KNN Algorithm
For Pattern Recognition

https://www.aiquest.org

Application of KNN

- Classification
- Regression
- Clustering
- Anomaly detection
- Recommendation systems

Note: KNN is considered a lazy learning algorithm because it does not perform an explicit training step but rather memorizes the training data and performs computations at the time of prediction. This allows KNN to quickly adapt to new data and handle dynamic environments since it does not require retraining the model when new data becomes available.

The Minkowski Distance

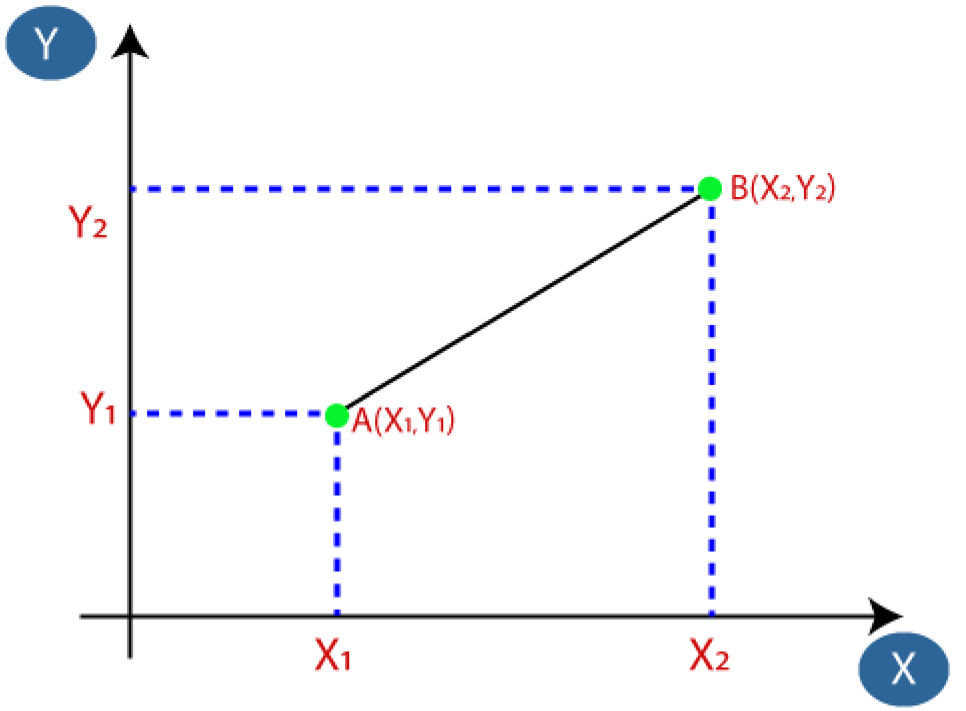
The Minkowski distance between two points, p and q, in a d-dimensional space is defined as:

$$d_M(p, q) = \left(\sum_{i=1}^{d} |x_i^{(p)} - x_i^{(q)}|^r\right)^{\frac{1}{r}}$$

where:

- $x_i^{(p)}$ and $x_i^{(q)}$ are the i-th feature values of points p and q, respectively.
- r is a positive parameter that determines the "degree" of the Minkowski distance. When $r = 1$, the Minkowski distance is equivalent to the Manhattan distance, and when $r = 2$, it is equivalent to the Euclidean distance.

Euclidean Distance Formula



Euclidean Distance between A_1 and B_2 = $\sqrt{(X_2-X_1)^2+(Y_2-Y_1)^2}$

Image: javatpoint.com

Perform K-NN Algorithm (Visually)

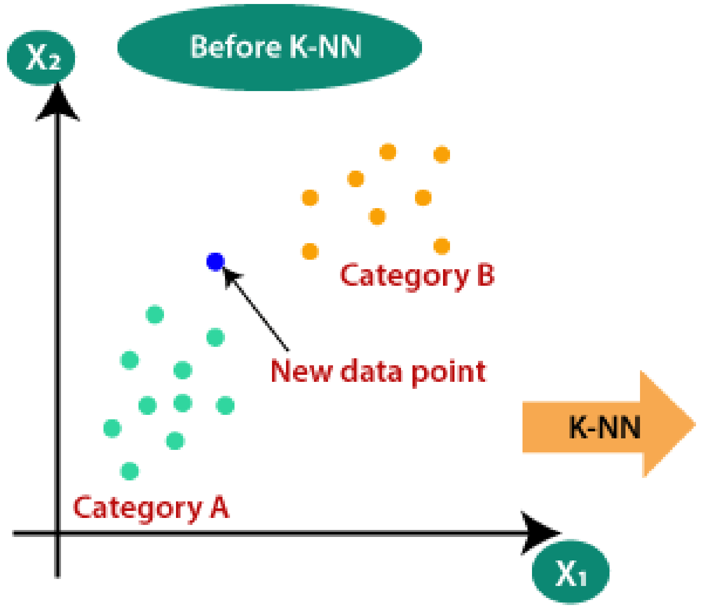


Image: javatpoint.com

Perform K-NN Algorithm (Visually)

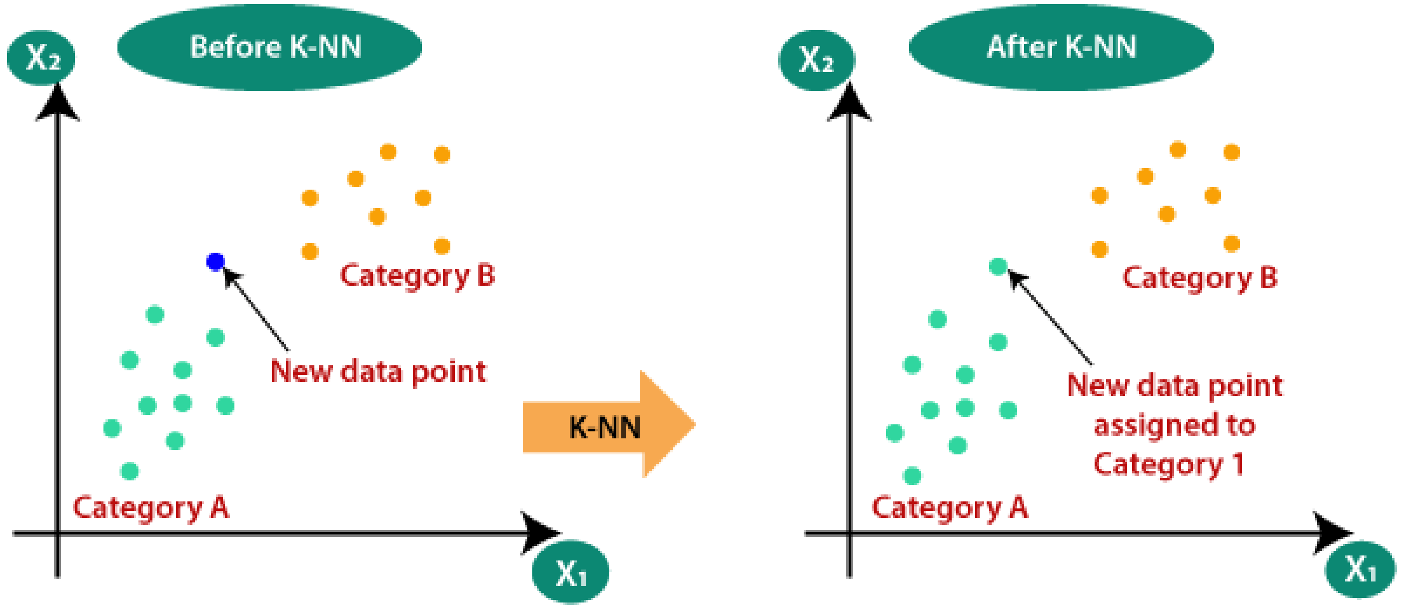


Image: javatpoint.com

Steps: K-NN Algorithm

Explanation of the K-Nearest Neighbors (KNN) algorithm:

1. **Collect the data:** Obtain a labeled dataset with instances and their corresponding class labels.
2. **Choose the value of K:** Determine the number of nearest neighbors to consider.
3. **Preprocess the data if needed:** Perform any necessary data preprocessing steps, such as normalization or handling missing values.
4. **Calculate distances:** Measure the distance between the new instance and all instances in the training dataset using a distance metric, commonly the Euclidean distance.
5. **Find K nearest neighbors:** Select the K instances with the shortest distances to the new instance.
6. **Determine the class:** For classification tasks, assign the class label that appears most frequently among the K nearest neighbors as the predicted class for the new instance.
7. Make predictions: Repeat steps 4 to 6 for all new instances in the test dataset to generate predictions.

That summarizes the core steps of the KNN algorithm. Please note that this simplified explanation omits the steps related to tuning the algorithm's parameters and evaluating its performance.

Perform K-NN for Classification (Visually)

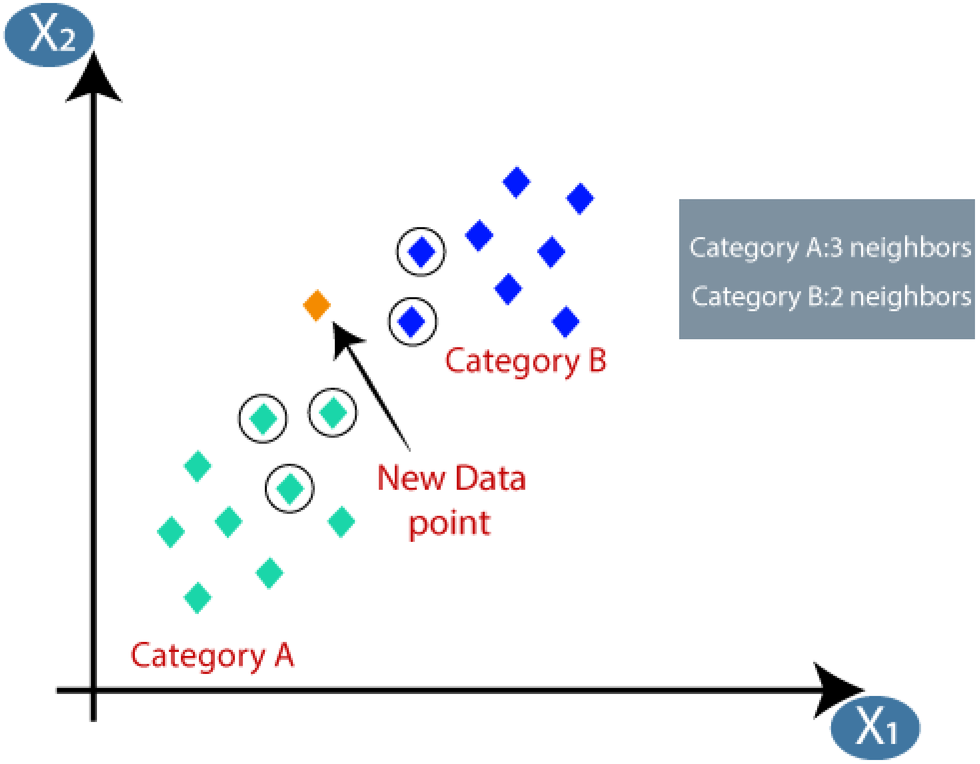
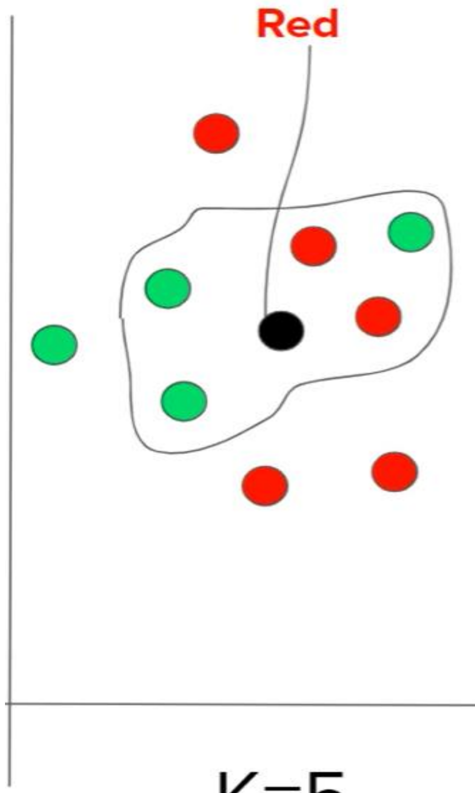


Image: javatpoint.com

K-NN Parameters: weights{‘uniform’, ‘distance’}

Point	Label	Distance	Weight
(x1,y1)	Red	0.2	5
(x2,y2)	Red	0.5	2
(x3,y3)	Green	0.7	1.4
(x4,y4)	Green	1.2	0.8
(x5,y5)	Green	1.5	0.6



Calculate Weight

Based on a **Weighing Function**

Distance Increases, Weight decreases

Simplest Weighing function

K=5

$$w_i = 1/d_i$$

1.4 + 0.8 + 0.6 = 2.8

5 + 2 = 7

Image: Google.com

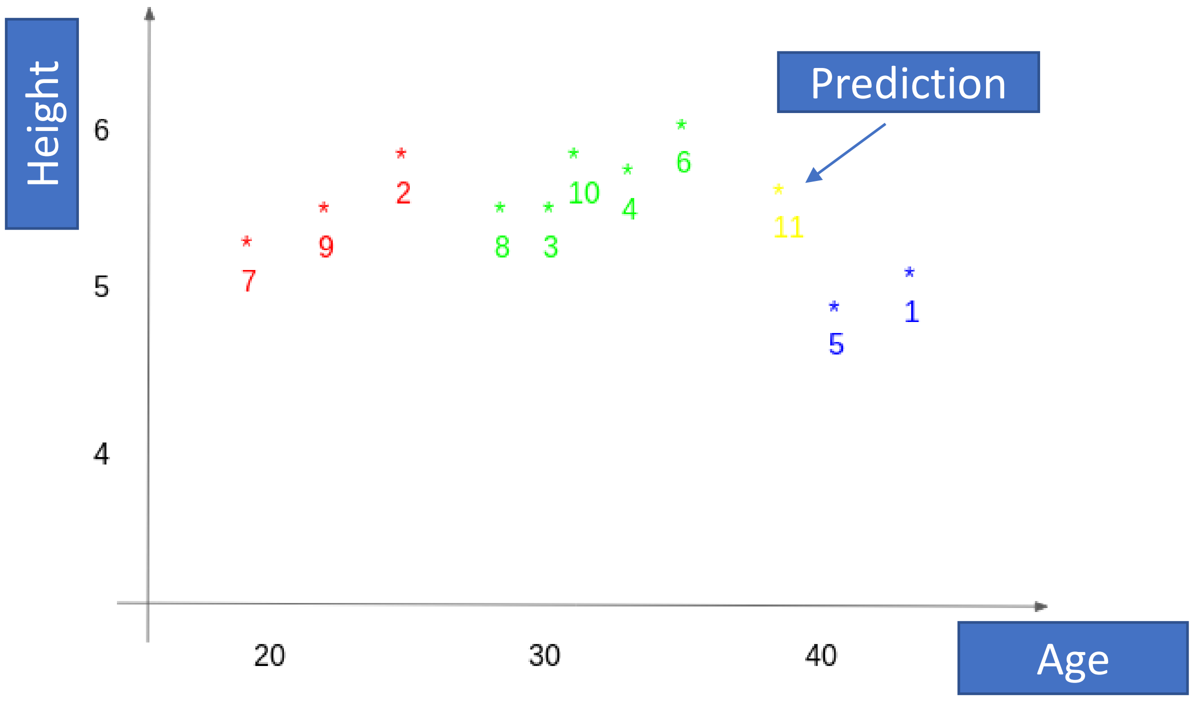
Let's Calculate for Regression!

AI Quest

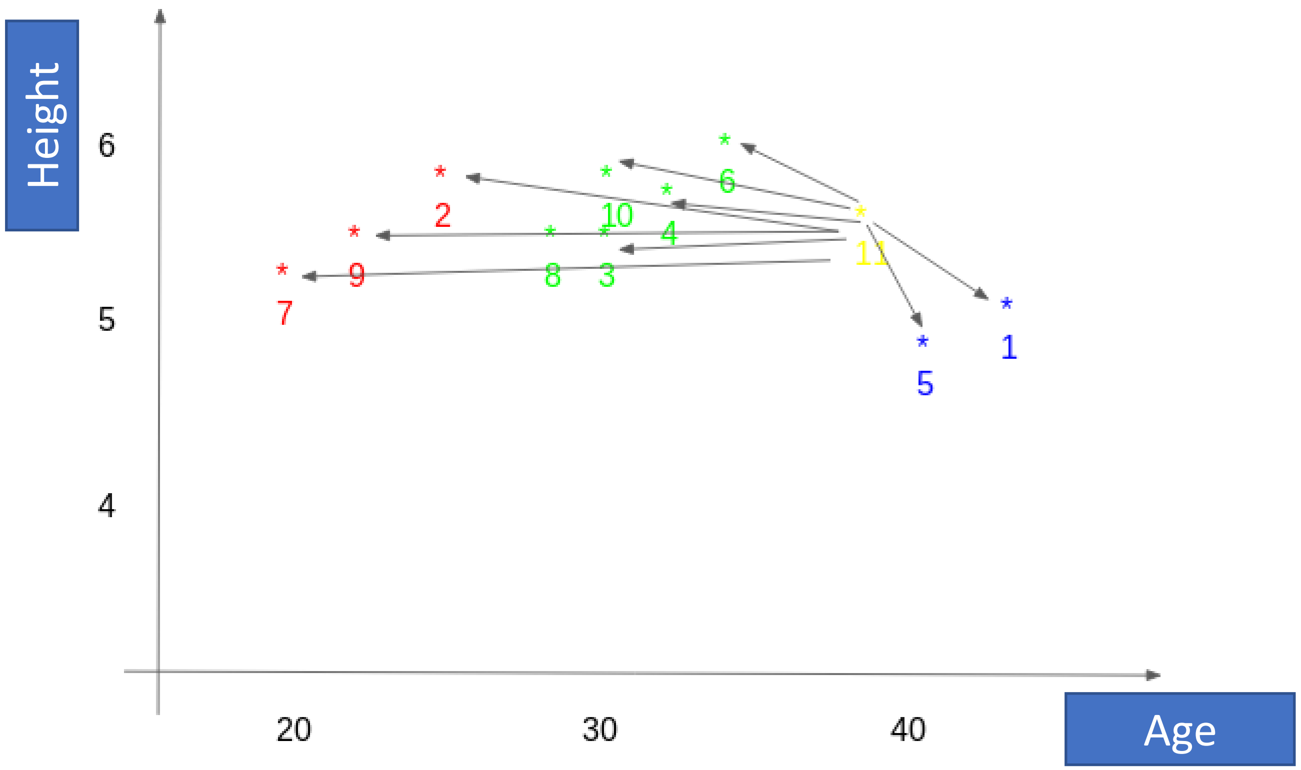
K-NN Algorithm: Regression

	A	B	C	D
1	id	age	height	weight
2	1	45	5	77
3	2	26	5.11	47
4	3	30	5.6	55
5	4	34	5.9	59
6	5	40	4.8	72
7	6	36	5.8	60
8	7	19	5.3	40
9	8	28	5.8	60
10	9	23	5.5	45
11	10	32	5.6	58
12	11	38	5.5	?

K-NN Algorithm: Regression



K-NN Algorithm: Regression



Analyticsvidhya.com

K-NN Algorithm: Regression

	A	B	C	D
1	id	age	height	weight
2	1	45	5	77
3	2	26	5.11	47
4	3	30	5.6	55
5	4	34	5.9	59
6	5	40	4.8	72
7	6	36	5.8	60
8	7	19	5.3	40
9	8	28	5.8	60
10	9	23	5.5	45
11	10	32	5.6	58
12		38	5.5	?

Distance,

$$d(p,q)^2 = (q_1 - p_1)^2 + (q_2 - p_2)^2$$

	A	B	C	E
1	id	age	height	distance
2	1	45	5	a
3	2	26	5.11	b
4	3	30	5.6	c
5	4	34	5.9	d
6	5	40	4.8	e
7	6	36	5.8	f
8	7	19	5.3	g
9	8	28	5.8	h
10	9	23	5.5	i
11	10	32	5.6	j
12	11	38	5.5	
13				

K-NN Algorithm: Regression

	A	B	C	D
1	id	age	height	weight
2	1	45	5	77
3	2	26	5.11	47
4	3	30	5.6	55
5	4	34	5.9	59
6	5	40	4.8	72
7	6	36	5.8	60
8	7	19	5.3	40
9	8	28	5.8	60
10	9	23	5.5	45
11	10	32	5.6	58
12		38	5.5	?

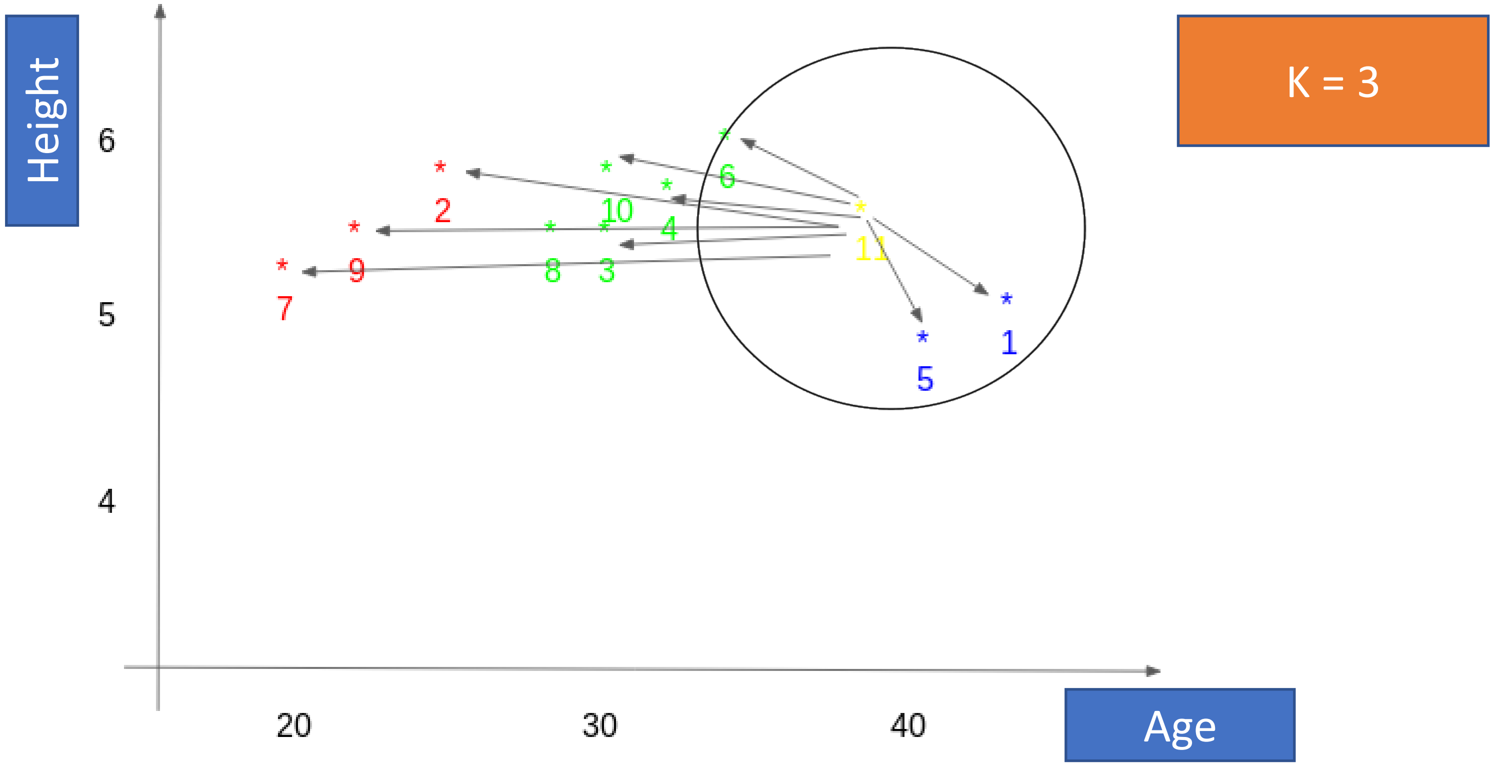
Distance,

$$d(p,q)^2 = (q_1 - p_1)^2 + (q_2 - p_2)^2$$

Sequence:

a > e > f > d > j > c > h > b > i > g

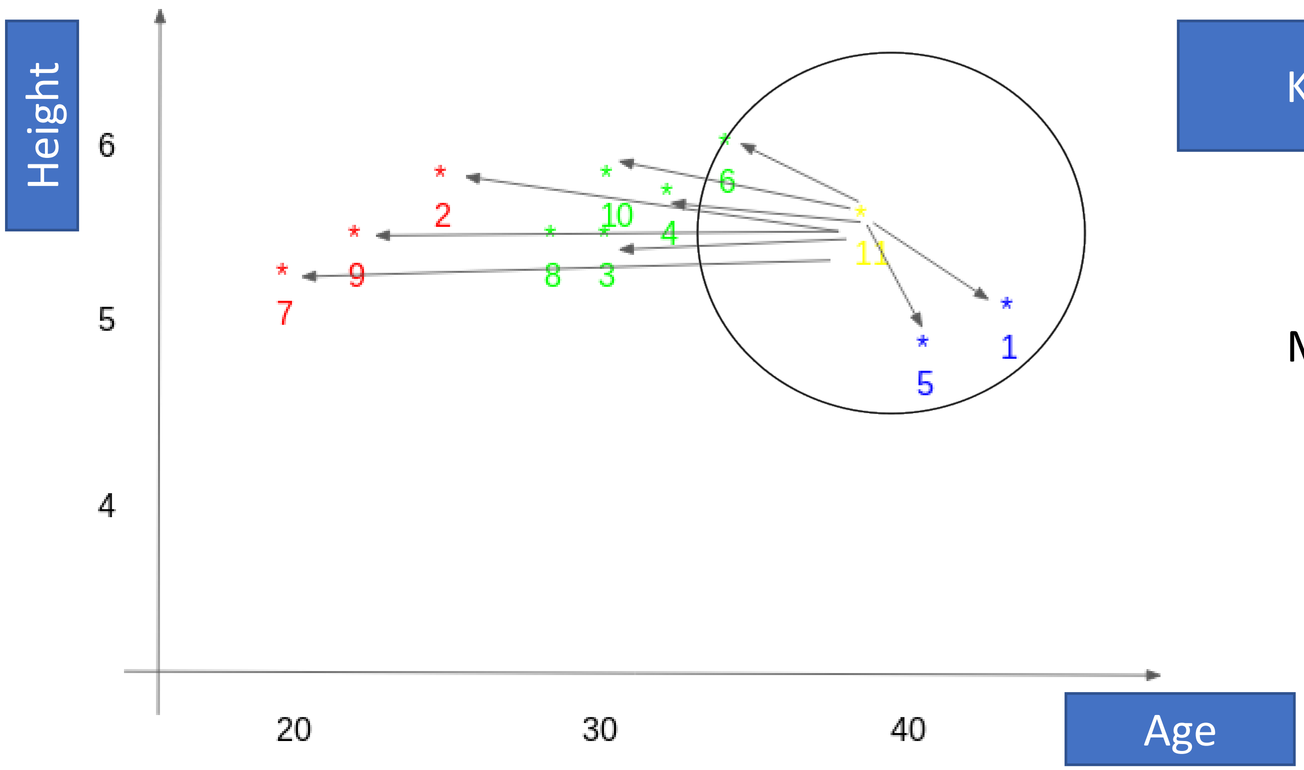
K-NN Algorithm: Regression



Analyticsvidhya.com

K-NN Algorithm: Regression

A	B	C	D
id	age	height	weight
1	45	5	77
2	26	5.11	47
3	30	5.6	55
4	34	5.9	59
5	40	4.8	72
6	36	5.8	60
7	19	5.3	40
8	28	5.8	60
9	23	5.5	45
10	32	5.6	58
	38	5.5	?



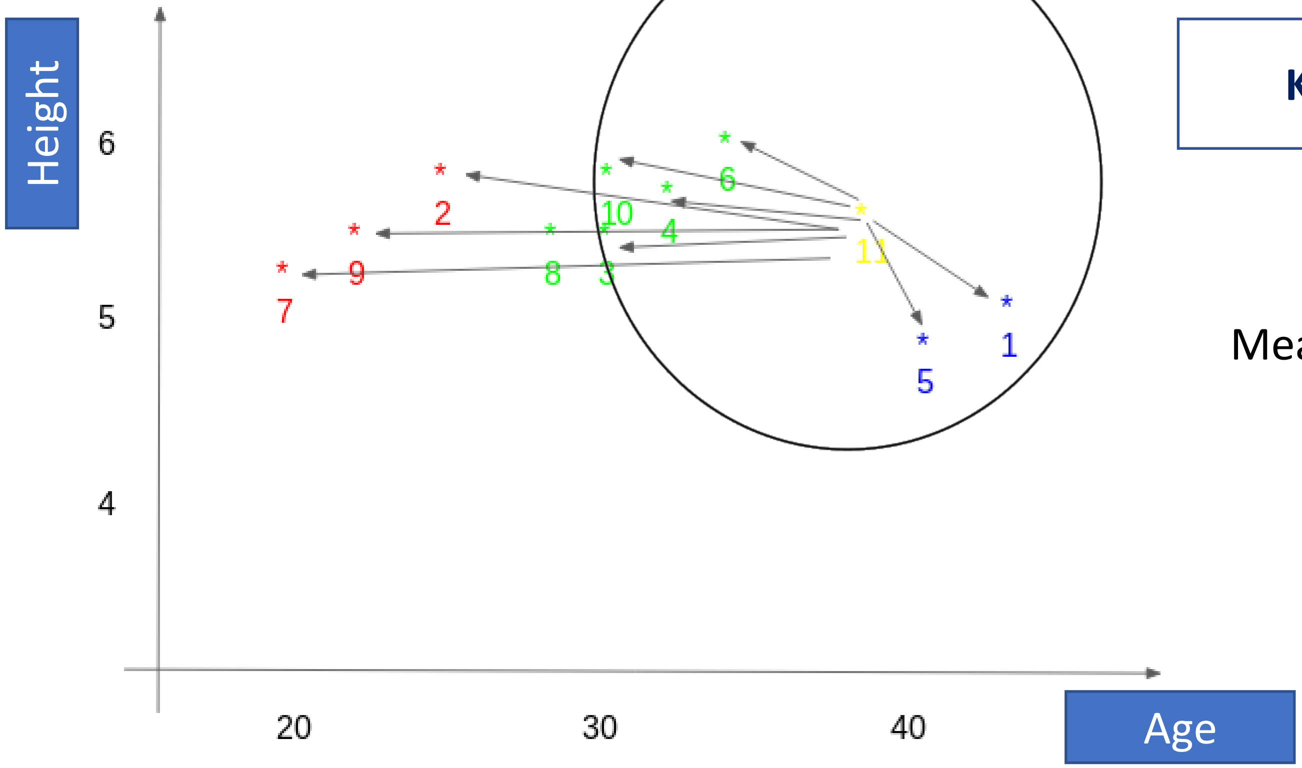
K = 3

Mean: (77+72+60)/3
= 69.6 kg

K-NN Algorithm: Regression

Sequence:
a > e > f > d > j > c > h > b > i > g

	A	B	C	D	E
1	id	age	height	weight	distance
2	1	45	5	77	a
3	2	26	5.11	47	b
4	3	30	5.6	55	c
5	4	34	5.9	59	d
6	5	40	4.8	72	e
7	6	36	5.8	60	f
8	7	19	5.3	40	g
9	8	28	5.8	60	h
10	9	23	5.5	45	i
11	10	32	5.6	58	j
12	11	38	5.5	?	
13					



K = 5

Mean: (77+72+60+59+58)/5
= 65.2 kg

Let’s Do Self Assessment!

www.aiquest.org