(Answers) ML\_mid.pdf (4/2022) (10 câu)

[**1. Questions 2**](#_p7arjvlfkaap)

[**1. Answers 2**](#_m8ql2xowgu4l)

[[Thắng]: 2](#_5qe9jhilw32y)

[**2. Questions 2**](#_1fdk20umwcw)

[**2. Answers 2**](#_twykqrr65ec)

[[Lâm]: 2](#_27uhpb190mkz)

[**3. Questions 2**](#_ndyhu47qsrr2)

[**3. Answers 3**](#_fdqabnvs1fks)

[[Lâm]: 3](#_x8xr8vqrb60)

[**4. Questions 3**](#_fzhnc3xp8bg8)

[**4. Answers 3**](#_8x8uqxlu1bbm)

[[Thắng]: 3](#_ggrzuytt5ecj)

[**5. Questions 3**](#_1ir2c4zchdcq)

[**5. Answers 4**](#_lyq4mqw1mops)

[[Thắng]: 4](#_kdo66xo3ypwd)

[**6. Questions 4**](#_2a6oz1qr1ca)

[**6. Answers 4**](#_41kb4x5892yy)

[[Lâm] 4](#_rfg1oortk5ll)

[[Author 2] 5](#_cnydmm2e30cb)

[**7. Questions 5**](#_kgri8s68m0pv)

[**7. Answers 5**](#_34ka3yd3o18)

[[Lâm] 5](#_3gycqixniaaz)

[[Author 2] 5](#_9480jib6x4k4)

[**8. Questions 5**](#_fgm1x7tfluht)

[**8. Answers 5**](#_bhl8n124o650)

[[Lâm] 5](#_nrpn8ic56pgu)

[[Trâm] 2^9 5](#_rnjm7jp9nkxn)

[**9. Questions 6**](#_ntda3wurt9dx)

[**9. Answers 6**](#_x0hea7h18f0q)

[[Lâm] 6](#_piz0hnu15ysa)

[[Author 2] 6](#_4p1m51ii5kcc)

[**10. Questions 6**](#_94qvtnmnl639)

[**10. Answers 7**](#_w73ace66hsg5)

[[Lâm] 7](#_kzx1dzu7tirk)

[[Author 2] 7](#_uwb79gn4bzib)

### **1. Questions**

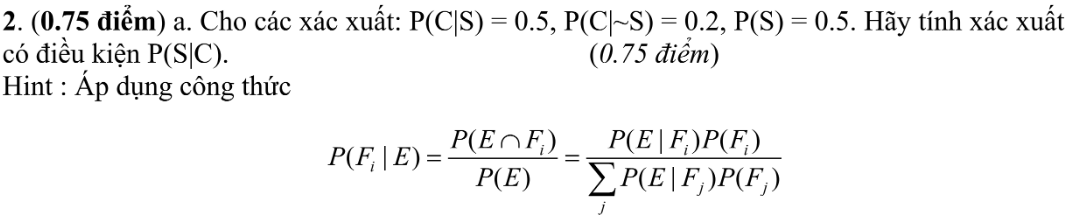
|  |
| --- |

### **1. Answers**

#### [Thắng]:

* **KNN**: đều phù hợp với cả 2 loại dữ liệu liên tục và dữ liệu rời rạc.
* **Decision Tree**: có thể áp dụng cho cả 2 loại dữ liệu liên tục và dữ liệu rời rạc. Tuy nhiên, đối với dữ liệu liên tục, thì cần áp dụng thêm một số phương pháp để xử lý dữ liệu. (pp rời nhánh liên tiếp, pp rời nhánh nhị phân, pp rời nhánh theo hàm số )
* **Naive Bayes**: có thể áp dụng cho cả 2 loại dữ liệu liên tục và dữ liệu rời rạc. Tuy nhiên, khi sử dụng cho dữ liệu liên tục, ta cần phải thực hiện các phương pháp tiền xử lý để đưa dữ liệu về dạng rời rạc hoặc sử dụng phân phối xác suất phù hợp.

### **2. Questions**



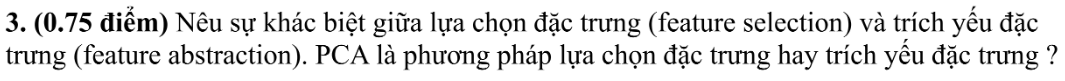
### **2. Answers**

#### **[Lâm]:**

mà mình đã có P(S), P(C|S) rồi nên chỉ cần tìm thêm P(C).

thay thế lên trên được

### **3. Questions**



### **3. Answers**

#### **[Lâm]:**

Lựa chọn đặc trưng (feature selection) là quá trình lựa chọn một tập con các thuộc tính đặc trưng từ một tập dữ liệu ban đầu để giảm chiều dữ liệu và cải thiện hiệu quả của các phương pháp học máy. Trích yếu đặc trưng (feature abstraction) là quá trình biến đổi hoặc kết hợp các thuộc tính đặc trưng ban đầu thành các thuộc tính mới có ý nghĩa cao hơn.

PCA (phân tích thành phần chính) là một phương pháp trích yếu đặc trưng, vì nó biến đổi các thuộc tính gốc thành các thành phần chính mới là tổ hợp tuyến tính của các thuộc tính gốc.

### **4. Questions**

|  |
| --- |

### **4. Answers**

#### [Thắng]:

* giả sử chúng ta tính trọng số theo cách sau:



N\_3 = 1/0.83 + 1/1.06 = 2.14

N\_2 = 1/1 + 1/1.02 + 1/1.08 = 2.9

vì N\_2 > N\_3 nên P sẽ phân vào lớp 2

### **5. Questions**

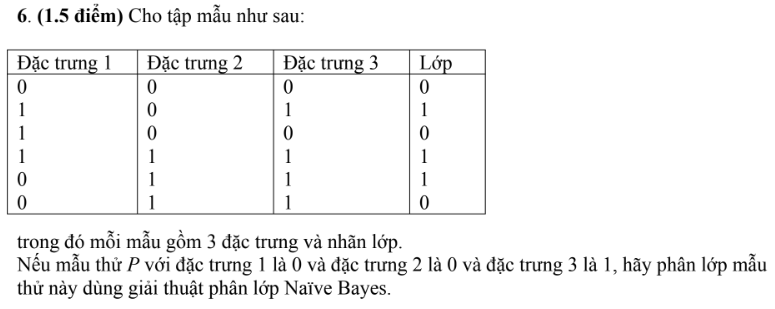
|  |
| --- |

### **5. Answers**

#### [Thắng]:

* **if** age = youth, student = no **then** result = no.
* **if** age = youth, student = yes **then** result = yes.
* **if** age = middle\_aged **then** result = yes.
* **if** age = senior, credit\_rating= fair **then** result = no.
* **if** age = senior, credit\_rating= excellent **then** result = yes.

### **6. Questions**



### **6. Answers**

#### [Lâm]

Gọi X là sự kiện DT1=0 và DT2=0 và DT3=1.

**P(L=0 | X)** = P(L=0) . P(DT1=0 | L=0) . P(DT2=0 | L=0) . P(DT3=1 | L=0)

P(L=0) = 3/6 = 0.5

P(DT1=0 | L=0) = ⅔

P(DT2=0 | L=0) = ⅔

P(DT3=1 | L=0) = ⅓

P(L=0 | X) = ½ \* ⅔ \* ⅔ \* ⅓ = 2/27

**P(L=1 | X)** = P(L=1) . P(DT1=0 | L=1) . P(DT2=0 | L=1) . P(DT3=1 | L=1)

P(L=1) = 3/6 = 0.5

P(DT1=0 | L=1) = ⅓

P(DT2=0 | L=1) = ⅓

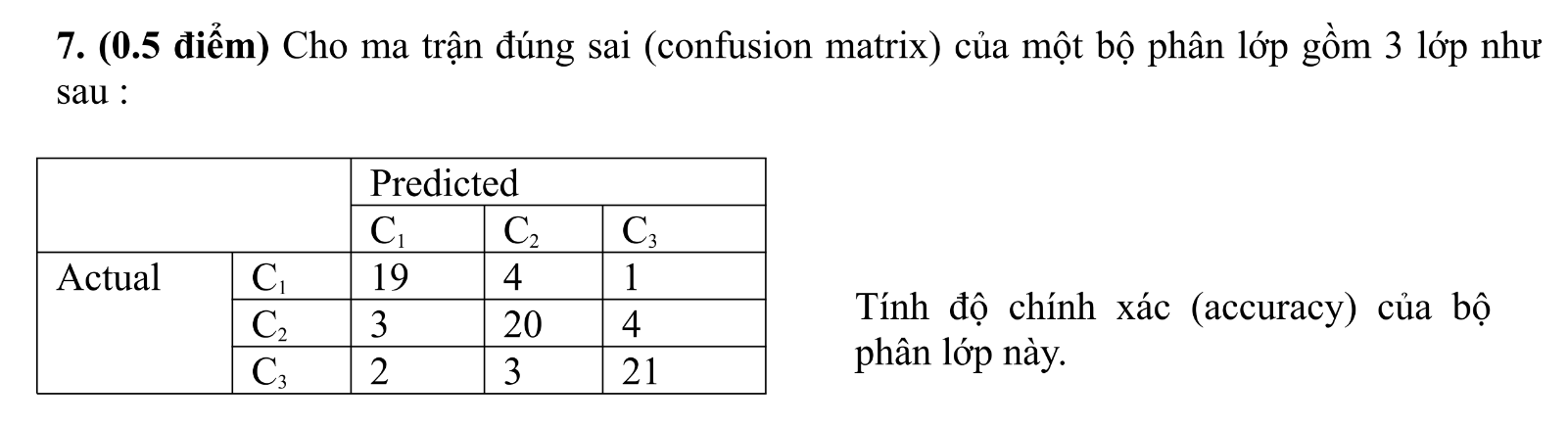
P(DT3=1 | L=1) = 3/3 = 1

P(L=1 | X) = ½ \* ⅓ \* ⅓ \* 1 = 1/18 < P(L=0 | X)

⇒ Mẫu thử P thuộc về lớp 0.

#### [Author 2]

### **7. Questions**



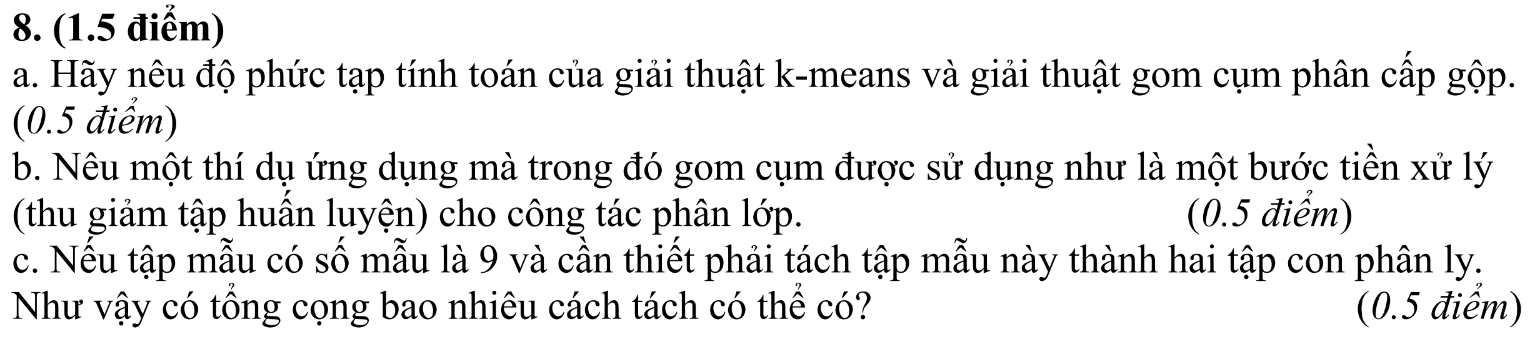
### **7. Answers**

#### [Lâm]

Accuracy = (Total correctly classified) / (Total number of samples) = (19+20+21)/(19+4+1+3+20+4+2+3+21) = 60/77

#### [Author 2]

### **8. Questions**



### **8. Answers**

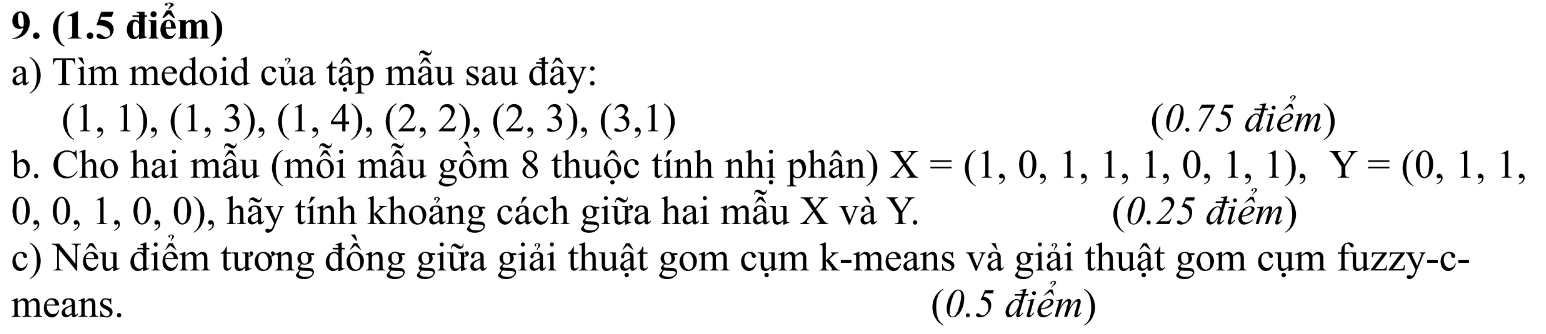
#### [Lâm]

8c. [combinatorics - How many ways to partition $n$ elements into two nonempty subsets? - Mathematics Stack Exchange](https://math.stackexchange.com/questions/3340723/how-many-ways-to-partition-n-elements-into-two-nonempty-subsets)

vì chỉ phân ra 2 tập thì ví dụ A=1,2,3 B=4,5,6,7,8,9 cũng giống B=1,2,3 A=4,5,6,7,8,9 và phân ra 2 tập thì không có tập rỗng nên -1 trường hợp tập rỗng.

#### [Trâm] 2^9

### **9. Questions**



### **9. Answers**

#### [Lâm]

9a. Ma trận khoảng cách

|  | 1 | 2 | 3 | 4 | 5 | 6 | Total |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 0 | 2 | 3 | 1.41 | 2.24 | 2 | 10.65 |
| 2 |  | 0 |  |  |  |  | 8.24 |
| 3 |  |  | 0 |  |  |  | 11.25 |
| 4 |  |  |  | 0 |  |  | 7.47 |
| 5 |  |  |  |  | 0 |  | 7.89 |
| 6 |  |  |  |  |  | 0 | 12.08 |

Medoid là mẫu số 4: (2, 2).

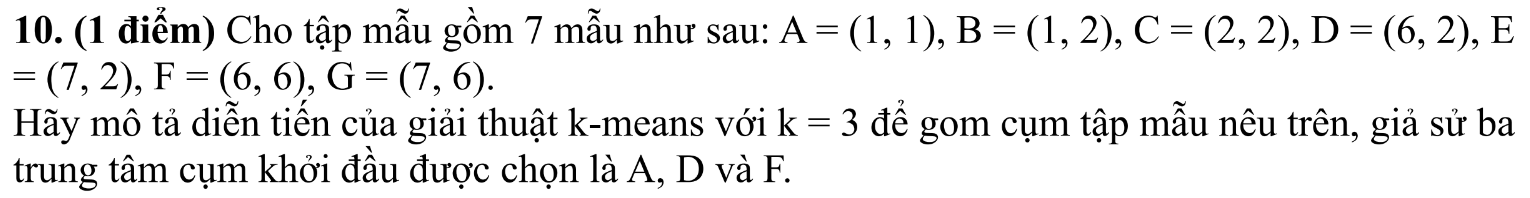
9b. Khoảng cách Euclid:

[Hamming distance](https://en.wikipedia.org/wiki/Hamming_distance) calculates the distance between two binary vectors, also referred to as binary strings or bitstrings for short.

Khoảng cách Hamming: 7

#### [Author 2]

### **10. Questions**



### **10. Answers**

#### [Lâm]

* Gán từng mẫu vào trung tâm cụm gần nhất:
  + Cluster 1: A(1,1), B(1,2), C(2,2)
  + Cluster 2: D(6,2), E(7,2)
  + Cluster 3: F(6,6), G(7,6)
* Tính các centroid mới:
  + Cluster 1: centroid\_cluster1 = (4/3, 5/3)
  + Cluster 2: centroid\_cluster2 = (6.5, 2)
  + Cluster 3: centroid\_cluster3 = (6.5, 6)
* Tiếp tục gán từng mẫu vào trung tâm cụm gần nhất:
  + Cluster 1: A(1,1), B(1,2), C(2,2)
  + Cluster 2: D(6,2), E(7,2)
  + Cluster 3: F(6,6), G(7,6)
  + Không thay đổi ⇒ Dừng giải thuật.

#### [Author 2]