|  |  |
| --- | --- |
| TRƯỜNG ĐẠI HỌC HẠ LONG  **KHOA CÔNG NGHỆ THÔNG TIN**  **Mã đề:** IT60802201  (01 *là số thứ tự)* | KỲ THI KẾT THÚC HỌC PHẦN NĂM HỌC 2021–2022  ĐỀ THI HỌC PHẦN: **Học máy**  LỚP KHOA HỌC MÁY TÍNH  *(Thời gian làm bài: 90 phút)* |

**Câu 1. (***3,5 điểm)***.** Cho tập cơ sở dữ liệu thời tiết dưới đây

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Ngày** | **Trời** | **Nhiệt độ** | **Độ ẩm** | **Gió** | **Chơi tennis** |
| D1 | Nắng | Cao | Cao | Yếu | Không |
| D2 | Nắng | Cao | Cao | Mạnh | Không |
| D3 | U ám | Cao | Cao | Yếu | Có |
| D4 | Mưa | Trung bình | Cao | Yếu | Có |
| D5 | Mưa | Thấp | Bình thường | Yếu | Có |
| D6 | Mưa | Thấp | Bình thường | Mạnh | Không |
| D7 | U ám | Thấp | Bình thường | Mạnh | Có |
| D8 | Nắng | Trung bình | Cao | Yếu | Không |

Dùng thuật toán Bayes dự báo trường hợp sau:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Trời** | **Nhiệt độ** | **Độ ẩm** | **Gió** | **Chơi tennis** |
| D1 | U ám | Thấp | Cao | Yếu | ? |

**Câu 2** *(3,5 điểm)* Cho bảng dữ liệu về hoa IRIS gồm 4 thuộc tính như bảng sau:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| STT | A1 | A2 | A3 | A4 |
| 1 | 5.40 | 3.40 | 1.70 | 0.20 |
| 2 | 7.90 | 3.80 | 6.40 | 2.00 |
| 3 | 4.40 | 2.90 | 1.40 | 0.20 |
| 4 | 4.30 | 3.00 | 1.10 | 0.10 |
| 5 | 6.70 | 3.00 | 5.20 | 2.30 |
| 6 | 6.20 | 3.40 | 5.40 | 2.30 |
| 7 | 7.00 | 3.20 | 4.70 | 1.40 |
| 8 | 7.20 | 3.20 | 6.00 | 1.80 |
| 9 | 6.30 | 2.90 | 5.60 | 1.80 |
| 10 | 5.80 | 2.80 | 5.10 | 2.40 |

Áp dụng giải thuật k-Means để phân cụm dữ liệu với *k*=3. Tâm khởi tạo ngẫu nhiên là:

C1(5.40, 3.40, 1.70, 0.20); C2 (5.80, 2.80, 5.10, 2.40); C3 (7.00, 3.20, 4.70, 1.40).

**Câu 3** *(3 điểm):*

1. Thế nào là hồi quy? Trình bày thuật toán hồi quy tuyến tính?
2. Tại sao người ta gọi thuật toán k-NN là thuật toán lười? “k” có ý nghĩa gi? Cách chọn “k”?
3. Lề (Margin) trong thuật toán SVM là gì? Vẽ sơ đồ minh họa và viết công thức đồng thời giải thích ý nghĩa của các biến, tham số trong công thức đó.

*Thí sinh không được sử dụng tài liệu*

*………, ngày ……. tháng 4 năm 2022*

**NGƯỜI DUYỆT ĐỀ NGƯỜI RA ĐỀ**

**Nguyễn Quang Hoan**

**HƯỚNG DẪN CHẤM ĐỀ MÃ SỐ IT60802201**

**HỌC PHẦN HỌC MÁY HỆ ĐẠI HỌC NGÀNH KHOA HỌC MÁY TÍNH**

*Năm học 2021 - 2022*

***Mã số đề:*** IT60802201

|  |  |  |
| --- | --- | --- |
| **Câu/Ý** | **ĐÁP ÁN-HƯỚNG DẪN CHẤM** | **Điểm** |
| **1/1** | Gọi bốn đặc trưng đầu vào là A1=”Trời”; A2=” Nhiệt độ”; A3=”Độ ẩm”,, A4=”Gió”; B=”Chơi Tennis”.  Bước 1. Sử dụng phương pháp “chía để trị” theo 4 đặc trưng đầu vào như bảng sau   |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | **Trời** | | | **Nhiệt độ** | | | **Độ ẩm** | | | **Gió** | | | **Chơi tennis** | | | Không | | Có | Không | | Có | Không | | Có | Không | | Có | Không | Có | | Nắng | 3 | 0 | Cao | 2 | 1 | Cao | 3 | 2 | Yếu | 2 | 3 | 4 | 4 | | Mưa | 0 | 2 | TB | 1 | 1 | BT | 1 | 2 | Mạnh | 2 | 1 | | U ám | 1 | 2 | Thấp | 1 | 2 |  | | |  | | | | A1 | | | A2 | | | A3 | | | A4 | | | B | | | A | | | | | | | | | | | | | ***Mỗi cột ở***  ***Bảng bên***  ***đúng***  ***0.2***  ***\*5***  ***=1*** |
| **1/2** | Bước 2. Tính xác suất tương ứng theo 4 đặc trưng đầu vào như bảng sau   |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | **Trời** | | | **Nhiệt độ** | | | **Độ ẩm** | | | **Gió** | | | **Chơi tennis** | | | Không | | Có | Không | | Có | Không | | Có | Không | | Có | Không | Có | | Nắng | 3/4 | 0/4 | Cao | 2/4 | 1/4 | Cao | 3/4 | 2/4 | Yếu | 2/4 | 3/4 | 4/8 | 4/8 | | Mưa | 0/4 | 2/4 | TB | 1/4 | 1/4 | BT | 1/4 | 2/4 | Mạnh | 2/4 | 1/4 | | U ám | 1/4 | 2/4 | Thấp | 1/4 | 2/4 |  | | |  | | | | A1 | | | A2 | | | A3 | | | A4 | | | B | | | A | | | | | | | | | | | | | ***Mỗi cột ở***  ***Bảng bên***  ***đúng***  ***0.2***  ***\*5***  ***=1*** |
| **1/3** | **Giả sử**  là độc lập từng đôi một, ta có thể viết: | ***0.25*** |
| **1/4** | *Trường hợp B =không. Thay vào (\*) và lấy các giá trị cần dự báo có:*  (\*) = P(A1­=U ám |B=Không).P(A2 =Thấp|B=Không).P(A3 =Cao|B=Không).P(A4 =Yếu|B= Không) = = 0 = P(B|A) | ***0.5*** |
| **1/5** | *Ngược lại: B = Có*  (\*) = P(A­=U ám|B=Có).P(A2 ­=Thấp|B=Có).P(A3 ­=Cao|B=Có).P(A4=Yếu|B=Có)  = = = P(|A) | ***0.5*** |
| **1/6** | Vậy: *P(B|A) > P(|A) B=Có; Có đi chơi Tennis.* | ***0.25*** |
| **Câu 2** |  |  |
| **2/1** | **Bước 1:** Tính khoảng cách Ơ clit và gán cụm theo bảng sau   |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | STT | A1 | A2 | A3 | A4 | d­­­­2(i,C­­­­­1­) | d­­­­2(i,C­­­­­2) | d­­­­2(i,C­­­­­3) | Gán cụm | | 1 | 5.40 | 3.40 | 1.70 | 0.20 | 0 | 4.10 | 3.61 | 1 | | 2 | 7.90 | 3.80 | 6.40 | 2.00 | 5.60 | 2.70 | 2.10 | 3 | | 3 | 4.40 | 2.90 | 1.40 | 0.20 | 1.60 | 4.50 | 4.40 | 1 | | 4 | 4.30 | 3.00 | 1.10 | 0.10 | 1.30 | 4.90 | 4.70 | 1 | | 5 | 6.70 | 3.00 | 5.20 | 2.30 | 4.30 | 0.90 | 1.10 | 2 | | 6 | 6.20 | 3.40 | 5.40 | 2.30 | 4.30 | 0.78 | 1.40 | 2 | | 7 | 7.00 | 3.20 | 4.70 | 1.40 | 3.61 | 1.66 | 0 | 3 | | 8 | 7.20 | 3.20 | 6.00 | 1.80 | 4.93 | 1.81 | 1.374 | 3 | | 9 | 6.30 | 2.90 | 5.60 | 1.80 | 4.34 | 0.93 | 1.24 | 2 | | 14 | 5.80 | 2.80 | 5.10 | 2.40 | 4.113 | 0 | 1.66 | 2 | | C­­­­­1­ | 5.40 | 3.40 | 1.70 | 0.20 |  | | | | | C­­­­­2 | 5.80 | 2.80 | 5.10 | 2.40 | | C­­­­­3 | 7.00 | 3.20 | 4.70 | 1.40 | | ***0.5*** |
| **2/2** | C­1 = = (4.27, 3.1, 1.4, 0.167) | ***0.33*** |
| **2/3** | C­2= =(6.25, 3.0, 5.30, 2.20) | ***0.33*** |
| **2/4** | C­3 = = (7.37, 3.4, 5.7, 1.73) | ***0.33*** |
| **2/5** | **Bước 2.** Lặp vòng 1: (Bảng 2)   |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | STT | A1 | A2 | A3 | A4 | d­­­­2(i,C­­­­­1­) | d­­­­2(i,C­­­­­2) | d­­­­2(i,C­­­­­3) | Nhóm gần nhất | | 1 | 5.40 | 3.40 | 1.70 | 0.20 | 1.2 | 4.24 | 4.71 | 1 | | 2 | 7.90 | 3.80 | 6.40 | 2.00 | 6.48 | 2.125 | 1 | 3 | | 3 | 5.80 | 2.70 | 5.10 | 1.90 | 0.24 | 4.47 | 5.46 | 1 | | 4 | 6.20 | 2.90 | 4.30 | 1.30 | 0.32 | 5.165 | 5.779 | 1 | | 5 | 4.40 | 2.90 | 1.40 | 0.20 | 4.99 | 0.47 | 1.088 | 2 | | 6 | 5.80 | 4.00 | 1.20 | 0.20 | 4.93 | 0.398 | 1.133 | 2 | | 7 | 6.10 | 2.80 | 4.00 | 1.30 | 4.45 | 1.2 | 1.33 | 3 | | 8 | 4.30 | 3.00 | 1.10 | 0.10 | 5.69 | 1.244 | 1.404 | 3 | | 9 | 6.70 | 3.00 | 5.20 | 2.30 | 4.9 | 0.503 | 1.187 | 2 | | 10 | 6.20 | 3.40 | 5.40 | 2.30 | 4.59 | 0.586 | 1.906 | 2 | | C­­­­­1­ | 6.40 | 2.70 | 1.90 | 0.50 |  | | | | | C­­­­­2 | 6.20 | 3.0 | 5.30 | 2.20 | | C­­­­­3 | 6.50 | 3.20 | 4.50 | 1.30 | | ***0.5*** |
| **2/6** | C­1­ = = ( 4.27, 3.1, 1.4, 0.167) | ***0.33*** |
| **2/7** | C­2­== (6.25, 3.0, 5.30, 2.20) | ***0.33*** |
| **2/8** | C­3 = =(7.37, 3.4, 5.7, 1.73) |  |
| **2/9** | **Bước 3**. Thuật toán dừng vì C1, C2, C3 không đổi | ***0.5*** |
| **Câu 3** |  |  |
| **3/1** | * Viết khái niệm * Nêu đinh nghĩa hồi quy * Trình bày thuật toán hồi quy tuyến tính? | ***0,33***  ***0,33***  ***0,33*** |
|  | -Thuật toán k-NN là thuật toán lười vì thuật toán đó không tạo ra mô hình học máy (Learning Model) và hàm đích (Target Function) nên thuật toán xếp loại đơn giản (Simple Learning) hay học lười (Lazy Learning).  - “k” có ý nghĩa là một số mẫu quan sát được chọn từ các mẫu ví dụ có khoảng cách gần nhấy với mẫu cần xác định tính chất.  - Cách chọn “*k*”. Vì *k* mẫu giống như số phiếu bầu chọn (thường là + hoặc – hay Yes hoặc No giống phương châm bỏ phiếu bình chọn, do đó thường chọn số lẻ | ***0,33***  ***0,33***  ***0,33*** |
| **3/3** | Lề (Margin) trong thuật toán SVM là khoảng cách (*Đinh nghìa lề.* Lề *(Margin)* là khoảng cách giữa hai siêu phẳng hỗ trợ song song.   * Sơ đồ minh họa:      * Công thức (sinh viên giải thích thêm các tham số, khái niệm chuẩn 2:     trong đó, là độ lớn của véc tơ trọng số ***W*** theo chuẩn 2 *(2-norm*). | ***0,33***  ***0,33***  ***0,33*** |

*………, ngày ……. tháng …… năm 2022*

**NGƯỜI DUYỆT ĐỀ NGƯỜI RA ĐỀ**

**Nguyễn Quang Hoan**

|  |  |
| --- | --- |
| TRƯỜNG ĐẠI HỌC HẠ LONG  **KHOA CÔNG NGHỆ THÔNG TIN**  **Mã đề:** IT60802202  (02 *là số thứ tự)* | KỲ THI KẾT THÚC HỌC PHẦN NĂM HỌC 2021–2022  ĐỀ THI HỌC PHẦN: **Học máy**  LỚP KHOA HỌC MÁY TÍNH  *(Thời gian làm bài: 90 phút)* |

**Câu 1. (***4 điểm)***.** Cho tập cơ sở dữ liệu thời tiết dưới đây:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Ngày** | **Trời** | **Nhiệt độ** | **Độ ẩm** | **Gió** | **Chơi tennis** |
| D1 | Nắng | Cao | Cao | Yếu | Không |
| D2 | Nắng | Cao | Cao | Mạnh | Không |
| D3 | U ám | Cao | Cao | Yếu | Có |
| D4 | Mưa | Trung bình | Cao | Yếu | Có |
| D5 | Nắng | Trung bình | Bình thường | Mạnh | Có |
| D6 | Mưa | Thấp | Bình thường | Mạnh | Không |
| D7 | U ám | Thấp | Bình thường | Mạnh | Có |
| D8 | Nắng | Trung bình | Cao | Yếu | Không |
| D9 | Nắng | Thấp | Bình thường | Yếu | Có |
| D10 | Mưa | Trung bình | Cao | Mạnh | Không |

Dùng thuật toán Quinlan phân lớp (có chơi Tennis không) “Có” và “Không”.

**Câu 2** *(3 điểm)* Cho bảng dữ liệu kích thước hoa IRIS gồm 4 thuộc tính như bảng sau:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **STT** | **A1** | **A2** | **A3** | **A4** |  | **STT** | **A1** | **A2** | **A3** | **A4** |  |
| 1 | 5.1 | 3.8 | 1.5 | 0.3 | Iris-setosa | 9 | 7.6 | 3.0 | 6.6 | 2.1 | Iris-virginica |
| 2 | 5.4 | 3.4 | 1.7 | 0.2 | Iris-setosa | 10 | 4.9 | 2.5 | 4.5 | 1.7 | Iris-virginica |
| 3 | 5.1 | 3.7 | 1.5 | 0.4 | Iris-setosa | 11 | 7.3 | 2.9 | 6.3 | 1.8 | Iris-virginica |
| 4 | 4.6 | 3.6 | 1.0 | 0.2 | Iris-setosa | 12 | 6.7 | 2.5 | 5.8 | 1.8 | Iris-virginica |
| 5 | 5.1 | 3.8 | 1.5 | 0.3 | Iris-setosa | 13 | 7.2 | 3.6 | 6.1 | 2.5 | Iris-virginica |
| 6 | 5.2 | 3.5 | 1.5 | 0.2 | Iris-setosa | 14 | 6.5 | 3.2 | 5.1 | 2.0 | Iris-virginica |
| 7 | 5.2 | 3.4 | 1.4 | 0.2 | Iris-setosa | 15 | 6.4 | 2.7 | 5.3 | 1.9 | Iris-virginica |
| 8 | 4.7 | 3.2 | 1.6 | 0.2 | Iris-setosa | 16 | ***5.3*** | ***7.5*** | ***5.2*** | ***0.4*** | ? |

Với: chiều dài đài hoa: A1, chiều rộng đài hoa A2, chiều dài cánh hoaA3, chiều rông cánh hoa A4, loại hoa: d; Sử dụng thuật toán K-NN với số TT: 16, k=7, dự báo là hoa gì?

**Câu 3** *(3 điểm):*

1. Thế nào là hồi quy? Trình bày thuật toán hồi quy tuyến tính?
2. Trình bày thuật toán k-Mean. Thuật toán k-Mean thuộc kiểu gì? Tọa độ mỗi cụm cố định hay thay đổi. Nếu thây đổi được tính theo cách nào?
3. Lề (Margin) trong thuật toán SVM là gì? Vẽ sơ đồ minh họa và viết công thức đồng thời giải thích ý nghĩa của các biến, tham số trong công thức đó.

*Thí sinh không được sử dụng tài liệu*

*………, ngày ……. tháng… năm 2022*

**NGƯỜI DUYỆT ĐỀ NGƯỜI RA ĐỀ**

**Nguyễn Quang Hoan**

**HƯỚNG DẪN CHẤM ĐỀ MÃ SỐ IT60802202**

**HỌC PHẦN HỌC MÁY HỆ ĐẠI HỌC NGÀNH KHOA HỌC MÁY TÍNH**

*Năm học 2021 - 2022*

***Mã số đề:*** IT60802202

|  |  |  |
| --- | --- | --- |
| **Câu/Ý** | **ĐÁP ÁN-HƯỚNG DẪN CHẤM** | **Điểm** |
| **1/1** | |  |  |  |  | | --- | --- | --- | --- | | A1: Trời | | | | | **N**ắng | U ám | Mưa | | D1 | D3 | D4 | | D2 | D7 | D6 | | D5 |  | D10 | | D8 |  |  | | D9 |  |  |  |  |  |  | | --- | --- | --- | | A2: Nhiệt Độ | | | | Cao | TB | Thấp | | D1 | D4 | D6 | | D2 | D5 | D7 | | D3 | D8 | D9 | |  | D10 |  |  |  |  | | --- | --- | | A3: Độ Ẩm | | | Cao | B. Thường | | D1 | D5 | | D2 | D6 | | D3 | D7 | | D4 | D9 | | D8 |  | | D10 |  |  |  |  | | --- | --- | | A4: Gió | | | Mạnh | Yếu | | D1 | D2 | |  |  | | D3 | D5 | | D4 | D6 | | D8 | D7 | | D9 | D10 | | ***0.25\*4***  ***=1*** |
| **1/2** | Bước 2. Bước 2:   * Thuộc tính: Trời   V(trời = nắng) = (T(nắng,có),T(nắng,không)) = (2/5,3/5)  ***V(trời = u ám) = (T(u ám,có),T(u ám,không)) = (2/2,0) = (1,0)***  V(trời = mưa) = (T(mưa,có),T(mưa,không)) = (1/3,2/3)   * Thuộc tính: Nhiệt Độ   V(nhiệt độ = cao) = (T(cao,có),T(cao,không)) = (1/3,2/3)  V(nhiệt độ = trung bình) = (T(trung bình,có),T(trung bình,không)) = (2/4,2/4)  V(nhiệt độ = thấp) = (T(thấp,có),T(thấp,không)) = (2/3,1/3)   * Thuộc tính: Độ ẩm   V(độ ẩm = cao) = (T(cao,có),T(cao,không)) = (2/6,4/6)  V(độ ẩm = bình thường) = (T(bình thường,có),T(bình thường,không)) = (3/4,1/4)   * Thuộc tính: Gió   V(gió = mạnh) = (T(mạnh,có),T(mạnh,không)) = (3/5,2/5)  V(gió = yếu) = (T(yếu,có),T(yếu,không)) = (2/5,3/5) | ***0.25***  ***0.25***  ***0.25***  ***0.25*** |
| **1/3** | **Luật 1: If “trời” = “u ám” then “chơi tennis” = có** | ***0.25*** |
| **1/4** | Bước 3 Tổ hợp chập 2 đặc trưng. Trường hợp 1: trong phân hoạch A1 = nắng  **Thuộc tính : Nhiệt độ**  V(nhiệt độ = cao) = (T(cao,có),T(cao,không)) = (0/2,2/2)  V(nhiệt độ = trung bình) = (T(trung bình,có),T(trung bình,không)) = (0/1,1/1)  V(nhiệt độ = thấp) = (T(thấp,có),T(thấp,không)) = (1/1,0/1)  **Thuộc tính : độ ẩm**  V(độ ẩm = cao) = (T(cao,có),T(cao,không)) = (0/3,3/3)  V(độ ẩm = bình thường) = (T(bình thường,có),T(bình thường,không)) = (1/1,0/1)  **Thuộc tính : Gió**  V(độ ẩm = cao) = (T(cao,có),T(cao,không)) = (1/3,2/3)  V(độ ẩm = bình thường) = (T(bình thường,có),T(bình thường,không)) = (0/1,1/1)  **=> thuộc tính nhiệt độ sẽ được phân hoạch** | ***0.5*** |
| **1/5** | Trường hợp 2: trong phân hoạch A1 = mưa  **Thuộc tính : Nhiệt độ**  V(nhiệt độ = cao) = (T(cao,có),T(cao,không)) = (0,0)  V(nhiệt độ = trung bình) = (T(trung bình,có),T(trung bình,không)) = (2/2,0/2)  V(nhiệt độ = thấp) = (T(thấp,có),T(thấp,không)) = (1/2,1/2)  **Thuộc tính : độ ẩm**  V(độ ẩm = cao) = (T(cao,có),T(cao,không)) = (1/1,0/1)  V(độ ẩm = bình thường) = (T(bình thường,có),T(bình thường,không)) = (2/3,1/3)  **Thuộc tính : Gió**  *V(gió = mạnh) = (T(mạnh,có),T(có,không)) = (3/3,0/3)*  *V(gió = yếu) = (T(yếu,có),T(yếu,không)) = (0/1,1/1)*  **=> thuộc tính gió sẽ được phân hoạch** | ***0.5*** |
| **1/6** | Luật 2:If trời = nắng And nhiệt độ = cao Then chơi tennis = không;  Luật 3: If trời = nắng And nhiệt độ = trung bình Then chơi tennis = không;  Luật 4: If trời = nắng And nhiệt độ = thấp Then chơi tennis = có;  Luật 5: If trời = mưa And gió= mạnh Then chơi tennis = có;  Luật 6: If trời = mưa And gió= yếu Then chơi tennis = không. | ***0.15\*5***  **=0.75** |
| **Câu 2** |  |  |
| **2/1** | Kí hiệu hoa Iris-virginica là “-”, hoa Iris-setosa là “+” | ***0.25*** |
| **2/2** | **Bước 1:** Lập được bảng gồm 7 cột và các giá trị đã cho từ cột (1) đến cột (4)   |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | STT | **x1**  **(1)** | **x2**  (2) | **x3**  (3) | **x4**  (4) | Y  (5) | Distance  (6) | Chọn  (7) | | 1 | 5.1 | 3.8 | 1.5 | 0.3 | + | 5.237366 | + | | 2 | 5.4 | 3.4 | 1.7 | 0.2 | + | 5.395368 |  | | 3 | 5.1 | 3.7 | 1.5 | 0.4 | + | 5.307542 | + | | 4 | 4.6 | 3.6 | 1 | 0.2 | + | 5.777543 |  | | 5 | 5.1 | 3.3 | 1.7 | 0.5 | + | 5.471746 |  | | 6 | 5.2 | 3.5 | 1.5 | 0.2 | + | 5.453439 |  | | 7 | 5.2 | 3.4 | 1.4 | 0.2 | + | 5.59464 |  | | 8 | 4.7 | 3.2 | 1.6 | 0.2 | + | 5.64358 |  | | 9 | 7.6 | 3 | 6.6 | 2.1 | - | 5.512713 |  | | 10 | 4.9 | 2.5 | 4.5 | 1.7 | - | 5.228767 | - | | 11 | 7.3 | 2.9 | 6.3 | 1.8 | - | 5.322593 | - | | 12 | 6.7 | 2.5 | 5.8 | 1.8 | - | 5.4111 |  | | 13 | 7.2 | 3.6 | 6.1 | 2.5 | - | 4.90306 | - | | 14 | 6.5 | 3.2 | 5.1 | 2 | - | 4.743416 | - | | 15 | 6.4 | 2.7 | 5.3 | 1.9 | - | 5.148786 | - | |  | 5.3 | 7.5 | 5.2 | 0.4 | ? | k= | 7 | | ***0.5*** |
| **2/3** | Sắp xếp theo thứ tự “+ “ đến “-“ như cột (5) bảng trên | ***0.25*** |
| **2/4** | Tính được cột (6). Mỗi giá trị tính đúng: 0,1 điểm | ***0.1\*15***  **=1.5** |
| **2/5** | Tính đúng cột (7) | ***0.25*** |
| **2/6** | Số “+”=2<Số “-“=5. Suy ra hoa đó là hoa Iris-virginica la “-” | ***0.25*** |
| **Câu 3** |  |  |
| **3/1** | * Viết khái niệm * Nêu đinh nghĩa hồi quy * Trình bày thuật toán hồi quy tuyến tính | ***0.25***  ***0.25***  ***0.5*** |
| **3/2a**  **3/2b**  **3/2c** | -Thuật toán k-Mean  *Input: Cho CSDL học, cho số cụm K; Cho tọa độ của K cụm*  *Output: Tọa độ K cụm*  *B1: Tính n khoảng cách (n là số mẫu học). Tính tọa độ trung bình Gán cụm theo thứ tự khoảng cách*  *B2: Lặp lại bước 1*  *B3: Giá trị tâm cụm không thay đổi thì dừng.*  Thuật toán K-Mean thuộc loại học không giám sát. Lý do: không có nhãn  Tọa độ của K được tính lại bằng cách lấy trung bình cho mỗi lần lặp | ***0,5***  ***0.25***  ***0.25*** |
| **3/3** | Lề (Margin) trong thuật toán SVM là khoảng cách (*Đinh nghìa lề.* Lề *(Margin)* là khoảng cách giữa hai siêu phẳng hỗ trợ song song   * Sơ đồ minh họa:      * Công thức (sinh viên giải thích thêm các tham số, khái niệm chuẩn 2:     trong đó, là độ lớn của véc tơ trọng số ***W*** theo chuẩn 2 *(2-norm*). | ***0,33***  ***0,33***  ***0,33*** |

*………, ngày ……. tháng …… năm……..*

**NGƯỜI DUYỆT ĐỀ NGƯỜI RA ĐỀ**

**Nguyễn Quang Hoan**

|  |  |
| --- | --- |
| TRƯỜNG ĐẠI HỌC HẠ LONG  **KHOA CÔNG NGHỆ THÔNG TIN**  **Mã đề:** IT60802203  (03 *là số thứ tự)* | KỲ THI KẾT THÚC HỌC PHẦN NĂM HỌC 2021–2022  ĐỀ THI HỌC PHẦN: **Học máy**  LỚP KHOA HỌC MÁY TÍNH  *(Thời gian làm bài: 90 phút)* |

**Câu 1: (3.5 điểm)** Cho tập cơ sở dữ liệu thời tiết dưới đây

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Ngày** | **Trời** | **Nhiệt độ** | **Độ ẩm** | **Gió** | **Chơi tennis** |
| D1 | Nắng | Cao | Cao | Yếu | Không |
| D2 | Nắng | Cao | Cao | Mạnh | Không |
| D3 | U ám | Cao | Cao | Yếu | Có |
| D4 | Mưa | Trung bình | Cao | Yếu | Có |
| D5 | U ám | Cao | Bình thường | Yếu | Có |
| D6 | Mưa | Thấp | Bình thường | Mạnh | Không |
| D7 | U ám | Thấp | Bình thường | Mạnh | Có |
| D8 | Mưa | Trung bình | Cao | Mạnh | Không |
| D9 | Nắng | Thấp | Bình thường | Yếu | Có |
| D10 | Mưa | Trung bình | Bình thường | Yếu | Có |
| **A** |  |  |  |  | **B** |
| **Vào** | | | | **Ra** |

1. Dùng thuật toán Qinlan lập bảng chia 4 gốc và đầu ra (có chơi Tennis không)
2. Xác định = và cho tỷ lệ khi bằng 1, Viết luật
3. Nêu các bước tiếp theo cần tính. Khi nào thì thuậ toán dừng?

**Câu 2: (3,5 điểm)** Cho tập dữ liệu về hoa iris (hoa Diên Vĩ) với các tham số theo thứ tự:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| i | Chiều dài đài hoa | Chiều rộng đài hoa | Chiều dài cánh hoa | Chiều rộng cánh hoa | Loại hoa |
| 1 | 5.4 | 3.9 | 1.3 | 0.4 | Iris |
| 2 | 5.1 | 3.5 | 1.4 | 0.3 | Iris |
| 3 | 5.7 | 3.8 | 1.7 | 0.3 | Iris |
| 4 | 4.9 | 2.4 | 3.3 | 1.0 | Iris |
| 5 | 6.6 | 2.9 | 4.6 | 1.3 | Iris |
| 6 | 5.2 | 2.7 | 3.9 | 1.4 | Iris |
| 7 | 5.0 | 2.0 | 3.5 | 1.0 | Iris |
| 8 | 5.1 | 3.8 | 1.5 | 0.3 | Iris |
| 9 | 5.4 | 3.4 | 1.7 | 0.2 | Iris |
| 10 | 5.1 | 3.7 | 1.5 | 0.4 | Iris |
| 11 | 4.6 | 3.6 | 1.0 | 0.2 | Iris |
| 12 | 5.1 | 3.3 | 1.7 | 0.5 | Iris |
| 13 | 7.6 | 3.0 | 6.6 | 2.1 | Iris |
| 14 | 4.9 | 2.5 | 4.5 | 1.7 | Iris |
| 15 | 6.1 | 2.9 | 4.7 | 1.4 | Iris |
| 16 | 7.3 | 2.9 | 6.3 | 1.8 | Iris |
|  | **5.1** | **3.7** | **1.5** | **0.4** | **Iris-Setosa** |
|  | **6.1** | **2.9** | **4.7** | **1.4** | **Iris-Versicolor** |
|  | **6.7** | **2.5** | **5.8** | **1.8** | **Iris-Virginica** |

1. Dùng thuật toán K-Trung bình phân tính khoảng cách và gán mã cho 3 cụm**:**
2. Tính lại tọa độ cụm bằng cách lấy trung bình.

**Câu 3** *(3 điểm):*

1. Thế nào là hồi quy? Trình bày thuật toán hồi quy tuyến tính?
2. Tại sao người ta gọi thuật toán k-NN là thuật toán lười? “k” có ý nghĩa gi? Cách chọn “k”?
3. Lề (Margin) trong thuật toán SVM là gì? Vẽ sơ đồ minh họa và viết công thức đồng thời giải thích ý nghĩa của các biến, tham số trong công thức đó.

*Thí sinh không được sử dụng tài liệu*

*………, ngày ……. tháng …… năm 2022*

**NGƯỜI DUYỆT ĐỀ NGƯỜI RA ĐỀ**

**Nguyễn Quang Hoan**

**HƯỚNG DẪN CHẤM ĐỀ MÃ SỐ IT60802203**

**HỌC PHẦN HỌC MÁY HỆ ĐẠI HỌC NGÀNH KHOA HỌC MÁY TÍNH**

*Năm học 2021 - 2022*

***Mã số đề:*** IT60802203

|  |  |  |
| --- | --- | --- |
| **Câu/Ý**  **1/1** | **ĐÁP ÁN-HƯỚNG DẪN CHẤM**  Bước 1: Chia theo các đặc tính đầu vào (i = ). Mỗi gốc xác định đúng: 0,25 điểm | **Điểm**  **0.25\*5=1.25** |

|  |  |  |
| --- | --- | --- |
| **Nắng** | **U ám** | **Mưa** |
| D1  D2  D9 | D3  D5  D7 | D4  D6  D8  D10 |

|  |  |  |
| --- | --- | --- |
| **Cao** | **Trung bình** | **Thấp** |
| D1  D2  D3  D5 | D4  D8  D10 | D6  D7  D9 |

|  |  |
| --- | --- |
| **Cao** | **Bình thường** |
| D1  D2  D3  D4  D8 | D5  D6  D7  D9  D10 |

|  |  |
| --- | --- |
| **Mạnh** | **Yếu** |
| D2  D6  D7  D8 | D1  ~~D3~~  ~~D4~~  ~~D5~~  ~~D9~~  ~~D10~~ |

|  |  |
| --- | --- |
| **Có** | **Không** |
| D3  D4  D5  D7  D9  D10 | D1  D2  D6  D8 |

Chơi Tennis

.Trời

.Nhiệt độ

.Độ ẩm

.Gió

|  |  |  |
| --- | --- | --- |
| **Câu/Ý**  **1/2** | **ĐÁP ÁN-HƯỚNG DẪN CHẤM**  Bước 2: Xác định công thức để chọn gốc = cho bốn gốc, mỗi gốc: 0.25 điểm khi xác định đúng. | **Điểm**  **0.25\*4=1** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | b |  |  | Tỷ số |
|  | 1. Nắng |  |  |  |
| 2. U ám |  |  |
| 3. Mưa |  |  |
|  | 1. Cao |  |  |  |
| 2. Trung bình |  |  |
| 3. Thấp |  |  |
|  | 1. Cao |  |  |  |
| 2. Bình thường |  |  |
|  | 1. Mạnh |  |  |  |
| 2. Yếu |  |  |

|  |  |  |
| --- | --- | --- |
| **Câu/Ý**  **1/3** | **ĐÁP ÁN-HƯỚNG DẪN CHẤM**  Chọn gốc (Max) và Bước 3: Viết luật | **Điểm**  **0.25** |

Luật 1:

|  |  |  |
| --- | --- | --- |
| **Câu/Ý**  **1/4** | **ĐÁP ÁN-HƯỚNG DẪN CHẤM**  **Các bước tiếp theo gồm:**  Bước 4: Tổ hợp chập 2 đặc trưng cho  **.**Trời=”Nắng”;  **.**Trời=”Mưa”;  Bước 5: Lặp lại như Bước 2 với  **.**Trời=”Nắng”;  **.**Trời=”Mưa”;  Bước 6: Lặp lại như Bước 3 với  **.**Trời=”Nắng”;  **.**Trời=”Mưa”; Viết các luật  Điều kiện dừng khi mọi dữ liệu (bản ghi hay ví dụ) đã bao hết trong các luật | **Điểm**  **0.25**  **0.25**  **0.25**  **0.25** |

**Câu 2:**

|  |  |  |
| --- | --- | --- |
| **Câu/Ý**  **2/1** | **ĐÁP ÁN-HƯỚNG DẪN CHẤM**  Đặt:Chiều dài đài hoa:A1, Chiều rộng đài hoa: A2, Chiều dài cánh hoa; A3; Chiều rộng cánh hoa A4 và lập bảng có cấu trúc như bảng dưới | **Điểm**  **0.25** |
| **2/2** | Tính khoảng cách ơclid ứng với A1, A2, A3, A4 cho ba cum | **0.5\*3=1.5** |
| **2/3** | Chọn để gán cụm cho cột “Nhóm gần nhất” (Xem bảng) | **0.25** |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **i** |  |  |  |  |  |  |  | **Nhóm gần nhất** |
| 1 | 5.4 | 3.9 | 1.3 | 0.4 | 0.41 | 3.75 | 5.08 | **1** |
| 2 | 5.1 | 3.5 | 1.4 | 0.3 | 0.24 | 3.66 | 5.02 | **1** |
| 3 | 5.7 | 3.8 | 1.7 | 0.3 | 0.65 | 3.34 | 4.66 | **1** |
| 4 | 4.9 | 2.4 | 3.3 | 1.0 | 2,31 | 1.95 | 3.18 | **2** |
| 5 | 6.6 | 2.9 | 4.6 | 1.3 | 3.65 | 0.52 | 1.36 | **2** |
| 6 | 5.2 | 2.7 | 3.9 | 1.4 | 2.79 | 1.3 | 2.46 | **2** |
| 7 | 5.0 | 2.0 | 3.5 | 1.0 | 2.69 | 1.90 | 3.01 | **2** |
| 8 | 5.1 | 3.8 | 1.5 | 0.3 | 0.14 | 3.64 | 4.99 | **1** |
| 9 | 5.4 | 3.4 | 1.7 | 0.2 | 0.51 | 3.34 | 4.68 | **1** |
| 10 | 5.1 | 3.7 | 1.5 | 0.4 | 0 | 3.59 | 4.94 | **1** |
| 11 | 4.6 | 3.6 | 1.0 | 0.2 | 0,74 | 4.23 | 5.64 | **1** |
| 12 | 5.1 | 3.3 | 1.7 | 0.5 | 0.44 | 3.31 | 4.66 | **1** |
| 13 | 7.6 | 3.0 | 6.6 | 2.1 | 5,96 | 2,52 | 1.34 | **3** |
| 14 | 4.9 | 2.5 | 4.5 | 1.7 | 3,49 | 1.32 | 2.22 | **2** |
| 15 | 6.1 | 2.9 | 4.7 | 1.4 | 3,59 | 0 | 1.37 | **2** |
| 16 | 7.3 | 2.9 | 6.3 | 1.8 | 5.52 | 2,04 | 0.88 | **3** |
|  | **5.1** | **3.7** | **1.5** | **0.4** |  |  |  |  |
|  | **6.1** | **2.9** | **4.7** | **1.4** |  |  |  |  |
|  | **6.7** | **2.5** | **5.8** | **1.8** |  |  |  |  |

|  |  |  |
| --- | --- | --- |
| **Câu/Ý**  **2/4** | **ĐÁP ÁN-HƯỚNG DẪN CHẤM**  **Tính lại tọa độ cụm bằng cách lấy trung bình như dưới đây cho ba cụm: 0.5\*3** | **Điểm**  **0.5\*3=1.5** |

|  |  |  |
| --- | --- | --- |
| **Câu 3** |  |  |
| **3/1** | * Viết khái niệm * Nêu đinh nghĩa hồi quy * Trình bày thuật toán hồi quy tuyến tính? | ***0,33***  ***0,33***  ***0,33*** |
| **3/2** | -Thuật toán k-NN là thuật toán lười vì thuật toán đó không tạo ra mô hình học máy (Learning Model) và hàm đích (Target Function) nên thuật toán xếp loại đơn giản (Simple Learning) hay học lười (Lazy Learning).  - “k” có ý nghĩa là một số mẫu quan sát được chọn từ các mẫu ví dụ có khoảng cách gần nhấy với mẫu cần xác định tính chất.  - Cách chọn “*k*”. Vì *k* mẫu giống như số phiếu bầu chọn (thường là + hoặc – hay Yes hoặc No giống phương châm bỏ phiếu bình chọn, do đó thường chọn số lẻ | ***0,33***  ***0,33***  ***0,33*** |
| **3/3** | * Lề (Margin) trong thuật toán SVM là khoảng cách (*Đinh nghìa lề.* Lề *(Margin)* là khoảng cách giữa hai siêu phẳng hỗ trợ song song. * Sơ đồ minh họa:      * Công thức (sinh viên giải thích thêm các tham số, khái niệm chuẩn 2:     trong đó, là độ lớn của véc tơ trọng số ***W*** theo chuẩn 2 *(2-norm*). | ***0,33***  ***0,33***  ***0,33*** |

*………, ngày ……. tháng …… năm 2022*

**NGƯỜI DUYỆT ĐỀ NGƯỜI RA ĐỀ**

**Nguyễn Quang Hoan**

|  |  |
| --- | --- |
| TRƯỜNG ĐẠI HỌC HẠ LONG  **KHOA CÔNG NGHỆ THÔNG TIN**  **Mã đề:** IT60802204  (04 *là số thứ tự)* | KỲ THI KẾT THÚC HỌC PHẦN NĂM HỌC 2021–2022  ĐỀ THI HỌC PHẦN: **Học máy**  LỚP KHOA HỌC MÁY TÍNH  *(Thời gian làm bài: 90 phút)* |

**Câu 1. (***3,5 điểm)***.** Cho tập cơ sở dữ liệu thời tiết dưới đây

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Ngày** | **Trời** | **Nhiệt độ** | **Độ ẩm** | **Gió** | **Chơi tennis** |
| D1 | Nắng | Cao | Cao | Yếu | Không |
| D2 | Nắng | Cao | Cao | Mạnh | Không |
| D3 | U ám | Cao | Cao | Yếu | Có |
| D4 | Mưa | Trung bình | Cao | Yếu | Có |
| D5 | Mưa | Thấp | Bình thường | Yếu | Có |
| D6 | Mưa | Thấp | Bình thường | Mạnh | Không |
| D7 | U ám | Thấp | Bình thường | Mạnh | Có |
| D8 | Nắng | Trung bình | Cao | Yếu | Không |

Dùng thuật toán Bayes dự báo trường hợp sau:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Trời** | **Nhiệt độ** | **Độ ẩm** | **Gió** | **Chơi tennis** |
| D1 | U ám | Thấp | Cao | Yếu | ? |

**Câu 2: (3,5 điểm)** Cho tập dữ liệu về hoa iris (hoa Diên Vĩ) với các tham số theo thứ tự:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| i | Chiều dài đài hoa | Chiều rộng đài hoa | Chiều dài cánh hoa | Chiều rộng cánh hoa | Loại hoa |
| 1 | 5.4 | 3.9 | 1.3 | 0.4 | Iris |
| 2 | 5.1 | 3.5 | 1.4 | 0.3 | Iris |
| 3 | 5.7 | 3.8 | 1.7 | 0.3 | Iris |
| 4 | 4.9 | 2.4 | 3.3 | 1.0 | Iris |
| 5 | 6.6 | 2.9 | 4.6 | 1.3 | Iris |
| 6 | 5.2 | 2.7 | 3.9 | 1.4 | Iris |
| 7 | 5.0 | 2.0 | 3.5 | 1.0 | Iris |
| 8 | 5.1 | 3.8 | 1.5 | 0.3 | Iris |
| 9 | 5.4 | 3.4 | 1.7 | 0.2 | Iris |
| 10 | 5.1 | 3.7 | 1.5 | 0.4 | Iris |
| 11 | 4.6 | 3.6 | 1.0 | 0.2 | Iris |
| 12 | 5.1 | 3.3 | 1.7 | 0.5 | Iris |
| 13 | 7.6 | 3.0 | 6.6 | 2.1 | Iris |
| 14 | 4.9 | 2.5 | 4.5 | 1.7 | Iris |
| 15 | 6.1 | 2.9 | 4.7 | 1.4 | Iris |
| 16 | 7.3 | 2.9 | 6.3 | 1.8 | Iris |
|  | **5.1** | **3.7** | **1.5** | **0.4** | **Iris-Setosa** |
|  | **6.1** | **2.9** | **4.7** | **1.4** | **Iris-Versicolor** |
|  | **6.7** | **2.5** | **5.8** | **1.8** | **Iris-Virginica** |

1. Dùng thuật toán K-Trung bình phân tính khoảng cách và gán mã cho 3 cụm**:**
2. Tính lại tọa độ cụm bằng cách lấy trung bình.

**Câu 3** *(3 điểm):*

a. Thế nào là hồi quy? Trình bày thuật toán hồi quy tuyến tính?

b. Tại sao người ta gọi thuật toán k-NN là thuật toán lười? “k” có ý nghĩa gi? Cách chọn “k”?

1. Lề (Margin) trong thuật toán SVM là gì? Vẽ sơ đồ minh họa và viết công thức đồng thời giải thích ý nghĩa của các biến, tham số trong công thức đó.

*Thí sinh không được sử dụng tài liệu*

*………, ngày ……. tháng …… năm 2022*

**NGƯỜI DUYỆT ĐỀ NGƯỜI RA ĐỀ**

**Nguyễn Quang Hoan**

**HƯỚNG DẪN CHẤM ĐỀ MÃ SỐ IT60802204**

**HỌC PHẦN HỌC MÁY HỆ ĐẠI HỌC NGÀNH KHOA HỌC MÁY TÍNH**

*Năm học 2021-2022*

***Mã số đề:*** IT60802204

|  |  |  |
| --- | --- | --- |
| **Câu/Ý** | **ĐÁP ÁN-HƯỚNG DẪN CHẤM** | **Điểm** |
| **1/1** | Gọi bốn đặc trưng đầu vào là A1=”Trời”; A2=” Nhiệt độ”; A3=”Độ ẩm”,, A4=”Gió”; B=”Chơi Tennis”.  Bước 1. Sử dụng phương pháp “chía để trị” theo 4 đặc trưng đầu vào như bảng sau   |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | **Trời** | | | **Nhiệt độ** | | | **Độ ẩm** | | | **Gió** | | | **Chơi tennis** | | | Không | | Có | Không | | Có | Không | | Có | Không | | Có | Không | Có | | Nắng | 3 | 0 | Cao | 2 | 1 | Cao | 3 | 2 | Yếu | 2 | 3 | 4 | 4 | | Mưa | 0 | 2 | TB | 1 | 1 | BT | 1 | 2 | Mạnh | 2 | 1 | | U ám | 1 | 2 | Thấp | 1 | 2 |  | | |  | | | | A1 | | | A2 | | | A3 | | | A4 | | | B | | | A | | | | | | | | | | | | | ***0.2\*5***  ***=1*** |
| **1/2** | Bước 2. Tính xác suất tương ứng theo 4 đặc trưng đầu vào như bảng sau   |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | **Trời** | | | **Nhiệt độ** | | | **Độ ẩm** | | | **Gió** | | | **Chơi tennis** | | | Không | | Có | Không | | Có | Không | | Có | Không | | Có | Không | Có | | Nắng | 3/4 | 0/4 | Cao | 2/4 | 1/4 | Cao | 3/4 | 2/4 | Yếu | 2/4 | 3/4 | 4/8 | 4/8 | | Mưa | 0/4 | 2/4 | TB | 1/4 | 1/4 | BT | 1/4 | 2/4 | Mạnh | 2/4 | 1/4 | | U ám | 1/4 | 2/4 | Thấp | 1/4 | 2/4 |  | | |  | | | | A1 | | | A2 | | | A3 | | | A4 | | | B | | | A | | | | | | | | | | | | | ***0.2\*5***  ***=1*** |
| **1/3** | **Giả sử**  là độc lập từng đôi một, ta có thể viết: | ***0.25*** |
| **1/4** | *Trường hợp B =không. Thay vào (\*) và lấy các giá trị cần dự báo có:*  (\*) = P(A1­=U ám |B=Không).P(A2 =Thấp|B=Không).P(A3 =Cao|B=Không).P(A4 =Yếu|B= Không) = = 0 = P(B|A) | ***0.5*** |
| **1/5** | *Ngược lại: B = Có*  (\*) = P(A­=U ám|B=Có).P(A2 ­=Thấp|B=Có).P(A3 ­=Cao|B=Có).P(A4=Yếu|B=Có)  = = = P(|A) | ***0.5*** |
| **1/6** | Vậy: *P(B|A) > P(|A) B=Có; Có đi chơi Tennis.* | ***0.25*** |
| **Câu 3** |  |  |
| **3/1** | * Viết khái niệm * Nêu đinh nghĩa hồi quy * Trình bày thuật toán hồi quy tuyến tính? | ***0,33***  ***0,33***  ***0,33*** |
| **3/2** | -Thuật toán k-NN là thuật toán lười vì thuật toán đó không tạo ra mô hình học máy (Learning Model) và hàm đích (Target Function) nên thuật toán xếp loại đơn giản (Simple Learning) hay học lười (Lazy Learning).  - “k” có ý nghĩa là một số mẫu quan sát được chọn từ các mẫu ví dụ có khoảng cách gần nhấy với mẫu cần xác định tính chất.  - Cách chọn “*k*”. Vì *k* mẫu giống như số phiếu bầu chọn (thường là + hoặc – hay Yes hoặc No giống phương châm bỏ phiếu bình chọn, do đó thường chọn số lẻ | ***0,33***  ***0,33***  ***0,33*** |
| **3/3** | Lề (Margin) trong thuật toán SVM là khoảng cách (*Đinh nghìa lề.* Lề *(Margin)* là khoảng cách giữa hai siêu phẳng hỗ trợ song song   * Sơ đồ minh họa:      * Công thức (sinh viên giải thích thêm các tham số, khái niệm chuẩn 2:     trong đó, là độ lớn của véc tơ trọng số ***W*** theo chuẩn 2 *(2-norm*). | ***0,33***  ***0,33***  ***0,33*** |

**Câu 2:**

|  |  |  |
| --- | --- | --- |
| **Câu/Ý**  **2/1** | **ĐÁP ÁN-HƯỚNG DẪN CHẤM**  Đặt:Chiều dài đài hoa:A1, Chiều rộng đài hoa: A2, Chiều dài cánh hoa; A3; Chiều rộng cánh hoa A4 và lập bảng có cấu trúc như bảng dưới | **Điểm**  **0.25** |
| **2/2** | Tính khoảng cách ơclid ứng với A1, A2, A3, A4 cho ba cum | **0.5\*3=1.5** |
| **2/3** | Chọn để gán cụm cho cột “Nhóm gần nhất” (Xem bảng) | **0.25** |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **i** |  |  |  |  |  |  |  | **Nhóm gần nhất** |
| 1 | 5.4 | 3.9 | 1.3 | 0.4 | 0.41 | 3.75 | 5.08 | **1** |
| 2 | 5.1 | 3.5 | 1.4 | 0.3 | 0.24 | 3.66 | 5.02 | **1** |
| 3 | 5.7 | 3.8 | 1.7 | 0.3 | 0.65 | 3.34 | 4.66 | **1** |
| 4 | 4.9 | 2.4 | 3.3 | 1.0 | 2,31 | 1.95 | 3.18 | **2** |
| 5 | 6.6 | 2.9 | 4.6 | 1.3 | 3.65 | 0.52 | 1.36 | **2** |
| 6 | 5.2 | 2.7 | 3.9 | 1.4 | 2.79 | 1.3 | 2.46 | **2** |
| 7 | 5.0 | 2.0 | 3.5 | 1.0 | 2.69 | 1.90 | 3.01 | **2** |
| 8 | 5.1 | 3.8 | 1.5 | 0.3 | 0.14 | 3.64 | 4.99 | **1** |
| 9 | 5.4 | 3.4 | 1.7 | 0.2 | 0.51 | 3.34 | 4.68 | **1** |
| 10 | 5.1 | 3.7 | 1.5 | 0.4 | 0 | 3.59 | 4.94 | **1** |
| 11 | 4.6 | 3.6 | 1.0 | 0.2 | 0,74 | 4.23 | 5.64 | **1** |
| 12 | 5.1 | 3.3 | 1.7 | 0.5 | 0.44 | 3.31 | 4.66 | **1** |
| 13 | 7.6 | 3.0 | 6.6 | 2.1 | 5,96 | 2,52 | 1.34 | **3** |
| 14 | 4.9 | 2.5 | 4.5 | 1.7 | 3,49 | 1.32 | 2.22 | **2** |
| 15 | 6.1 | 2.9 | 4.7 | 1.4 | 3,59 | 0 | 1.37 | **2** |
| 16 | 7.3 | 2.9 | 6.3 | 1.8 | 5.52 | 2,04 | 0.88 | **3** |
|  | **5.1** | **3.7** | **1.5** | **0.4** |  |  |  |  |
|  | **6.1** | **2.9** | **4.7** | **1.4** |  |  |  |  |
|  | **6.7** | **2.5** | **5.8** | **1.8** |  |  |  |  |

|  |  |  |
| --- | --- | --- |
| **Câu/Ý**  **2/4** | **ĐÁP ÁN-HƯỚNG DẪN CHẤM**  Tính lại tọa độ cụm bằng cách lấy trung bình như dưới đây cho ba cụm: 0.5\*3=1,5 điểm | **Điểm**  **0.5\*3=1.5** |

*………, ngày ……. tháng …… năm 2022*

**NGƯỜI DUYỆT ĐỀ NGƯỜI RA ĐỀ**

**Nguyễn Quang Hoan**