Midterm Exam CS513

Name: Lan Chang

Question 1

$A = \{1, 2, 3\}$

$B = \{1, 2, 5, 6\}$

$C = \{1, 3, 5\}$

a) $\overline{(A \cup B)} = \{4\}$

$\overline{A} \cap \overline{B} = \{4\}$

True.

b) $\overline{(A \cap C)} = \{2, 4, 5, 6\}$

$\overline{A} \cup \overline{C} = \{2, 4, 5, 6\}$

True.

c) $\overline{(A \cap C)} - B = \{4\}$

$\overline{A} \cup (C - B) = \{3, 4, 5, 6\}$

False.

Question 2

$$d(x, y) = \left(\sum_i (x_i - y_i)^2 \right)^{3/2}$$

A distance metric or distance function is a real-valued function d, such that for any coordinates x, y, and z:

1. $d(x, y) \geq 0, \text{and } d(x, y) = 0 \text{ if and only if } x = y$
2. $d(x, y) = d(y, x)$
3. $d(x, z) \leq d(x, y) + d(y, z)$

For statement 3:

Distance from (0, 0) to (0, 1), $d1 = ((0-0)^2 + (0-1)^2)^{3/2} = 1$.

Distance from (0, 1) to (1, 1), $d2 = ((0-1)^2 + (1-1)^2)^{3/2} = 1$.

Distance from (0, 0) to (1, 1), $d3 = ((0-1)^2 + (0-1)^2)^{3/2} = 2\sqrt{2} \approx 2.828$.

$d1 + d2 = 2 < 2.828 = d3$, so **this function is not a proper distance function.**

Question 3

See Midterm_q3.R

Question 4

	International Plan	Voice Plan	Churn False	Churn True	Row Total
	no	no	1878	302	2180
	no	yes	786	44	830
Sub-Total			2664	346	3010
	yes	no	130	101	231
	yes	yes	56	36	92
Sub-Total			186	137	323
Grand-Total			2850	483	3333

$P(\text{Churn} = \text{True}) = 483/3333 \approx \mathbf{0.1449}$
$P(\text{Churn} = \text{False}) = 2850/3333 \approx \mathbf{0.8551}$
$P(\text{International Plan} = \text{Yes}) = 323/3333 \approx \mathbf{0.0969}$
$P(\text{Voice Plan} = \text{Yes}) = (830 + 92)/3333 \approx \mathbf{0.2766}$
$P(\text{International Plan} = \text{Yes}, \text{Voice Plan} = \text{Yes}) = 92/3333 \approx \mathbf{0.0276}$

Are "Voice Plan" and "International plan" independent?
$P(\text{VPlan} = \text{Yes}) = 922/3333$
$P(\text{IPlan} = \text{Yes}) = 323/3333$
$P(\text{VPlan} = \text{Yes}) \times P(\text{IPlan} = \text{Yes}) = (922 \times 323)/(3333 \times 3333) = 297806/11108889$
$P(\text{IPlan} = \text{Yes}, \text{VPlan} = \text{Yes}) = 92/3333 = 306636/11108889$
$P(\text{VPlan} = \text{Yes}) \times P(\text{IPlan} = \text{Yes}) \neq P(\text{IPlan} = \text{Yes}, \text{VPlan} = \text{Yes})$
They are not independent.

$$P((\text{International Plan} = \text{Yes}, \text{Voice Plan} = \text{Yes})/\text{Churn} = \text{True}) = \frac{P(\text{International Plan} = \text{Yes}, \text{Voice Plan} = \text{Yes}, \text{Churn} = \text{True})}{P(\text{Churn} = \text{True})}$$
$= (36/3333)/(483/3333) = 36/483 \approx \mathbf{0.0745}$

$$P((\text{International Plan} = \text{Yes}, \text{Voice Plan} = \text{Yes})/\text{Churn} = \text{False}) = \frac{P(\text{International Plan} = \text{Yes}, \text{Voice Plan} = \text{Yes}, \text{Churn} = \text{False})}{P(\text{Churn} = \text{False})}$$
$= (56/3333)/(2850/3333) = 56/2850 \approx \mathbf{0.0196}$

$$P(\text{Churn} = \text{True}/(\text{International Plan} = \text{Yes}, \text{Voice Plan} = \text{Yes})) = \frac{P(\text{International Plan} = \text{Yes}, \text{Voice Plan} = \text{Yes}, \text{Churn} = \text{True})}{P(\text{International Plan} = \text{Yes}, \text{Voice Plan} = \text{Yes})}$$
$= (36/3333)/(92/3333) = 36/92 \approx \mathbf{0.3913}$

$$P(\text{Churn} = \text{False}/(\text{International Plan} = \text{Yes}, \text{Voice Plan} = \text{Yes})) = \frac{P(\text{International Plan} = \text{Yes}, \text{Voice Plan} = \text{Yes}, \text{Churn} = \text{False})}{P(\text{International Plan} = \text{Yes}, \text{Voice Plan} = \text{Yes})}$$
$= (56/3333)/(92/3333) = 56/92 \approx \mathbf{0.6087}$

Question 5

a)

Customer	Day Calls	Eve Calls	Night Calls
A	110	99	91
B	123	103	103
C	71	88	89
D	113	122	121
E	98	101	118
X	114	110	?

After normalization:

Customer	Day Calls	Eve Calls
A	0.628571	0.565714
B	0.702857	0.588571
C	0.405714	0.502857
D	0.645714	0.697143
E	0.56	0.577143
X	0.651429	0.628571

Distance:

	distance	1/(d^2)
A	0.066884	223.5401
B	0.065153	235.5769
C	0.276007	13.12688
D	0.068809	211.2069
E	0.1049	90.87537

k=1, unweighted vote:

B

$\text{NightCalls(X)} = \mathbf{103}$

k=2, unweighted vote:

A, B

$\text{NightCalls(X)} = (91 + 103)/2 = \mathbf{97}$

k=3, weighted vote:

A, B, D

$$\text{NightCalls(X)} = \frac{91 \times 223.5401 + 103 \times 235.5769 + 121 \times 211.2069}{223.5401 + 235.5769 + 211.2069} \approx 104.67 \approx \mathbf{105}$$

b)

Customer	Day Calls	Eve Calls	Night Calls
A	110	99	Low
B	123	103	Medium
C	71	88	Low
D	113	122	High
E	98	101	Medium
F	114	110	High
G	114	110	Medium
H	114	110	Medium
X	114	110	?

k=3, weighted vote:

3 no distance: F, G, H

There are two Medium and one High.

NightCalls(X) = **Medium**