Introduction to Machine Learning
Fall 2018

Lecturer: Jie Wang — Homework 1
Posted: Sep. 10, 2018 — Due: Sep. 17, 2018
Name: ziqing li — ID: PB15061290

Notice, to get the full credits, please present your solutions step by step.

Exercise 1: 2pts

Show that $(1-\epsilon)^m \leq e^{-m\epsilon}$, where $m \in \mathbb{N}$.

Solution: $m\ln(1-\epsilon) \leq -m\epsilon$
$\ln(1-\epsilon)+\epsilon \leq 0$
Now we need prove of $1-\epsilon \leq e^{-\epsilon}$
so we have a function $f(x) = 1-x-e^{-x}$
$f'(x) = -1+e^{-x}$
because $0 \leq \epsilon \leq 1$
so $f'(x) < 0$
when$x = 0, f(x) = 0$,so wo have $f(x) < f(0) = 0$.
so $1-x \leq e^{-x}$
get

■

Exercise 2: Markov inequality 2pts

Let X be a nonnegative randome variable on $\mathbb{R}$. Then, for all $t > 0$, show that

$$\mathbf{P}(X \geq t) \leq \frac{\mathbf{E}[X]}{t}.$$

Solution: for any event E,we let I_E that $I_E = 1$ if E occurs and $I_E = 0$ if the other
so we have $I_{(E\geq a)} = 1$ if $E \geq$ a occurs,$I_{(E\geq a)} = 0$ if $E < a$
so we have $aI_{(X\geq a)} \leq X$
then make E for all side,we have:
$\mathbf{E}(aI_{(X\geq a)}) \leq \mathbf{E}(X)$
the left equals: $a\mathbf{P}(X \geq a)$
so wo have it

■

Exercise 3: VC-dimension 2pts

Assume that the instance space $X = \mathbb{R}^2$ and the hypothesis space H be the set of all linear threshold functions defined on $\mathbb{R}^2$. Find $VC(H)$ and prove it.

Solution: when d=2,N=3 and the instance are not in one line,we can shatter it ,so the VC-dimension is 3
now we prove it for space$X = \mathbb{R}^{(N-1)}$,the VC-dimension is N(not in one line)
first we have $y = xw$, for x,we add a bias Column vector,so it a N^2 matrix,then y is a label matric for $1 * N$,
because we can make w from $y = xw$,so it can be classification,so the VC-dimension is N
then for d=2 the VC-dimemsion is 3.

■

Exercise 4: Learning intervals 4pts

Let the target concept class be $C = \{[a, b] : a < b, a, b \in \mathbb{R}\}$ and the hypotheses class $H = C$, and the version space be $VS_{H,D}$. Each $c \in C$ labels the points inside the interval positive and the others negative. A consistent learner will pick a consistent hypothesis—if any—$h \in H$ according to a set of i.i.d. samples $\{(x_1, c(x_1)), (x_2, c(x_2), \ldots, (x_m, c(x_m)\}$ that obey an unknown distribution $\mathcal{D}$. Please find

$$\mathbf{P}[\exists\, h \in VS_{H,D} \text{ and } error_{\mathcal{D}}(h) > \epsilon],$$

and the corresponding sample complexity.

Solution: we have PAC learnable,but now we need prove

$$\mathbf{P}[\exists\, h \in VS_{H,D} \text{ and } error_{\mathcal{D}}(h) > \epsilon]$$

that is proving the C is not PAC learnable,that is Target c is not in hypothesis H,the same as

$$prove \forall h \in H, error_{\mathcal{D}}(h) > \epsilon$$

but now we have a example $(x_3, +), (x_4, -), (x_5, +), x_3 < x_4 < x_5$
we can not find a $h \in H$ make the example seperate. prove it. ■