**Chapter 3 (k-Nearest-Neighbor Classifier)**

### 

[**3.1 Question 1**](#_b8l95b2iavv4)

[**3.1 Giải: 2**](#_e2wxp3hcrlxa)

[**3.2 Question 2**](#_53w20n93rane)

[**3.2 Giải: 3**](#_9ov4s8omjmoh)

[**3.3 Question 3**](#_om19um2jd91u)

[**3.4 Question 3**](#_iit9gipe2cgi)

[**3.5 Question 3**](#_qhhkyr5g1cbg)

[**3.6 Question 4**](#_dd63ogjn37y7)

[**3.3 Question 6**](#_2cva7rc29wkw)

[**3.3 Answer 7**](#_t93smxjp0kra)

[**3.4 Question 7**](#_dvxa1z91frfc)

[**3.4 Answer 8**](#_gvotbsg9s5rz)

[**3.5 Question 12**](#_j8j15ags7n69)

[**3.5 Answer 12**](#_8pdmunflwj5i)

[**3.6 Question 13**](#_ki8ork53l8z7)

[**3.6 Answer 13**](#_4hl6i8q61ula)

#### **3.1 Question**

Consider the set of two-dimensional patterns:

(1, 1, 1), (1, 2, 1), (1, 3, 1), (2, 1, 1), (2, 2, 1), (2, 3, 1), (2, 3.5, 1),

(2.5, 2, 1), (3.5, 1, 1), (3.5, 2, 1), (3.5, 3, 2), (3.5, 4, 2), (4.5, 1, 2),

(4.5, 2, 2), (4.5, 3, 2), (5, 4, 2), (5, 5, 2), (6, 3, 2), (6, 4, 2), (6, 5, 2)

where each pattern is represented by feature 1, feature 2 and the class.

1. If the pattern *P* is (3.8, 3.1), find the class of *P* using the 1-nearest-neighbor algorithm.
2. Find the class of *P* using the *k*-nearest-neighbor algorithm (*k* = 3).

*Hãy xem xét tập hợp các mẫu hai chiều:*

*(1, 1, 1), (1, 2, 1), (1, 3, 1), (2, 1, 1), (2, 2, 1), (2, 3, 1), (2, 3.5, 1),  
(2.5, 2, 1), (3.5, 1, 1), (3.5, 2, 1), (3.5, 3, 2), (3.5, 4, 2), (4.5, 1, 2),  
(4.5, 2, 2), (4.5, 3, 2), (5, 4, 2), (5, 5, 2), (6, 3, 2), (6, 4, 2), (6, 5, 2)*

*trong đó mỗi mẫu được đại diện bởi tính năng 1, tính năng 2 và lớp.*

1. *Nếu mẫu P là (3.8, 3.1), hãy tìm lớp của P bằng thuật toán 1 lân cận gần nhất.*
2. *Tìm lớp của P bằng thuật toán k-láng giềng gần nhất (k = 3).*

#### **3.1 Giải:**

1. Tính toán khoảng cách từ điểm P đến tất cả các điểm khác đã có sẵn, ở đây dùng khoảng cách Euclide:

* Điểm (1, 1, 1): 3,5
* Điểm (1, 2, 1): 3,00832179129826
* Điểm (1, 3, 1): 2,80178514522438
* Điểm (2, 1, 1): 2,76586333718787
* Điểm (2, 2, 1): 2,1095023109729
* Điểm (2, 3, 1): 1,80277563773199
* Điểm (2, 3.5, 1): 1,84390889145858
* Điểm (2.5, 2, 1): 1,70293863659264
* Điểm (3.5, 1, 1): 2,12132034355964
* Điểm (3.5, 2, 1): 1,14017542509914
* Điểm (3.5, 3, 2): 0,316227766016838
* Điểm (3.5, 4, 2): 0,948683298050514
* Điểm (4.5, 1, 2): 2,21359436211787
* Điểm (4.5, 2, 2): 1,30384048104053
* Điểm (4.5, 3, 2): 0,707106781186548
* Điểm (5, 4, 2): 1,5
* Điểm (5, 5, 2): 2,24722050542442
* Điểm (6, 3, 2): 2,20227155455452
* Điểm (6, 4, 2): 2,37697286480094
* Điểm (6, 5, 2): 2,90688837074973

Vậy, khi k = 1, class của P là 2

1. Khi k = 3, với khoảng cách Euclide được tinh ở trên, ta có thể thấy class của P là 2.

*Cách giải: Tính khoảng cách từ P đến tất cả các điểm trong 3 class trên.*

1. *Với k=1, class của P là class của điểm gần P nhất*
2. *Với k=3, class của P là class chiếm đa số trong 3 điểm gần P nhất*

#### **3.2 Question**

Consider the set of two-dimensional patterns:

(1, 1, 1), (1, 2, 1), (1, 3, 1), (2, 1, 1), (2, 2, 1), (2, 3, 1), (2, 3.5, 1),

(2.5, 2, 1), (3.5, 1, 1), (3.5, 2, 1), (3.5, 3, 2), (3.5, 4, 2), (4.5, 1, 2)

(4.5, 2, 2), (4.5, 3, 2), (5, 4, 2), (5, 5, 2), (6, 3, 2), (6, 4, 2), (6, 5, 2)

where each pattern is represented by feature 1, feature 2 and the class.

Find the centroids of the two classes. Use the minimum-distance classifier to find the class of the pattern *P* = (3.8, 3.1).

*Hãy xem xét tập hợp các mẫu hai chiều:*

*(1, 1, 1), (1, 2, 1), (1, 3, 1), (2, 1, 1), (2, 2, 1), (2, 3, 1), (2, 3.5, 1),*

*(2.5, 2, 1), (3.5, 1, 1), (3.5, 2, 1), (3.5, 3, 2), (3.5, 4, 2), (4.5, 1, 2)*

*(4.5, 2, 2), (4.5, 3, 2), (5, 4, 2), (5, 5, 2), (6, 3, 2), (6, 4, 2), (6, 5, 2)*

*trong đó mỗi mẫu được đại diện bởi tính năng 1, tính năng 2 và lớp.*

*Tìm trọng tâm của hai lớp. Sử dụng bộ phân loại khoảng cách tối thiểu để tìm lớp của mẫu P = (3.8, 3.1).*

#### **3.2 Giải:**

Điểm trung tâm của class 1 sẽ là (2.05, 2.05) (Tính trung bình của các điểm trên class 1)

Điểm trung tâm của class 2 sẽ là (4.85, 3.4)

Vậy, class của P theo phương pháp tính này sẽ là **Class 2**. *(Khoảng cách P đến trung tâm class 1 là 2,04083316319586 so với khoảng cách đến trung tâm class 2 là 1,09201648339208)*

*Cách giải: Tính giá trị trung bình của 2 feature để tìm ra centroid của các class.*

*Tính khoảng cách từ P đến 2 điểm centroid này. Class của P là class của centroid gần P hơn.*

#### **3.3 Question**

Consider the set of two-dimensional patterns:

(1, 1, 1), (1, 2, 1), (1, 3, 1), (2, 1, 1), (2, 2, 1), (2, 3, 1), (2, 3.5, 1),

(2.5, 2, 1), (3.5, 1, 1), (3.5, 2, 1), (3.5, 3, 2), (3.5, 4, 2), (4.5, 1, 2)

(4.5, 2, 2), (4.5, 3, 2), (5, 4, 2), (5, 5, 2), (6, 3, 2), (6, 4, 2), (6, 5, 2)

where each pattern is represented by feature 1, feature 2 and the class.

a) Find the condensed set using the condensed nearest neighbor algorithm.

b) If the patterns of Class 2 appears first, given the set

(3.5, 3, 2), (3.5, 4, 2), (4.5, 1, 2) (4.5, 2, 2), (4.5, 3, 2), (5, 4, 2), (5, 5, 2),

(6, 3, 2), (6, 4, 2), (6, 5, 2), (1, 1, 1), (1, 2, 1), (1, 3, 1), (2, 1, 1), (2, 2, 1),

(2, 3, 1), (2, 3.5, 1), (2.5, 2, 1), (3.5, 1, 1), (3.5, 2, 1),

find the condensed set using the condensed nearest neighbor algorithm.

#### **3.4 Question**

Consider the set of two-dimensional patterns:

(1, 1, 1), (1, 2, 1), (1, 3, 1), (2, 1, 1), (2, 2, 1), (2, 3, 1), (2, 3.5, 1),

(2.5, 2, 1), (3.5, 1, 1), (3.5, 2, 1), (3.5, 3, 2), (3.5, 4, 2), (4.5, 1, 2)

(4.5, 2, 2), (4.5, 3, 2), (5, 4, 2), (5, 5, 2), (6, 3, 2), (6, 4, 2), (6, 5, 2)

where each pattern is represented by feature 1, feature 2 and the class.

Rank the data set using the Naïve Rank Numerosity Reduction algorithm.

#### **3.5 Question**

State the difference between *prototype selection* and *prototype abstraction.* Is Naïve Rank Numerosity Reduction algorithm belongs to the category of prototype abstraction?

*Nêu sự khác biệt giữa lựa chọn nguyên mẫu và trừu tượng hóa nguyên mẫu. Thuật toán Naïve Rank Numerosity Reduction có thuộc danh mục trừu tượng hóa nguyên mẫu không?*

#### **3.6 Question**

We use a weighted 5-NN classifier to determine the class of object P. Assume that the distances between P and the five nearest neighbors (X1, X2, X3, X4 and X5) be d1 = 1, d2 = 3, d3 = 4, d4 = 5 and d5 =8. If X1, X2 belong to class + and the three others belong to class -. Which class is P assigned to?

Phạm Minh Tú

**3.1** Consider the set of two-dimensional patterns:

(1, 1, 1), (1, 2, 1), (1, 3, 1), (2, 1, 1), (2, 2, 1), (2, 3, 1), (2, 3.5, 1),

(2.5, 2, 1), (3.5, 1, 1), (3.5, 2, 1), (3.5, 3, 2), (3.5, 4, 2), (4.5, 1, 2)

(4.5, 2, 2), (4.5, 3, 2), (5, 4, 2), (5, 5, 2), (6, 3, 2), (6, 4, 2), (6, 5, 2)

where each pattern is represented by feature 1, feature 2 and the class.

(a) If the pattern *P* is (3.8, 3.1), find the class of P using the 1-nearest-neighbor algorithm.

(b) Find the class of P using the *k*-nearest-neighbor algorithm (*k* = 3).

(a) d(P, x1) = 3.500

d(P, x2) = 3.008

d(P, x3) = 2.802

d(P, x4) = 2.766

d(P, x5) = 2.110

d(P, x6) = 1.803

d(P, x7) = 1.844

d(P, x8) = 1.703

d(P, x9) = 2.121

d(P, x10) = 1.140

d(P, x11) = 0.316

d(P, x12) = 0.949

d(P, x13) = 2.214

d(P, x14) = 1.304

d(P, x15) = 0.707

d(P, x16) = 1.500

d(P, x17) = 2.247

d(P, x18) = 2.202

d(P, x19) = 2.377

d(P, x20) = 2.907

Class of P is 2

Để bấm máy nhanh các bạn có thể bấm máy biểu thức tổng quát vào máy tính casio sau đó CACL để thay số.

(b) 3 nearest neighbors are x11, x12, x15, and 3 of them come from class 2

**3.2** Consider the set of two-dimensional patterns:

(1, 1, 1), (1, 2, 1), (1, 3, 1), (2, 1, 1), (2, 2, 1), (2, 3, 1), (2, 3.5, 1),

(2.5, 2, 1), (3.5, 1, 1), (3.5, 2, 1), (3.5, 3, 2), (3.5, 4, 2), (4.5, 1, 2)

(4.5, 2, 2), (4.5, 3, 2), (5, 4, 2), (5, 5, 2), (6, 3, 2), (6, 4, 2), (6, 5, 2)

where each pattern is represented by feature 1, feature 2 and the class.

Find the centroids of the two classes. Use the minimum-distance classifier to find the class of the pattern *P* = (3.8, 3.1).

Centroids of class 1 is c1

x\_(c\_1 ) =(1+1+1+2+2+2+2+2.5+3.5+3.5)/10=2.05

y\_(c\_1 ) =(1+2+3+1+2+3+3.5+2+1+2)/10=2.05

Centroids of class 2 is c2

x\_(c\_2 ) =(3.5+3.5+4.5+4.5+4.5+5+5+6+6+6)/10=4.85

y\_(c\_2 ) =(3+4+1+2+3+4+5+3+4+5)/10=3.4

d(P,c1) = 2.04083

d(P,c2) = 1.09202

P belongs to class2

#### **3.3 Question**

Consider the set of two-dimensional patterns:

(1, 1, 1), (1, 2, 1), (1, 3, 1), (2, 1, 1), (2, 2, 1), (2, 3, 1), (2, 3.5, 1),

(2.5, 2, 1), (3.5, 1, 1), (3.5, 2, 1), (3.5, 3, 2), (3.5, 4, 2), (4.5, 1, 2)

(4.5, 2, 2), (4.5, 3, 2), (5, 4, 2), (5, 5, 2), (6, 3, 2), (6, 4, 2), (6, 5, 2)

where each pattern is represented by feature 1, feature 2 and the class.

a) Find the condensed set using the condensed nearest neighbor algorithm.

b) If the patterns of Class 2 appears first, given the set

(3.5, 3, 2), (3.5, 4, 2), (4.5, 1, 2) (4.5, 2, 2), (4.5, 3, 2), (5, 4, 2), (5, 5, 2),

(6, 3, 2), (6, 4, 2), (6, 5, 2), (1, 1, 1), (1, 2, 1), (1, 3, 1), (2, 1, 1), (2, 2, 1),

(2, 3, 1), (2, 3.5, 1), (2.5, 2, 1), (3.5, 1, 1), (3.5, 2, 1),

find the condensed set using the condensed nearest neighbor algorithm.

*Hãy xem xét tập hợp các mẫu hai chiều:*

*(1, 1, 1), (1, 2, 1), (1, 3, 1), (2, 1, 1), (2, 2, 1), (2, 3, 1), (2 , 3.5, 1),*

*(2.5, 2, 1), (3.5, 1, 1), (3.5, 2, 1), (3.5, 3, 2), (3.5, 4, 2), (4.5, 1, 2)*

*(4.5, 2, 2), (4.5, 3, 2), (5, 4, 2), (5, 5, 2), (6, 3, 2), (6, 4, 2), (6 , 5, 2)*

*trong đó mỗi mẫu được đại diện bởi tính năng 1, tính năng 2 và lớp.*

*a) Tìm tập cô đọng bằng thuật toán láng giềng gần nhất cô đọng.*

*b) Nếu các mẫu của Loại 2 xuất hiện trước, với tập hợp*

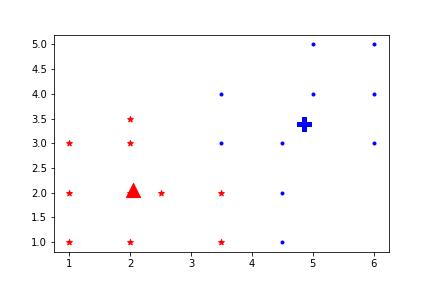
*(3.5, 3, 2), (3.5, 4, 2), (4.5, 1, 2) (4.5, 2, 2), (4.5, 3, 2), (5, 4, 2), (5, 5, 2),*

*(6, 3, 2), (6, 4, 2), (6, 5, 2), (1, 1, 1), (1, 2, 1), (1, 3, 1), (2 , 1, 1), (2, 2, 1),*

*(2, 3, 1), (2, 3.5, 1), (2.5, 2, 1), (3.5, 1, 1), (3.5, 2, 1),*

*tìm tập hợp thu gọn bằng thuật toán hàng xóm gần nhất thu gọn.*

#### **3.3 Answer**



Theo thuật toán ta có lời giải như sau:

a) X1, X10, X6, X8, X12

b) X1, X10, X15, X18, X3

#### **3.4 Question**

Consider the set of two-dimensional patterns:

(1, 1, 1), (1, 2, 1), (1, 3, 1), (2, 1, 1), (2, 2, 1), (2, 3, 1), (2, 3.5, 1),

(2.5, 2, 1), (3.5, 1, 1), (3.5, 2, 1), (3.5, 3, 2), (3.5, 4, 2), (4.5, 1, 2)

(4.5, 2, 2), (4.5, 3, 2), (5, 4, 2), (5, 5, 2), (6, 3, 2), (6, 4, 2), (6, 5, 2)

where each pattern is represented by feature 1, feature 2 and the class.

Rank the data set using the Naïve Rank Numerosity Reduction algorithm.

*Hãy xem xét tập hợp các mẫu hai chiều:*

*(1, 1, 1), (1, 2, 1), (1, 3, 1), (2, 1, 1), (2, 2, 1), (2, 3, 1), (2 , 3.5, 1),*

*(2.5, 2, 1), (3.5, 1, 1), (3.5, 2, 1), (3.5, 3, 2), (3.5, 4, 2), (4.5, 1, 2)*

*(4.5, 2, 2), (4.5, 3, 2), (5, 4, 2), (5, 5, 2), (6, 3, 2), (6, 4, 2), (6 , 5, 2)*

*trong đó mỗi mẫu được đại diện bởi tính năng 1, tính năng 2 và lớp.*

*Xếp hạng tập dữ liệu bằng thuật toán Giảm số thứ hạng Naïve.*

#### **3.4 Answer**

· First iteration: S = {}

rank(x1) = 2

rank(x2) = 3

rank(x3) = 2

rank(x4) = 2

rank(x5) = 1

rank(x6) = 1

rank(x7) = 1

rank(x8) = 1

rank(x9) = -1 +

rank(x10) = -3

rank(x11) = 0 +

rank(x12) = 1

rank(x13) = -1

rank(x14) = 0 +

rank(x15) = 2

rank(x16) = 2

rank(x17) = 2

rank(x18) = 1

rank(x19) = 3

rank(x20) = 2

· Second iteration: S = {x10}

rank(x1) = 2

rank(x2) = 3

rank(x3) = 2

rank(x4) = 2

rank(x5) = 1

rank(x6) = 1

rank(x7) = 1

rank(x8) = 1

rank(x9) = -2

rank(x11) = 2

rank(x12) = 1

rank(x13) = -1 +

rank(x14) = 2

rank(x15) = 2

rank(x16) = 2

rank(x17) = 2

rank(x18) = 1

rank(x19) = 3

rank(x20) = 2

· Thirth iteration: S = {x9, x10}

rank(x1) = 2

rank(x2) = 3

rank(x3) = 2

rank(x4) = 2

rank(x5) = 1, p = 2

rank(x6) = 1, p = 2

rank(x7) = 1, p = 2

rank(x8) = 1, p = 2

rank(x11) = 2 +

rank(x12) = 1, p = 1

rank(x13) = 1, p = 1

rank(x14) = 2

rank(x15) = 2

rank(x16) = 2

rank(x17) = 2

rank(x18) = 1, p = 1

rank(x19) = 3

rank(x20) = 2

· Fourth iteration: S = {x12, x9, x10}

rank(x1) = 2

rank(x2) = 3

rank(x3) = 2

rank(x4) = 2

rank(x5) = 1, p = 2

rank(x6) = 1, p = 2

rank(x7) = 1, p = 2

rank(x8) = 1, p = 2

rank(x11) = 1, p = 1

rank(x13) = 1, p = 1

rank(x14) = 2

rank(x15) = 2

rank(x16) = 2

rank(x17) = 2

rank(x18) = 1, p = 1

rank(x19) = 3

rank(x20) = 2

Fifth iteration get x11, sixth iteration get x13, seventh iteration get x14, eighth iteration get x15, ninth iteration get x18.

· Tenth iteration S = {x5, x18, x15, x14, x13, x11, x12, x9, x10}

rank(x1) = 2

rank(x2) = 3 +

rank(x3) = 2

rank(x4) = 2 +

rank(x5) = 1, p = 2

rank(x6) = 1, p = 2

rank(x7) = 1, p = 2

rank(x8) = 1, p = 2 +

rank(x16) = 2

rank(x17) = 2

rank(x19) = 3

rank(x20) = 2

· Eleventh iteration S = {x4, x5, x18, x15, x14, x13, x11, x12, x9, x10}

rank(x1) = 2 +

rank(x2) = 2

rank(x3) = 2

rank(x4) = 1, p = 1

rank(x6) = 1, p = 2

rank(x7) = 1, p = 2

rank(x8) = 2 +

rank(x16) = 2

rank(x17) = 2

rank(x19) = 3

rank(x20) = 2

· Twelve iteration S = {x1, x4, x5, x18, x15, x14, x13, x11, x12, x9, x10}

rank(x1) = 1, p = 1

rank(x2) = 2 +

rank(x3) = 2

rank(x6) = 1, p = 2

rank(x7) = 1, p = 2

rank(x8) = 1, p = 3

rank(x16) = 2

rank(x17) = 2

rank(x19) = 3

rank(x20) = 2

Thirteenth iteration get x2, fourteenth iteration get x3, fifteenth iteration get x6, sixteenth iteration get x7, seventeenth iteration get x8, eighteenth iteration get x16, nineteenth iteration get x17 and finally are x19 and x20 respectively.

S = {x20, x19, x17, x16, x8, x7, x6, x3, x2, x1, x4, x5, x18, x15, x14, x13, x11, x12, x9, x10}

*Lần lặp thứ 13 được x2, lần thứ 14 được x3, lượt thứ 15 được x6, lượt thứ 16 được x7, lượt thứ 17 được x8, lượt thứ 18 được x16, lượt thứ 19 được x17 và cuối cùng lần lượt là x19 và x20.*

*S = {x20, x19, x17, x16, x8, x7, x6, x3, x2, x1, x4, x5, x18, x15, x14, x13, x11, x12, x9, x10}*

#### **3.5 Question**

State the difference between *prototype selection* and *prototype abstraction.* Is Naïve Rank Numerosity Reduction algorithm belongs to the category of prototype abstraction?

*Nêu sự khác biệt giữa lựa chọn nguyên mẫu và trừu tượng hóa nguyên mẫu. Thuật toán Naïve Rank Numerosity Reduction có thuộc danh mục trừu tượng hóa nguyên mẫu không?*

#### **3.5 Answer**

**Prototype selection** là phương pháp lấy mẫu hoặc một tập mẫu để đại diện cho một tập lớn hơn với các phần tử được lấy ra là con của tập mẫu cho trước

**Prototype abstraction** là phương pháp lấy mẫu hoặc một tập mẫu để đại diện cho một tập lớn hơn với các phần tử được lấy ra không là con của tập mẫu cho trước, mà được tính ra từ tập mẫu cho trước

Thuật toán Rank Numerosity Reduction thuộc về lựa chọn nguyên mẫu (prototype selection)

#### **3.6 Question**

We use a weighted 5-NN classifier to determine the class of object P. Assume that the distances between P and the five nearest neighbors (X1, X2, X3, X4 and X5) be d1 = 1, d2 = 3, d3 = 4, d4 = 5 and d5 =8. If X1, X2 belong to class + and the three others belong to class -. Which class is P assigned to?

*Chúng tôi sử dụng bộ phân loại 5-NN có trọng số để xác định lớp của đối tượng P. Giả sử rằng khoảng cách giữa P và năm láng giềng gần nhất (X1, X2, X3, X4 và X5) là d1 = 1, d2 = 3, d3 = 4 , d4 = 5 và d5 =8. Nếu X1, X2 thuộc lớp + và 3 lớp còn lại thuộc lớp -. P được giao cho lớp nào?*

#### **3.6 Answer**

*(Slide 26, nặng tính toán, khá chua tí)*

w1 = (8-1)/(8-1) =1

w2 = (8-3)/(8-1) = 5/7

w3 = (8-4)/(8-1) = 4/7

w4 = (8-5)/(8-1) = 3/7

w5 = (8-8)/(8-1) = 0

sum+ = 1 + 5/7 = 12/7

sum- = 4/7+3/7 = 1 < sum+

P thuộc về lớp +