy behavior policy: at each state  $s_t$ , with probability  $1 - \epsilon$  choose what the current Q table says is the best action:  $\arg \max_a Q(s_t, a)$ ; Break ties arbitrarily. Otherwise (with probability  $\epsilon$ ) uniformly chooses between move and stay. Use  $\epsilon = 0.5$ .

	stay	move
A	9.82054	8.69072
B	9.73766	8.79053

3. Reset and repeat the above, but with a deterministic greedy behavior policy: at each state  $s_t$  use the best action  $a_t \in \arg \max_a Q(s_t, a)$  indicated by the current Q table. If there is a tie, prefer move.

	stay	move
A	0	0
B	0	0

4. Without doing simulation, use Bellman equation to derive the true Q table induced by the MDP.  $V^*(s) = E[r(s, \pi^*(s)) + \gamma E_{s'|s, \pi^*(s)}[V^*(s')]] = 1 + 0.9 \times V^*(s)$   
 So,  $V^*(s) = 10$   $Q(s, move) = E[r(s, move)] + \gamma E_{s'|s, move}[V^*(s')] = 0 + 0.9 \times 10 = 9$   
 $Q(s, stay) = E[r(s, stay)] + \gamma E_{s'|s, stay}[V^*(s')] = 1 + 0.9 \times 10 = 10$

	stay	move
A	10	9
B	10	9