

CSE 482 HW 4 PART 1

1. Euclidean Distance = $\sqrt{(x_{1_{\text{test}}} - x_1)^2 + (x_{2_{\text{test}}} - x_2)^2}$

(a)

Data Point	x_1	x_2	class	Euclidean Distance
p1	0.3	0.2	+	0.364005
p2	0.2	0.45	+	0.10
p3	0.5	0.2	+	0.460977229
p4	0.1	0.1	+	0.460977229
p5	0.4	0.1	+	0.4924428901
p6	0.25	0.8	-	0.2549509757
p7	0.3	0.5	-	0.1118033989
p8	0.4	0.8	-	0.3201562119
p9	0.15	0.7	-	0.158113883
p10	0.3	0.7	-	0.1802775638

(b) Data point p2 because it has the smallest distance so test instance is '+' class.

(c) The five nearest neighbors are p2, p7, p9, p10, and p6. The test instance is '-' class.

(d) The 5-nearest neighbor approach is more reliable because it uses more data points.

$$c) \log \left(\frac{P(\hat{y}=1 | x_1, x_2)}{P(\hat{y}=-1 | x_1, x_2)} \right) = -122.1774 x_2 - 73.36 x_1 + 73.3023$$

$$-122.1774(0.55) - 73.36(0.2) + 73.3023 = -8.56727$$

The test instance will have '-' class label.

2(a)

Iter	Cluster Assignment of data points (enter A, B, or C)								Centroid Locations		
	0.10	0.20	0.80	0.90	1.00	1.30	1.80	1.90	A	B	C
0	-	-	-	-	-	-	-	-	0.10	0.20	1.90
1	A	B	B	B	B	C	C	C	0.10	0.725	1.67
2	A	A	B	B	B	C	C	C	0.15	0.9	1.67
3	A	A	B	B	B	C	C	C	0.15	0.9	1.67

(b)

$$SSE = \sum_{i=1}^k \sum_{x_i \in C_j} d(x_i, c_j)^2$$

$$SSE = (0.15 - 0.10)^2 + (0.15 - 0.20)^2 + (0.9 - 0.8)^2 + (0.9 - 0.9)^2 + (0.9 - 1.00)^2 \\ + (1.67 - 1.30)^2 + (1.67 - 1.80)^2 + (1.67 - 1.90)^2$$

$$SSE = 0.2317$$

(c)

Data	cluster Assignment of data points								Centroid Locations		
	0.10	0.20	0.80	0.90	1.00	1.30	1.80	1.90	A	B	C
0	-	-	-	-	-	-	-	-	0.80	1.00	1.80
1	A	A	A	A	B	B	C	C	0.50	1.15	1.85
2	A	A	A	B	B	B	C	C	0.367	1.07	1.85
3	A	A	B	B	B	B	C	C	0.15	1.00	1.85
4	A	A	B	B	B	B	C	C	0.15	1.00	1.85

(d)

$$SSE = (0.1 - 0.15)^2 + (0.2 - 0.15)^2 + (0.8 - 1)^2 + (0.9 - 1)^2 + (1 - 1)^2 + (1.3 - 1)^2 + (1.8 - 1.85)^2 + (1.9 - 1.85)^2$$

$$SSE = 0.15$$

3. a)

Itemset	support
{Bread, eggs}	$\frac{4}{10}$
{Bread, milk}	$\frac{5}{10}$
{Bread, butter}	$\frac{5}{10}$
{butter, milk}	$\frac{5}{10}$

b)

{Bread, eggs, milk}
 {Bread, butter, eggs}
 {Bread, butter, milk}

c)

Itemset	support
{Bread, eggs, milk}	$\frac{3}{10}$
{Bread, butter, eggs}	$\frac{3}{10}$
{Bread, butter, milk}	$\frac{4}{10}$

d) e)

Rules	Support	confidence
{Bread} \rightarrow {Butter, milk}	$\frac{4}{10}$	$\frac{4}{7} = 0.57$
{butter} \rightarrow {Bread, milk}	$\frac{4}{10}$	$\frac{4}{6} = 0.67$
{milk} \rightarrow {Bread, butter}	$\frac{4}{10}$	$\frac{4}{8} = 0.5$
{Bread, butter} \rightarrow {milk}	$\frac{4}{10}$	$\frac{4}{5} = 0.8$
{Bread, milk} \rightarrow {butter}	$\frac{4}{10}$	$\frac{4}{5} = 0.8$
{butter, milk} \rightarrow {Bread}	$\frac{4}{10}$	$\frac{4}{5} = 0.8$

Support ≥ 0.4 and confidence ≥ 0.7

{Bread, butter} \rightarrow {milk}
{Bread, milk} \rightarrow {butter}
{Butter, milk} \rightarrow {bread}