

Introduction to Machine Learning

Fall 2018

Lecturer: Jie Wang
Posted: Sep. 10, 2018
Name: Shilong Zhang

Homework 1
Due: Sep. 17, 2018
ID: PB14214061

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: There may be another condition lost, the $\epsilon \in [-\infty, 1]$, otherwise if $\epsilon = 2$ and $m = 2$, the inequality is broken.

And with my condition, the question will be easy, we all know that

$$e^x \leq (1 - x), \forall x \in [-\infty, 1]$$

This is easy to prove. We can make the function

$$f(x) = e^{-x} - 1 + x$$

$f(x) = 0$ when $x = 0$, and $f'(x) = 1 - e^{-x}$, and we can know

$$f'(x) \begin{cases} \leq 0 & \text{if } x \leq 0 \\ > 0 & \text{if } x > 0 \end{cases} \quad (1)$$

So

$$0 \leq f(x)_{\min} \text{ when } x \in [-\infty, 1] \quad (1)$$

and for $m = 0$, $m = 1$, the inequality is ok, and we can assume that the inequality is ok for all $m \leq N$, and for $m = N + 1$,

$$(1 - \epsilon)^{N+1} = (1 - \epsilon)^N * (1 - \epsilon) \leq e^{-(N\epsilon)} * (1 - \epsilon)$$

then use (1) we have proved,

$$e^{-(N\epsilon)} * (1 - \epsilon) \leq e^{-(N\epsilon)} * e^{-\epsilon} = e^{-(N+1)\epsilon}$$

The inequality is ok for $N+1$, so we prove it!

■

Exercise 2: Markov inequality 2pts

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

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

Solution: For nonnegative random variable, we assume $f(x)$ is the probability distribution function of the x ; for any $t \geq 0$

$$E(X) = \int_0^\infty xf(x)dx \geq \int_t^\infty xf(x)dx \geq t \int_t^\infty f(x)dx = tP(x \geq t)$$

This exactly the what we need. ■

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: Put your solutions here. ■

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 $\{(x_1, c(x_1)), (x_2, c(x_2)), \dots, (x_m, c(x_m))\}$ 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: Put your solutions here. ■