Assignment - 3

Assignment - 3

1.

a. fasle
b. false
c. false
d. True
e. false
f. fasle
g. fasle

2.

Using the matrices, we can compute the accuracy and cost for each model:

Accuracy for Model M1:

Accuracy = (True Positives + True Negatives) / Total

Accuracy = (100 + 200) / (100 + 50 + 150 + 200)

Accuracy = 300 / 500 = 0.6

Cost for Model M1:

Cost = (True Positives * Cost of False Negatives + False Positives * Cost of False Positives + True Negatives * Cost of True Negatives) / Total

Cost = (100 * 50 + 50 * 1 + 200 * 0) / (100 + 50 + 150 + 200)

Cost = 5050 / 500 = 10.1

Therefore, the accuracy of Model M1 is **0.6** and the cost of Model M1 is **10.1**.

Accuracy for Model M2:

Accuracy = (True Positives + True Negatives) / Total

Accuracy = (200 + 200) / (200 + 90 + 10 + 200)

Accuracy = 400 / 500 = 0.8

Cost for Model M2:

Cost = (True Positives * Cost of False Negatives + False Positives * Cost of False Positives + True Negatives * Cost of True Negatives) / Total

Cost = (200 * 50 + 90 * 1 + 200 * 0) / (200 + 90 + 10 + 200)

Cost = 10090/500 = 20.18

Therefore, the accuracy of Model M2 is **0.8** and the cost of Model M2 is **20.18**.

3.a.

PERMUTATION 1, (h1)

		S1	S2	S3	S4
1	A	1	0	1	0
2	E	0	1	0	1
3	B	1	0	0	1
4	G	1	0	1	0
5	F	1	0	1	0
6	C	0	1	0	1
7	D	0	1	0	1
		1	**2**	**1**	**2**

PERMUTATION 2, (h2)

		S1	S2	S3	S4
1	E	0	1	0	1
2	B	1	0	0	1
3	C	0	1	0	1
4	F	1	0	1	0
5	G	1	0	1	0
6	A	1	0	1	0
7	D	0	1	0	1
		2	**1**	**4**	**1**

PERMUTATION 3, (h3)

		S1	S2	S3	S4
1	D	0	1	0	1
2	B	1	0	0	1
3	F	1	0	1	0
4	G	1	0	1	0
5	A	1	0	1	0
6	E	0	1	0	1
7	C	0	1	0	1
		2	**1**	**3**	**1**

The Signature matrix is as follows,

	S1	S2	S3	S4
h1	1	2	1	2
h2	2	1	4	1
h3	2	1	3	1

3.b.

(1). the original representation in M

Sim(S1 , S1) = Sim(S2 , S2) = Sim(S3 , S3) = Sim(S4 , S4) = 1

Sim(S1 , S2) = 0/7 = 0 = Sim(S2 , S1)

Sim(S1 , S3) = 3/4 = 0.75 = Sim(S3 , S1)

Sim(S1 , S4) = 1/7 = 0.14 = Sim(S4 , S1)

Sim(S2 , S3) = 0/6 = 0 = Sim(S3 , S2)

Sim(S2 , S4) = 3/4 = 0.75 = Sim(S4 , S2)

Sim(S3 , S4) = 0/7 = 0 = Sim(S4 , S3)

	S1	S2	S3	S4
S1	1	0	0.75	0.14
S2	0	1	0	0.75
S3	0.75	0	1	0
S4	0.14	0.75	0	1

(2). Minhashing generated by the permutations in question (a)

Permutation 1:

Sim(S1 , S1) = Sim(S2 , S2) = Sim(S3 , S3) = Sim(S4 , S4) = 1

Sim(S1 , S2) = 0/7 = 0 = Sim(S2 , S1)

Sim(S1 , S3) = 3/4 = 0.75 = Sim(S3 , S1)

Sim(S1 , S4) = 1/7 = 0.14 = Sim(S4 , S1)

Sim(S2 , S3) = 0/6 = 0 = Sim(S3 , S2)

Sim(S2 , S4) = 3/4 = 0.75 = Sim(S4 , S2)

Sim(S3 , S4) = 0/7 = 0 = Sim(S4 , S3)

	S1	S2	S3	S4
S1	1	0	0.75	0.14
S2	0	1	0	0.75
S3	0.75	0	1	0
S4	0.14	0.75	0	1

Permutation 2:

Sim(S1 , S1) = Sim(S2 , S2) = Sim(S3 , S3) = Sim(S4 , S4) = 1

Sim(S1 , S2) = 0/7 = 0 = Sim(S2 , S1)

Sim(S1 , S3) = 3/4 = 0.75 = Sim(S3 , S1)

Sim(S1 , S4) = 1/7 = 0.14 = Sim(S4 , S1)

Sim(S2 , S3) = 0/6 = 0 = Sim(S3 , S2)

Sim(S2 , S4) = 3/4 = 0.75 = Sim(S4 , S2)

Sim(S3 , S4) = 0/7 = 0 = Sim(S4 , S3)

	S1	S2	S3	S4
S1	1	0	0.75	0.14
S2	0	1	0	0.75
S3	0.75	0	1	0
S4	0.14	0.75	0	1

Permutation 3:

Sim(S1 , S1) = Sim(S2 , S2) = Sim(S3 , S3) = Sim(S4 , S4) = 1

Sim(S1 , S2) = 0/7 = 0 = Sim(S2 , S1)

Sim(S1 , S3) = 3/4 = 0.75 = Sim(S3 , S1)

Sim(S1 , S4) = 1/7 = 0.14 = Sim(S4 , S1)

Sim(S2 , S3) = 0/6 = 0 = Sim(S3 , S2)

Sim(S2 , S4) = 3/4 = 0.75 = Sim(S4 , S2)

Sim(S3 , S4) = 0/7 = 0 = Sim(S4 , S3)

	S1	S2	S3	S4
S1	1	0	0.75	0.14
S2	0	1	0	0.75
S3	0.75	0	1	0
S4	0.14	0.75	0	1

3.c.

x		S1	S2
0	A	1	0
1	B	1	0
2	C	0	1
3	D	0	1
4	E	0	1
5	F	1	0
6	G	1	0

Using two hashing functions: h(x) = (x + 1)mod 7 and $g(x)$ = (2x + 3)mod 7

For x = 0,

h(0) = (0 + 1)mod 7 = 1mod 7 = 1

g(0) = (2*0 + 3)mod 7 = 3mod 7 = 3

For x = 1,

h(1) = (1 + 1)mod 7 = 2mod 7 = 2

g(1) = (2*1 + 3)mod 7 = 5mod 7 = 5

For x = 2,

h(2) = (2 + 1)mod 7 = 3mod 7 = 3

g(2) = (2*2 + 3)mod 7 = 7mod 7 = 0

For x = 3,

h(3) = (3 + 1)mod 7 = 4mod 7 = 4

g(3) = (2*3 + 3)mod 7 = 9mod 7 = 2

For x = 4,

h(4) = (4 + 1)mod 7 = 5mod 7 = 5

g(4) = (2*4 + 3)mod 7 = 11mod 7 = 4

For x = 5,

h(5) = (5 + 1)mod 7 = 6mod 7 = 6

g(5) = (2*5 + 3)mod 7 = 13mod 7 = 5

For x = 6,

h(6) = (6 + 1)mod 7 = 7mod 7 = 2

g(6) = (2*6 + 3)mod 7 = 15mod 7 = 5

Signatures after hashing are,

h(x)	S1	S2
0	1	0
1	1	0
2	1	0
3	0	1
4	0	1
5	0	1
6	1	0

g(x)	S1	S2
0	0	1
1	1	0
2	0	1
3	1	0
4	0	1
5	1	0
6	1	0