

# Bài tập KNN

## Sinh viên làm bài trên giấy, và nộp lại cho giảng viên sau 15 phút. Trên giấy làm bài sinh viên ghi rõ họ tên và mã số sv

- Cho tập dữ liệu như hình bên. Người ta mong muốn xây dựng một phần mềm có khả năng dự báo thiếu cân của một người nào đó dựa trên chiều cao và cân nặng. Anh/chị hãy xác định loại bài toán và hãy giải thích.
- Anh chị hãy nhận xét về sự tương quan giữa chiều cao và cân nặng trong tập dữ liệu này.
- Hãy áp dụng giải thuật knn với  $k=3$ , dự báo cho trường hợp anh Nguyễn Văn A cao 170 cm, cân nặng 57kg. Khi áp dụng KNN, dữ liệu này có cần chuẩn hóa hay không? Tại sao?

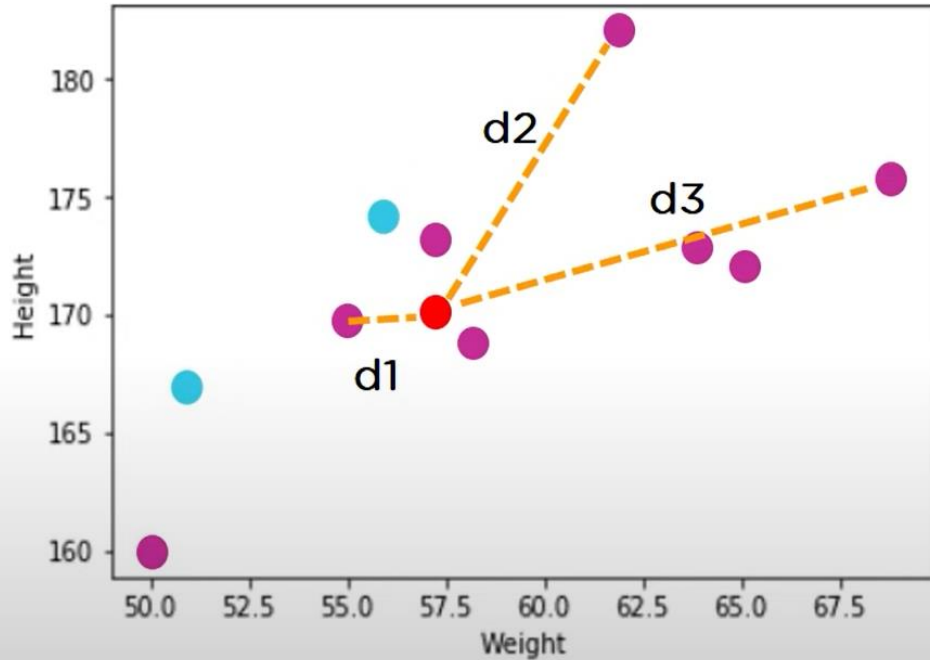
Weight(x2)	Height(y2)	Class
51	167	Underweight
62	182	Normal
69	176	Normal
64	173	Normal
65	172	Normal
56	174	Underweight
58	169	Normal
57	173	Normal
55	170	Normal

Weight(x2)	Height(y2)	Class
51	167	Underweight
62	182	Normal
69	176	Normal
64	173	Normal
65	172	Normal
56	174	Underweight
58	169	Normal
57	173	Normal
55	170	Normal

57 kg	170 cm	?
-------	--------	---



Let's calculate it to understand clearly:



$$\text{dist}(\mathbf{d1}) = \sqrt{(170-167)^2 + (57-51)^2} \approx 6.7$$

$$\text{dist}(\mathbf{d2}) = \sqrt{(170-182)^2 + (57-62)^2} \approx 13$$

$$\text{dist}(\mathbf{d3}) = \sqrt{(170-176)^2 + (57-69)^2} \approx 13.4$$

Similarly, we will calculate Euclidean distance of unknown data point from all the points in the dataset

Now, let's calculate the nearest neighbor at  $k=3$

Weight(x2)	Height(y2)	Class	Euclidean Distance
51	167	Underweight	6.7
62	182	Normal	13
69	176	Normal	13.4
64	173	Normal	7.6
65	172	Normal	8.2
56	174	Underweight	4.1
58	169	Normal	1.4
57	173	Normal	3
55	170	Normal	2



57 kg	170 cm	?
-------	--------	---

Height (in cms)	Weight (in kgs)	T Shirt Size
158	58	M
158	59	M
158	63	M
160	59	M
160	60	M
163	60	M
163	61	M
160	64	L
163	64	L
165	61	L
165	62	L
165	65	L
168	62	L
168	63	L
168	66	L
170	63	L
170	64	L
170	68	L

Nguyễn Văn A, chiều cao = 164, cân nặng 64.  
 Hãy dự đoán size áo cho anh Nguyễn Văn A. Sử dụng giải thuật knn với  $k = 3$ .