

**Iris flower dataset** là một bộ dữ liệu nhỏ. Bộ dữ liệu này bao gồm thông tin của ba loại hoa Iris (một loài hoa lan) khác nhau: **Iris setosa**, **Iris virginica** và **Iris versicolor**. Mỗi loại có 50 bông hoa được đo với dữ liệu là 4 thông tin: chiều dài, chiều rộng đài hoa (sepal), và chiều dài, chiều rộng cánh hoa (petal). Dưới đây là ví dụ về hình ảnh của ba loại hoa.



Iris setosa



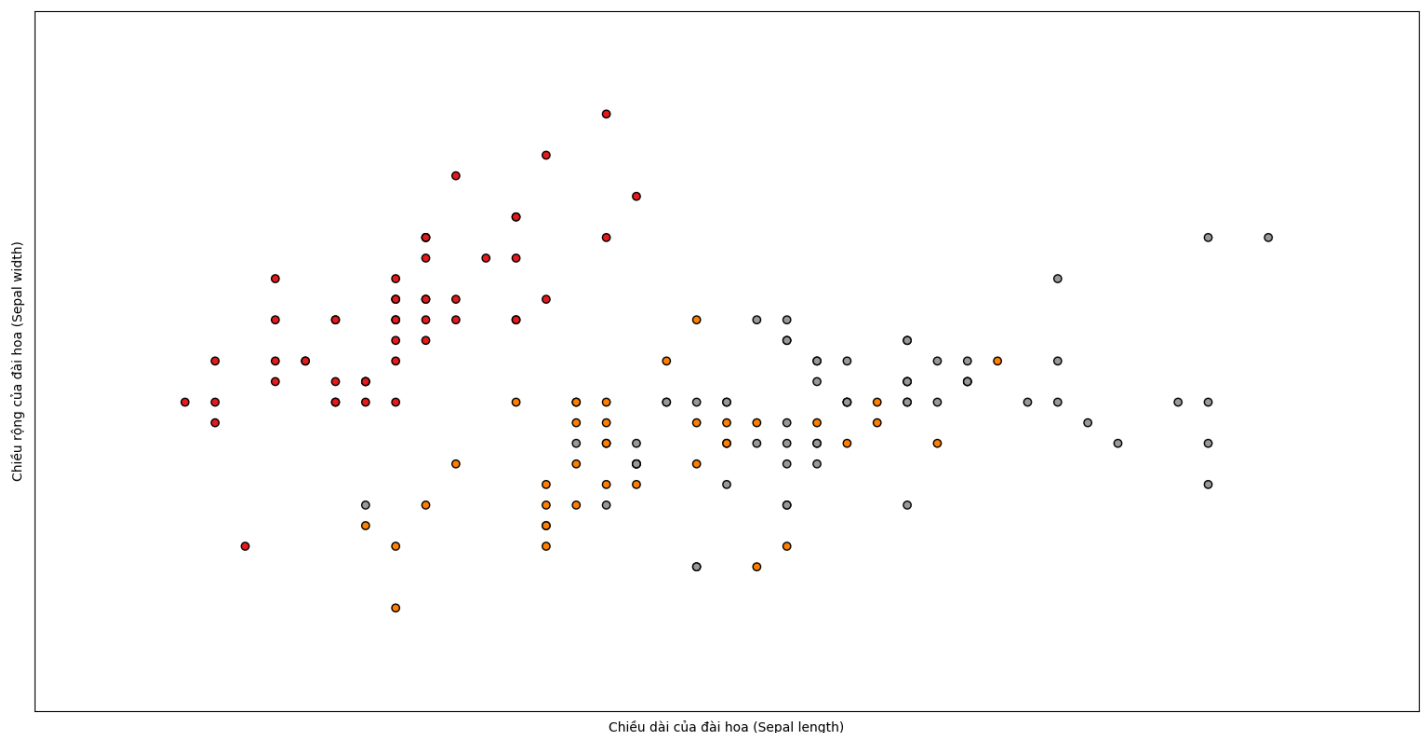
Iris versicolor



Iris virginica

Trong bài tập này, để đơn giản hóa, chúng ta chỉ lấy 2 thông tin đầu tiên của mỗi loại hoa: Chiều dài, chiều rộng của đài hoa (sepal).

Biểu diễn chiều dài, chiều rộng đài hoa trên không gian 2 chiều (chiều dài là hoành độ, chiều rộng là tung độ)



Từ các dữ liệu có sẵn, xây dựng 1 bộ phân loại (Object classifier) cho 3 loại hoa này, theo nguyên tắc sau

### **Bài 1a: Áp dụng với các bạn chưa/ít kinh nghiệm về lập trình**

Tính trung bình cộng chiều dài, chiều rộng đài hoa của từng loại (3 loại hoa)

### **Bài 2a: Áp dụng với các bạn chưa/ít kinh nghiệm về lập trình**

Xác định chiều dài, chiều rộng đài hoa lớn nhất của từng loại (3 loại hoa)

### **Bài 3a: Áp dụng với các bạn chưa/ít kinh nghiệm về lập trình**

Nhập input đầu vào là 2 thông số, chiều dài & chiều rộng của 1 bông hoa bất kỳ (theo đơn vị mm). Xác định xem bông hoa đấy thuộc loại hoa nào (**Iris setosa**, **Iris virginica** hay **Iris versicolor**). Loại của bông hoa được xác định là class của

- a) Điểm có tọa độ gần nó nhất
- b) Trung bình cộng nhỏ nhất đến tất cả các điểm của từng loại

### **Bài 1b: Áp dụng cho các bạn đã có kinh nghiệm về lập trình**

In ra chiều dài đài hoa lớn nhất của 10 bông hoa mỗi loại

### **Bài 2b: Áp dụng cho các bạn đã có kinh nghiệm về lập trình**

Thống kê trong 50 bông hoa có tổng chiều dài + chiều rộng đài hoa lớn nhất, bao nhiêu bông thuộc về từng loại

### **Bài 3b: Áp dụng cho các bạn đã có kinh nghiệm về lập trình**

Nhập input đầu vào là 2 thông số, chiều dài & chiều rộng của 1 bông hoa bất kỳ (theo đơn vị mm). Xác định xem bông hoa đấy thuộc loại hoa nào (**Iris setosa**, **Iris virginica** hay **Iris versicolor**). Loại của bông hoa được xác định là class của

- a) Điểm có tọa độ gần nó nhất. Nếu có 2 điểm thuộc 2 loại hoa khác nhau cùng nằm gần nhất với input, thì xác định sang điểm kế tiếp
- b) 7 điểm gần nó nhất (Vote đa số). Nếu không xác định được, lấy tiếp điểm thứ 8,9... cho đến khi xác định được class

- c) Lấy tọa độ của input là tâm của 1 đường tròn, có bán kính 2 mm. Vote đa số dựa trên tất cả các điểm nằm trên và trong đường tròn. Nếu có 2 class có số điểm bằng nhau, xét đến trung bình cộng khoảng cách của input đến tất cả các điểm trong vòng tròn của mỗi class.