**35. Phạm Tuân**

Câu 1. Cho tập dữ liệu về hoa iris (Hoa Diên Vĩ hoặc Hoa Loa kèn) với các tham số theo thứ tự: chiều dài đài hoa, chiều rộng đài hoa, chiều dài cánh hoa, chiều rông cánh hoa, loại hoa.

4.9,2.5,4.5,1.7,Iris-virginica

7.3,2.9,6.3,1.8,Iris-virginica

6.7,2.5,5.8,1.8,Iris-virginica

7.2,3.6,6.1,2.5,Iris-virginica

6.5,3.2,5.1,2.0,Iris-virginica

6.4,2.7,5.3,1.9,Iris-virginica

6.8,3.0,5.5,2.1,Iris-virginica

4.8,3.1,1.6,0.2,Iris-setosa

5.4,3.4,1.5,0.4,Iris-setosa

5.2,4.1,1.5,0.1,Iris-setosa

5.5,4.2,1.4,0.2,Iris-setosa

4.9,3.1,1.5,0.1,Iris-setosa

5.0,3.2,1.2,0.2,Iris-setosa

5.5,3.5,1.3,0.2,Iris-setosa

**Với: 5.3, 3.5, 5.1, 0.3, ?-? (dự báo là hoa gì?) sử dụng K-nn với k=5.**

Cho tập dữ liệu về hoa iris (Hoa Diên Vĩ hoặc Hoa Loa kèn) với các tham số theo thứ tự: chiều dài đài hoa, chiều rộng đài hoa, chiều dài cánh hoa, chiều rông cánh hoa, loại hoa.

4.9,3.1,1.5,0.1,Iris-setosa

5.0,3.2,1.2,0.2,Iris-setosa

5.5,3.5,1.3,0.2,Iris-setosa

4.9,3.1,1.5,0.1,Iris-setosa

4.4,3.0,1.3,0.2,Iris-setosa

5.0,3.3,1.4,0.2,Iris-setosa

5.8,2.6,4.0,1.2,Iris-versicolor

5.0,2.3,3.3,1.0,Iris-versicolor

7.0,3.2,4.7,1.4,Iris-versicolor

6.5,2.8,4.6,1.5,Iris-versicolor

6.3,3.3,4.7,1.6,Iris-versicolor

6.0,2.7,5.1,1.6,Iris-versicolor

5.4,3.0,4.5,1.5,Iris-versicolor

Với: 5.4, 8.2, 6.5, 1.1, ?-? (dự báo là hoa gì?) sử dụng K-nn với k=5.Trích dẫn từ bộ dữ liệu hoa Iris, ta lập ra 1 bảng dữ liệu nhỏ và tiến hành thực hiện theo thuật toán K-NN:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Chiều dài đài hoa** | **Chiều rộng đài hoa** | **Chiều dài cánh hoa** | **Chiều rông cánh hoa** | **Loại hoa** |
| 4,9 | 2,5 | 4,5 | 1,7 | Iris-virginica |
| 7,3 | 2,9 | 6,3 | 1,8 | Iris-virginica |
| 6,7 | 2,5 | 5,8 | 1,8 | Iris-virginica |
| 7,2 | 3,6 | 6,1 | 2,5 | Iris-virginica |
| 6,5 | 3,2 | 5,1 | 2,0 | Iris-virginica |
| 6,4 | 2,7 | 5,3 | 1,9 | Iris-virginica |
| 6,8 | 3,0 | 5,5 | 2,1 | Iris-virginica |
| 4,8 | 3,1 | 1,6 | 0,2 | Iris-setosa |
| 5,4 | 3,4 | 1,5 | 0,4 | Iris-setosa |
| 5,2 | 4,1 | 1,5 | 0,1 | Iris-setosa |
| 5,5 | 4,2 | 1,4 | 0,2 | Iris-setosa |
| 4,9 | 3,1 | 1,5 | 0,1 | Iris-setosa |
| 5,0 | 3,2 | 1,2 | 0,2 | Iris-setosa |
| 5,5 | 3,5 | 1,3 | 0,2 | Iris-setosa |
| 5,3 | 3,5 | 5,1 | 0,3 | ? hoa gì |

Ta tiến hành thường hóa dữ liệu, bảng dữ liệu sẽ bị biến đổi thành bảng sau:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Chiều dài đài hoa** | **Chiều rộng đài hoa** | **Chiều dài cánh hoa** | **Chiều rông cánh hoa** | **Loại hoa** |
| 0,67 | 0,60 | 0,71 | 0,68 | Iris-virginica |
| 1,00 | 0,69 | 1,00 | 0,72 | Iris-virginica |
| 0,92 | 0,60 | 0,92 | 0,72 | Iris-virginica |
| 0,99 | 0,86 | 0,97 | 1,00 | Iris-virginica |
| 0,89 | 0,76 | 0,81 | 0,80 | Iris-virginica |
| 0,88 | 0,64 | 0,84 | 0,76 | Iris-virginica |
| 0,93 | 0,71 | 0,87 | 0,84 | Iris-virginica |
| 0,66 | 0,74 | 0,25 | 0,08 | Iris-setosa |
| 0,74 | 0,81 | 0,24 | 0,16 | Iris-setosa |
| 0,71 | 0,98 | 0,24 | 0,04 | Iris-setosa |
| 0,75 | 1,00 | 0,22 | 0,08 | Iris-setosa |
| 0,67 | 0,74 | 0,24 | 0,04 | Iris-setosa |
| 0,68 | 0,76 | 0,19 | 0,08 | Iris-setosa |
| 0,75 | 0,83 | 0,21 | 0,08 | Iris-setosa |
| 0,73 | 0,83 | 0,81 | 0,12 | ? hoa gì |

Ta tính toán khoảng cách Euclide được bảng khoảng cách tương ứng :

|  |  |
| --- | --- |
| **Tên loài hoa** | **Khoảng cách** |
| Iris-virginica | 0,6194 |
| Iris-virginica | 0,6997 |
| Iris-virginica | 0,6820 |
| Iris-virginica | 0,9316 |
| Iris-virginica | 0,7022 |
| Iris-virginica | 0,6851 |
| Iris-virginica | 0,7620 |
| Iris-setosa | 0,5694 |
| Iris-setosa | 0,5734 |
| Iris-setosa | 0,5945 |
| Iris-setosa | 0,6125 |
| Iris-setosa | 0,5869 |
| Iris-setosa | 0,6257 |
| Iris-setosa | 0,6052 |

Với K = 5 hàng xóm gần nhất là 5 ứng với 5 khoảng cách nhỏ nhất lần lượt là : 0,5694

0,5734

0,5945

0,5869

0,6052

tương ứng với tên loài: Iris-setosa

. Như vậy ta kết luận loài cần phân lớp thuộc loài Iris-setosa

.

**Chiều rộng đài**