

BÀI TẬP 01

TRƯỜNG MINH ÁNH - 1112010

Contents

Câu 1: What types of learning, if any, best describe the following three scenarios:	2
Câu 2: Which of the following problems are best suited for the learning approach?	2
Câu 3: We have 2 opaque bags, each containing 2 balls. One bag has 2 black balls and the other has a black ball and a white ball. You pick a bag at random and then pick one of the balls in that bag at random. When you look at the ball, it is black. You now pick the second ball from that same bag. What is the probability that this ball is also black?	3
Câu 4: We draw only one such sample. Compute the probability that $\vartheta = 0$. The closest answer is (closest is the answer that makes the expression ϑ closest to 0):	3
Câu 5: We draw 1,000 independent samples. Compute the probability that (at least) one of the samples has $\vartheta = 0$. The closest answer is:	4
Câu 6: Which hypothesis g agrees the most with the possible target functions in terms of the above score?	4
Câu 7: Take $N = 10$. Run the Perceptron Learning Algorithm to find g and measure the difference between f and g as $Pr(f(x) \neq g(x))$ (you can either calculate this exactly, or approximate it by generating a sufficiently large separate set of points to evaluate it). Repeat the experiment for 1000 runs (as specified above) and take the average. Start the PLA with the weight vector \mathbf{w} being all zeros, and at each iteration have the algorithm choose a point randomly from the set of misclassified points.	5
Câu 8: Which of the following is closest to $Pr(f(x) \neq g(x))$ for $N = 10$?	5
Câu 9: Now, try $N = 100$. How many iterations does it take on average for the PLA to converge for $N=100$ training points? Pick the value closest to your results.	6
Câu 10: Which of the following is closest to $Pr(f(x) \neq g(x))$ for $N = 100$?	6

Câu 1: What types of learning, if any, best describe the following three scenarios:

- (i) A coin classification system is created for a vending machine. In order to do this, the developers obtain exact coin specifications from the U.S. Mint and derive a statistical model of the size, weight, and denomination, which the vending machine then uses to classify its coins.
- (ii) Instead of calling the U.S. Mint to obtain coin information, an algorithm is presented with a large set of labeled coins. The algorithm uses this data to infer decision boundaries which the vending machine then uses to classify its coins.
- (iii) A computer develops a strategy for playing Tic-Tac-Toe by playing repeatedly and adjusting its strategy by penalizing moves that eventually lead to losing.

Trả lời:

[d] (i) Not learning, (ii) Supervised Learning, (iii) Reinforcement Learning

Giải thích:

- (i): Sử dụng việc tìm hàm thống kê kích thước theo mệnh giá của đồng tiền để phân lớp → không phải learning
- (ii): Sử dụng dữ liệu về tiền xu cùng nhãn để thực hiện việc tạo mô hình phân lớp → Supervised Learning
- (iii): Trong quá trình chơi các ván cờ, máy tính sẽ học cách đánh giá được một trạng thái của bàn cờ → tương ứng với Reinforcement Learning gồm input thử, kết quả phép thử, và đánh giá

Câu 2: Which of the following problems are best suited for the learning approach?

- (i) Classifying numbers into primes and non-primes.
- (ii) Detecting potential fraud in credit card charges.
- (iii) Determining the time it would take a falling object to hit the ground.
- (iv) Determining the optimal cycle for traffic lights in a busy intersection.

Trả lời: [a] (ii) and (iv)

Giải thích:

- (i): Là bài toán phân lớp nhưng đã biết được hàm f để phân lớp số nguyên tố hoặc không phải số nguyên tố
→ Không sử dụng máy học
- (ii): Phát hiện tiềm năng gian lận tín dụng là một bài toán hiện:
 - Không có hàm f cụ thể để phát hiện có gian lận tín dụng hay không.

- Tồn tại một quy luật để phát hiện việc gian lận qua dữ liệu trong quá khứ (bằng chứng là trong một số trường hợp con người có thể phân tích để phát hiện việc gian lận)
- Có thể thu thập được dữ liệu

→ Có thể sử dụng máy học

(iii): Bài toán xác định thời gian một vật thể rơi xuống đất đã có hàm giải quyết tổng quát

→ Không sử dụng máy học

(iv): Xác định vòng xoay tối ưu cho đèn giao thông là một bài toán hiện:

- Không có hàm f cụ thể để phát hiện có gian lận tín dụng hay không.
- Tồn tại một quy luật để phát hiện việc gian lận qua dữ liệu trong quá khứ (bằng chứng là trong một số trường hợp con người có thể phân tích để phát hiện việc gian lận)
- Có thể thu thập được dữ liệu

→ Có thể sử dụng máy học

Câu 3: We have 2 opaque bags, each containing 2 balls. One bag has 2 black balls and the other has a black ball and a white ball. You pick a bag at random and then pick one of the balls in that bag at random. When you look at the ball, it is black. You now pick the second ball from that same bag. What is the probability that this ball is also black?

Trả lời: [d] $2/3$

Giải thích:

Gọi $P(A)$: Xác suất bốc được bi đen

Gọi $P(B)$: Xác suất bốc được bi trắng

$$\Rightarrow P(A) = 1/4; P(B) = 3/4$$

Ta có:

$P(AB)$: Xác suất bốc 2 lần liên tiếp, bốc được bi đen bi trắng

$P(B/A)$: Xác suất bốc được bi trắng sau khi bốc được bi đen

$$P(AB) = 1/2$$

$$P(B/A) = P(AB) / P(B) = (1/2)/(3/4) = 2/3$$

Câu 4: We draw only one such sample. Compute the probability that $\vartheta = 0$. The closest answer is (closest is the answer that makes the expression j your answer given option j closest to 0):

Trả lời: [b] 3.405×10^{-4}

Giải thích:

Ta có xác suất rút được bi xanh: $1 - 0.55 = 0.45$

Xác suất không rút được bi đỏ trong 10 lần rút (rút toàn bi xanh): $0.45^{10} \approx 3.405 \times 10^{-4}$

Câu 5: We draw 1,000 independent samples. Compute the probability that (at least) one of the samples has $\vartheta = 0$. The closest answer is:

[c] 0.289

Giải thích:

Xác suất một mẫu có ít nhất 1 bi đỏ: $1 - 3.405 \times 10^{-4} = 0.9996595$

Xác suất một nghìn mẫu toàn các mẫu có ít nhất 1 bi đỏ: $0.9996595^{1000} = 0.7113732776$

Xác suất một nghìn mẫu có một mẫu không chứa bi đỏ: $1 - 0.7113732776 = 0.28862672239$

Câu 6: Which hypothesis g agrees the most with the possible target functions in terms of the above score?

Trả lời: [e] They are all equivalent.

Giải thích:

	PTF1	PTF2	PTF3	PTF4	PTF5	PTF6	PTF7	PTF8
101	0	0	0	1	1	1	0	1
110	0	0	1	0	1	0	1	1
111	0	1	0	0	0	1	1	1

[a]: cho kết quả là của 3 trường hợp là 111

→ Score = $1*3 + 3*2 + 3*1 + 1*0 = 12$

[b]: cho kết quả là của 3 trường hợp là 000

→ Score = $1*3 + 3*2 + 3*1 + 1*0 = 12$

[c]: cho kết quả là của 3 trường hợp là 001

→ Score = $1*3 + 3*2 + 3*1 + 1*0 = 12$

[d]: cho kết quả là của 3 trường hợp là 110

→ Score = $1*3 + 3*2 + 3*1 + 1*0 = 12$

→ Tất cả đều tương đương

Câu 7: Take $N = 10$. Run the Perceptron Learning Algorithm to find g and measure the difference between f and g as $Pr(f(x) \neq g(x))$ (you can either calculate this exactly, or approximate it by generating a sufficiently large separate set of points to evaluate it). Repeat the experiment for 1000 runs (as specified above) and take the average. Start the PLA with the weight vector w being all zeros, and at each iteration have the algorithm choose a point randomly from the set of misclassified points.

How many iterations does it take on average for the PLA to converge for $N = 10$ training points? Pick the value closest to your results (again, closest is the answer that makes the expression $|\text{your answer} - \text{given option}|$ closest to 0).

Trả lời: [b]15

Cách sử dụng: Gọi hàm `PLAFunction()` theo hình dưới, với I là số lần duyệt trung bình

```
>> [I , Pr] = PLAFunction()

I =

    14.1510

Pr =

    0.1085
```

Figure 1 – Câu 7

Câu 8: Which of the following is closest to $Pr(f(x) \neq g(x))$ for $N = 10$?

Trả lời: [c] 0.1

Cách sử dụng: Gọi hàm `PLAFunction()` theo hình dưới, với Pr là giá trị $Pr(f(x) \neq g(x))$ cần tìm

```
>> [I , Pr] = PLAFunction()

I =

    14.1510

Pr =

    0.1085
```

Figure 2 – Câu 8

Câu 9: Now, try $N = 100$. How many iterations does it take on average for the PLA to converge for $N=100$ training points? Pick the value closest to your results.

Trả lời: [b] 100

Cách sử dụng: Gọi hàm `PLAFunction100()` theo hình dưới, với I là số lần duyệt trung bình

```
>> [I , Pr] = PLAFunction100()
```

```
I =
|
194.5410
```

```
Pr =

0.0161
```

Figure 3 - Câu 9

Câu 10: Which of the following is closest to $Pr(f(x) \neq g(x))$ for $N = 100$?

Trả lời: [b] 0.01

Cách sử dụng: Gọi hàm `PLAFunction100()` theo hình dưới, với Pr là giá trị $Pr(f(x) \neq g(x))$ cần tìm

```
>> [I , Pr] = PLAFunction100()
```

```
I =
|
194.5410
```

```
Pr =

0.0161
```

Figure 4 - Câu 10