

1. Two numbers are selected independently at random in the interval $[0, 1]$. You are told that the smaller one is less than $\frac{1}{3}$. The probability that the larger one is greater than $\frac{3}{4}$ is ____.

(2 marks)

2. Consider the Rosenbrock function on \mathbb{R}^2 defined as

$$f(x, y) = 100(y - x^2)^2 + (1 - x)^2.$$

The Hessian (second derivative) of the function at the point $(1, 1)$ is ____.

(2 marks)

3. For a hypothesis set \mathcal{H} , let $d_{VC} = N$. Find all the correct statements from below:

- (a) For any set of N points, we can get all dichotomies
- (b) There is a set of N points for which we can get all dichotomies
- (c) There are no set of N points for which we can get all dichotomies
- (d) For any set of $N + 1$ points, we can get all dichotomies
- (e) There is a set of $N + 1$ points for which we can get all dichotomies
- (f) There are no set of $N + 1$ points for which we can get all dichotomies
- (g) The growth function $m_{\mathcal{H}}(N + 1) = 2^{N+1}$
- (h) The growth function $m_{\mathcal{H}}(N + 1) < 2^{N+1}$

(2 marks)

4. Consider the “ k positive intervals” hypothesis set \mathcal{H}_k , where $h : \mathbb{R} \rightarrow \{-1, +1\}$ and $h(x) = +1$ if the point falls inside any of k arbitrarily chosen intervals and -1 otherwise.

- (a) For $k = 1$, the growth function $m_{\mathcal{H}_1}(N)$ is ____.
- (b) For $k = 2$, the growth function $m_{\mathcal{H}_2}(N)$ is ____.
- (c) The VC dimension of \mathcal{H}_k is ____.

(2+2+3=7 marks)

5. Consider the following data set of the classification problem with feature space \mathbb{R}^2 .

Example	x	y	Label
1	3	4	Red
2	2	2	Red
3	4	4	Red
4	1	4	Red
5	2	1	Blue
6	4	3	Blue
7	4	1	Blue

- (a) The equation for the maximum margin classifier is ____.
- (b) The margin of the maximum margin classifier is ____.
- (c) The support vectors are ____.

(d) The dual optimization problem is _____.

(2+1+1+2=6 marks)

6. Present a brief argument that the polynomial kernel defined as $K(x, y) = (1 + x^T y)^2$ is indeed a valid kernel. (3 marks)

7. Consider Boolean functions taking input/output values as -1 and $+1$ (in place of the usual 0 and 1).

(a) The "smallest" MLP that implements the function $f(x_1, x_2) = x_1 \cdot x_2$ is _____.

(b) The "smallest" MLP that implements the function $g(x_1, x_2) = x_1 \oplus x_2$ is _____.

(2+2=4 marks)

8. Consider the following data set comprising of three binary input attributes (A, B, C).

Example	A	B	C	Label
1	1	1	1	1
2	1	0	0	1
3	1	1	0	0
4	0	0	1	0

(a) The decision tree obtained by greedily growing till zero training error, while maximizing the information gain at each step is _____.

(b) The decision tree of smallest depth that has zero training error is _____.

(2+2=4 marks)