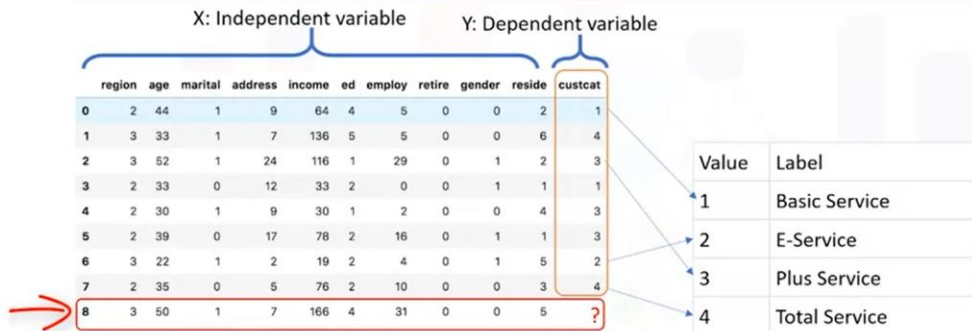
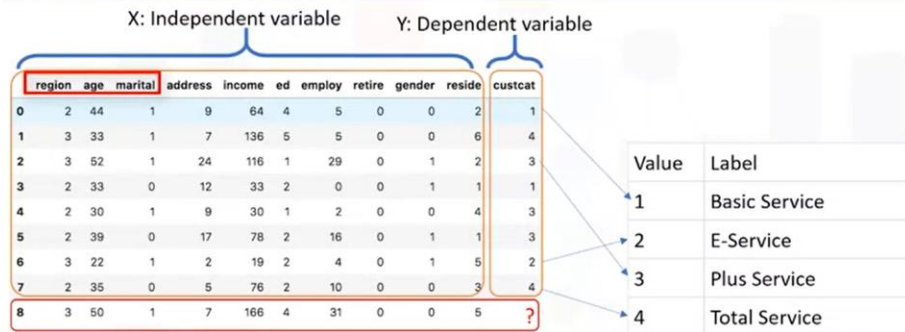


Intro to KNN



Intro to KNN



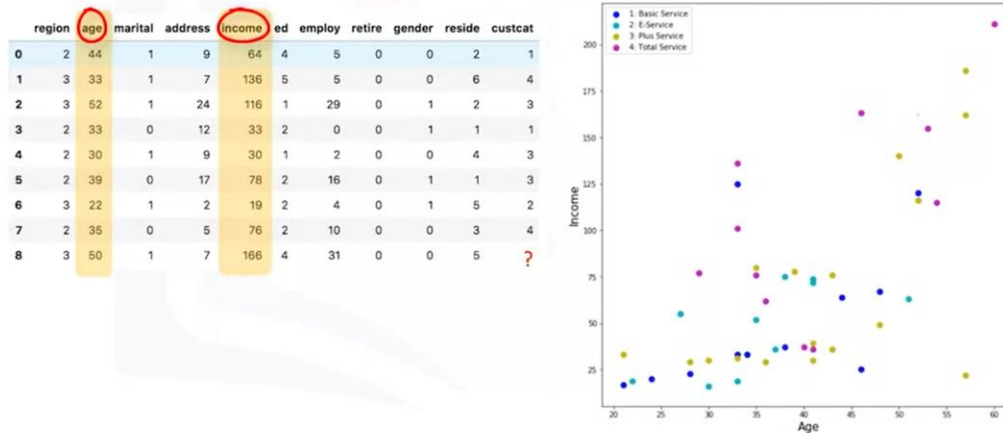
Intro to KNN

X: Independent variable Y: Dependent variable

	region	age	marital	address	income	ed	employ	retire	gender	reside	custcat
0	2	44	1	9	64	4	5	0	0	2	1
1	3	33	1	7	136	5	5	0	0	6	4
2	3	52	1	24	116	1	29	0	1	2	3
3	2	33	0	12	33	2	0	0	1	1	1
4	2	30	1	9	30	1	2	0	0	4	3
5	2	39	0	17	78	2	16	0	1	1	3
6	3	22	1	2	19	2	4	0	1	5	2
7	2	35	0	5	76	2	10	0	0	3	4
8	3	50	1	7	166	4	31	0	0	5	?

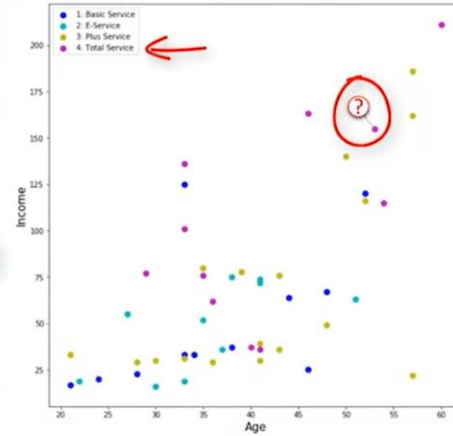
Value	Label
1	Basic Service
2	E-Service
3	Plus Service
4	Total Service

Determining the class using 1st KNN



Determining the class using 1st KNN

	region	age	marital	address	income	ed	employ	retire	gender	reside	custcat
0	2	44	1	9	64	4	5	0	0	2	1
1	3	33	1	7	136	5	5	0	0	6	4
2	3	52	1	24	116	1	29	0	1	2	3
3	2	33	0	12	33	2	0	0	1	1	1
4	2	30	1	9	30	1	2	0	0	4	3
5	2	39	0	17	78	2	16	0	1	1	3
6	3	22	1	2	19	2	4	0	1	5	2
7	2	35	0	5	76	2	10	0	0	3	4
8	3	50	1	7	166	4	31	0	0	5	?

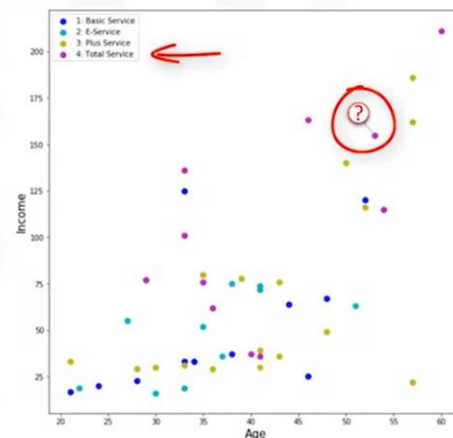


IBM Developer

SKILLS NETWORK 

Determining the class using 1st KNN

	region	age	marital	address	income	ed	employ	retire	gender	reside	custcat
0	2	44	1	9	64	4	5	0	0	2	1
1	3	33	1	7	136	5	5	0	0	6	4
2	3	52	1	24	116	1	29	0	1	2	3
3	2	33	0	12	33	2	0	0	1	1	1
4	2	30	1	9	30	1	2	0	0	4	3
5	2	39	0	17	78	2	16	0	1	1	3
6	3	22	1	2	19	2	4	0	1	5	2
7	2	35	0	5	76	2	10	0	0	3	4
8	3	50	1	7	166	4	31	0	0	5	?



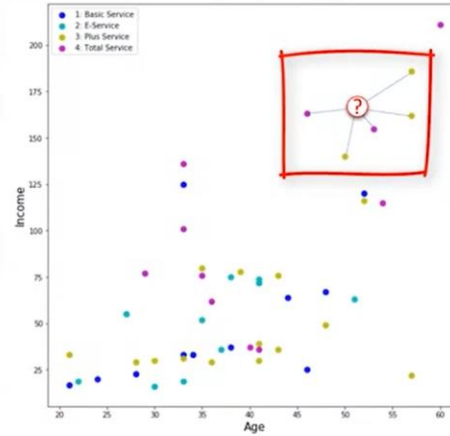
1-NN → 4: Total Service

IBM Developer

SKILLS NETWORK 

Determining the class using the 5 KNNs

	region	age	marital	address	income	ed	employ	retire	gender	reside	custcat
0	2	44	1	9	64	4	5	0	0	2	1
1	3	33	1	7	136	5	5	0	0	6	4
2	3	52	1	24	116	1	29	0	1	2	3
3	2	33	0	12	33	2	0	0	1	1	1
4	2	30	1	9	30	1	2	0	0	4	3
5	2	39	0	17	78	2	16	0	1	1	3
6	3	22	1	2	19	2	4	0	1	5	2
7	2	35	0	5	76	2	10	0	0	3	4
8	3	50	1	7	166	4	31	0	0	5	?



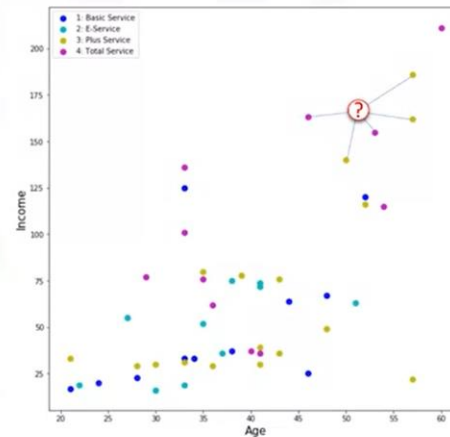
IBM Developer

SKILLS NETWORK 

Determining the class using the 5 KNNs

	region	age	marital	address	income	ed	employ	retire	gender	reside	custcat
0	2	44	1	9	64	4	5	0	0	2	1
1	3	33	1	7	136	5	5	0	0	6	4
2	3	52	1	24	116	1	29	0	1	2	3
3	2	33	0	12	33	2	0	0	1	1	1
4	2	30	1	9	30	1	2	0	0	4	3
5	2	39	0	17	78	2	16	0	1	1	3
6	3	22	1	2	19	2	4	0	1	5	2
7	2	35	0	5	76	2	10	0	0	3	4
8	3	50	1	7	166	4	31	0	0	5	?

5-NN → 3: Plus Service



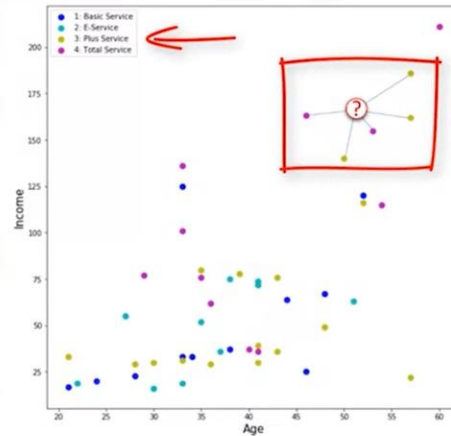
IBM Developer

SKILLS NETWORK 

Determining the class using the 5 KNNs

	region	age	marital	address	income	ed	employ	retire	gender	reside	custcat
0	2	44	1	9	64	4	5	0	0	2	1
1	3	33	1	7	136	5	5	0	0	6	4
2	3	52	1	24	116	1	29	0	1	2	3
3	2	33	0	12	33	2	0	0	1	1	1
4	2	30	1	9	30	1	2	0	0	4	3
5	2	39	0	17	78	2	16	0	1	1	3
6	3	22	1	2	19	2	4	0	1	5	2
7	2	35	0	5	76	2	10	0	0	3	4
8	3	50	1	7	166	4	31	0	0	5	?

5-NN → 3: Plus Service

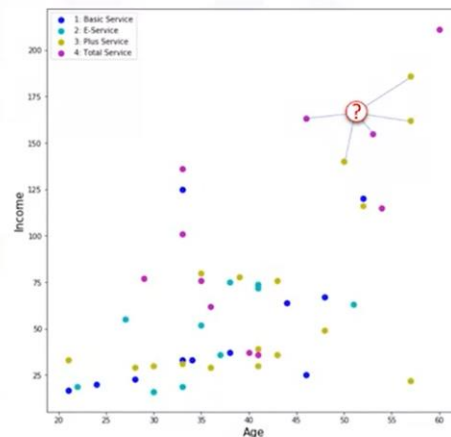


IBM Developer

SKILLS NETWORK

What is K-Nearest Neighbor (or KNN)?

- A method for **classifying** cases based on their similarity to other cases
- Cases that are near each other are said to be “**neighbors**”
- Based on **similar** cases with same **class labels** are near each other



IBM Developer

SKILLS NETWORK

The K-Nearest Neighbors algorithm

1. Pick a value for K.
2. Calculate the distance of unknown case from all cases.
3. Select the K-observations in the training data that are “nearest” to the unknown data point.
4. Predict the response of the unknown data point using the most popular response value from the K-nearest neighbors.

IBM Developer

SKILLS NETWORK 

Calculating the similarity/distance in a 1-dimensional space



Customer 1

Age
34



Customer 2

Age
30

$$\text{Dis}(x_1, x_2) = \sqrt{\sum_{i=0}^n (x_{1i} - x_{2i})^2}$$

$$\text{Dis}(x_1, x_2) = \sqrt{(34 - 30)^2} = 4$$

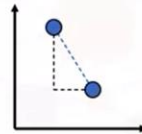
IBM Developer

SKILLS NETWORK 

Calculating the similarity/distance in a 2-dimensional space



Customer 1	
Age	Income
34	190



Customer 2	
Age	Income
30	200

$$\begin{aligned}\text{Dis}(x_1, x_2) &= \sqrt{\sum_{i=0}^n (x_{1i} - x_{2i})^2} \\ &= \sqrt{(34 - 30)^2 + (190 - 200)^2} = 10.77\end{aligned}$$

IBM Developer

SKILLS NETWORK 

Calculating the similarity/distance in a multi-dimensional space



Customer 1		
Age	Income	Education
34	190	3



Customer 2		
Age	Income	Education
30	200	8

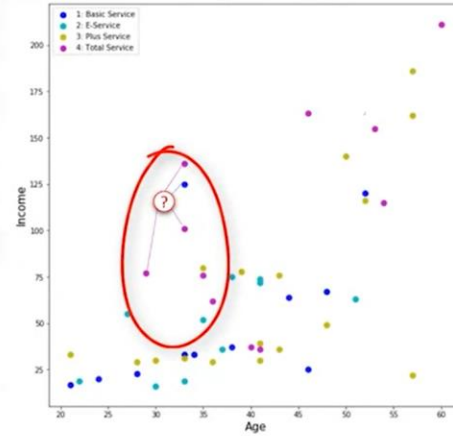
$$\begin{aligned}\text{Dis}(x_1, x_2) &= \sqrt{\sum_{i=0}^n (x_{1i} - x_{2i})^2} \\ &= \sqrt{(34 - 30)^2 + (190 - 200)^2 + (3 - 8)^2} = 11.87\end{aligned}$$

IBM Developer

SKILLS NETWORK 

What is the best value of K for KNN?

- K =1 class 1



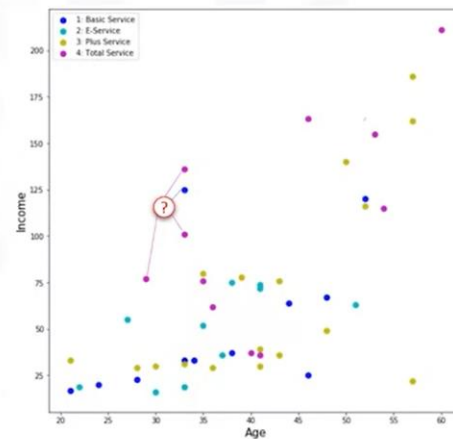
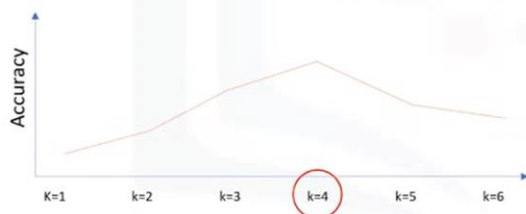
IBM Developer

SKILLS NETWORK

This is a bad prediction

What is the best value of K for KNN?

- K =1 class 1
- K =20 ?



IBM Developer

SKILLS NETWORK

Computing continuous targets using KNN

- KNN can also be used for regression

