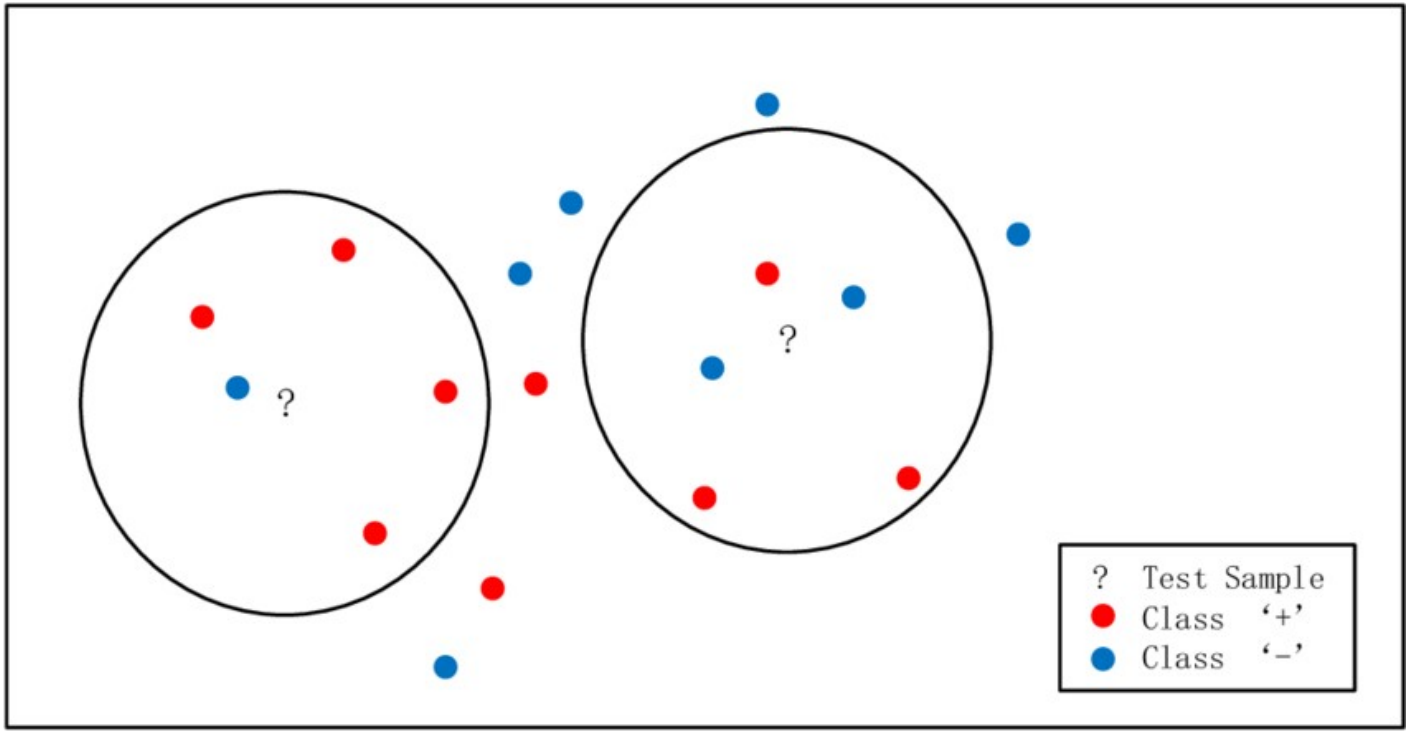
Getting Started with ML

Raunak Joshi

K-Nearest Neighbors



Cheng, Debo & Zhang, Shichao & Deng, Zhenyun & Zhu, Yonghua & Zong, Ming. (2014). kNN Algorithm with Data-Driven k Value. 499-512. 10.1007/978-3-319-14717-8_39.

So how we do it ?

- Requires K value.
- Uses Distance Formulas.
- Uses a Voting Mechanism.

Numerical for Conceptualization

Car Mileage	Car Efficient Speed	Outcome
25	100	Good
11	80	Bad
6	70	Bad
8	100	Good
9	100	Good

Numerical for Conceptualization

Car Mileage	Car Efficient Speed	Outcome
25	100	Good
11	80	Bad
6	70	Bad
8	100	Good
9	100	Good

Step 1 – Select K value. This K value is responsible for finding the Nearest Estimators with respect to all the data.

Selecting K = 3 for this problem.

Numerical for Conceptualization

Car Mileage	Car Efficient Speed	Outcome
25	100	Good
11	80	Bad
6	70	Bad
8	100	Good
9	100	Good

Step 2- Select a distance estimation algorithm from Euclidean, Manhattan or Minkowski.

We are selecting Euclidean for this problem.

$$Euclidean(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Car Mileage (x_i)	Car Efficient Speed (y_i)	Outcome
25	100	Good
11	80	Bad
6	70	Bad
8	100	Good
9	100	Good

$$Euclidean(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Assume there is a new car with Mileage 25 *kmpl and Efficient Speed of* 60 *kmph. Find the if Car is good or bad.*

$I)\ \sqrt{(x_i - y_i)^2 + (x_i - y_i)^2} = \sqrt{(25-25)^2 + (60-100)^2} = \sqrt{1600} = 40$ ⟵ Good ~ 1

$II)\ \sqrt{(x_i - y_i)^2 + (x_i - y_i)^2} = \sqrt{(25-11)^2 + (60-80)^2} = 2\sqrt{149} = 24.41$ ⟵ Bad ~ 0

$III)\ \sqrt{(x_i - y_i)^2 + (x_i - y_i)^2} = \sqrt{(25-6)^2 + (60-70)^2} = \sqrt{461} = 21.47$ ⟵ Bad ~ 0

$IV)\ \sqrt{(x_i - y_i)^2 + (x_i - y_i)^2} = \sqrt{(25-8)^2 + (60-100)^2} = \sqrt{1889} = 43.462$

$V)\ \sqrt{(x_i - y_i)^2 + (x_i - y_i)^2} = \sqrt{(25-9)^2 + (60-100)^2} = \sqrt{1856} = 43.081$

Now comes the voting system which has outcome [Good, Bad, Bad]. Since the Bad values are more in the predicted value, the Bad is the predicted outcome for our problem.

Conclusion ~ The car with Mileage of 25 kmpl and Efficient Speed of 60 kmph is termed as Bad by KNN.