

→ Naive Bayesian example :-

Ex	Color	Type	Origin	Stolen
1	R	S	D	Y
2	R	S	D	N
3	R	S	D	Y
4	Y	S	D	N
5	Y	S	I	Y
6	Y	X	I	N
7	Y	X	I	Y
8	Y	X	D	N
9	R	X	I	N
10	R	S	I	Y

$X = (\text{Red, SUV, Domestic}) \rightarrow (Y/N)$

$$p(\text{Yes}) = \frac{5}{10} = 0.5$$

$$p(\text{No}) = \frac{5}{10} = 0.5$$

Color	Yes	No
Red	3/5	2/5
Yellow	2/5	3/5

Type	Yes	No
Sports	4/5	2/5
SUV	1/5	3/5

Origin	Yes	No
Dom	2/5	3/5
Imp	3/5	2/5

$$P(\text{Yes} / \text{New Instance}) = P(\text{Yes}) * P(\text{color} = \text{Red} / \text{Yes})$$

$$* P(\text{Type} = \text{SUV} / \text{Yes}) *$$

$$P(\text{Origin} = \text{Domestic} / \text{Yes})$$

$$= 0.5 * \frac{3}{5} * \frac{1}{5} * \frac{2}{5} = \frac{3}{125} = 0.024$$

$$P(\text{No} / \text{New Instance}) = P(\text{No}) *$$

$$P(\text{color} = \text{Red} / \text{No}) * P(\text{Type} = \text{SUV} / \text{No})$$

$$* P(\text{origin} = \text{domestic} / \text{No})$$

$$= \frac{5}{10} * \frac{2}{5} * \frac{3}{5} * \frac{3}{5} = 0.072$$

$$P(\text{No}) > P(\text{Yes})$$

$\therefore$  'No'



## → Silhouette Coefficient

Cluster label:-

<u>point</u>	<u>cluster label</u>
P1	1
P2	1
P3	2
P4	2

Dissimilarity matrix:-

point	P1	P2	P3	P4
P1	0	0.10	0.65	0.55
P2	0.10	0	0.70	0.60
P3	0.65	0.70	0	0.30
P4	0.55	0.60	0.30	0

calculate silhouette coeff for each point;  
for each of 2 clusters; overall clustering.

$$\text{Silhouette coefficient} = 1 - \frac{a}{b}$$

where:-  $a$  = avg distance of a point to other points in its cluster.

$b$  = min avg of distance of a point to points in another cluster.

point P1 :-

$$a = \frac{0.1}{1} = 0.1$$



distance from cluster 1 i.e. from P1 to  
P2 = 0.10 (in the matrix)

$$b = \frac{0.65 + 0.55}{2} = 0.6$$

↓

min avg distance from P1 to P3  
and P1 to P4.

$$SC = 1 - \frac{a}{b} = 1 - \frac{0.1}{0.6} = 0.833.$$

point P2 :-

$$a = \frac{0.1}{1} = 0.1 \quad b = \frac{0.70 + 0.60}{2}$$

$$SC = 1 - \frac{0.1}{\frac{0.70 + 0.60}{2}} = 0.846.$$

point P3 :-

$$SC = 1 - \frac{0.30}{\frac{0.65 + 0.70}{2}} = 0.556.$$



point P4:-

$$Sc = 1 - \frac{0.30}{\frac{0.55 + 0.60}{2}} = 0.478$$

$$P1: Sc = 0.833$$

$$P2: Sc = 0.846$$

$$P3: Sc = 0.556$$

$$P4: Sc = 0.478$$

$$\text{avg } Sc \text{ for cluster 1} = \frac{0.833 + 0.846}{2} = 0.84$$

$$\text{cluster 2} = \frac{0.556 + 0.478}{2} = 0.517$$

$$Sc \text{ for entire cluster} = \frac{0.84 + 0.517}{2} = 0.68$$