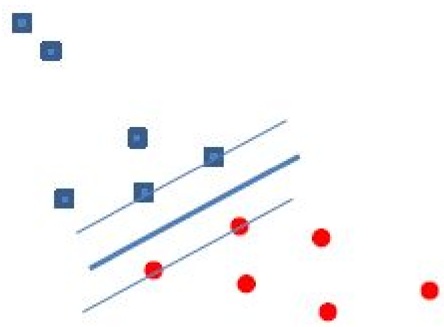
1) **(a) Draw (approximately) the SVM line separator.**



(b) Suppose we find (1/2)* w^2 to be 2 in the SVM optimization. What is the margin, i.e. the distance of closest points to the line?

½ w^2 = 2
w^2 = 4
|| w || = 2
Margin = 1 / ||w|| = ½

(c) Now consider the dataset in Fig 2 (the red points are shifted below). Will (1/2)*w 2 be smaller or greater than previously? Explain.

The margin (1/||w||) is greater than the previous margin, therefore is w is smaller, which means ½ w^2 is smaller too.

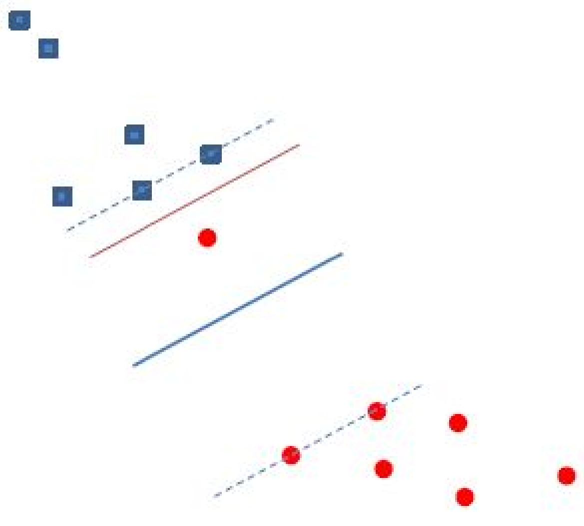
(d) Using a ruler, and the fact that (1/2)*w^2 was 2 previously, find (approximately) the magnitude of the new line coefficient vector, w'

The distance is approximately 4 times greater than the previous distance.
1/||w|| = 2
w = 0.5
½ w^2 = 0.125

(e) Consider the dataset in Fig 3 (with one additional red circle quite close to the blue squares). Assuming optimization using slack variables and C=1, draw a line that does not perfectly separate the points, but which is nonetheless better than the line that perfectly separates the points. (Draw it in the figure, and explain why).

The cost of the Blue:
½ (0.5)^2 + 1.5
= 1.625
Only 1 slack is non Zero because there is only one point that doesn't follow the rule.

The cost of the red:
(½)(2)^2 = 2
All slacks are zero because it perfectly separates the points



(f) Why would we rather prefer the line in (e) to the line that perfectly separates the points?

Because the cost of the blue line is less than the red line.

2. Consider the task of building a classifier from random data, where the attribute values are generated randomly irrespective of the class labels. Assume the data set contains records from two classes, “+” and “−.” Half of the data set is used for training while the remaining half is used for testing.

(a) Suppose there are an equal number of positive and negative records in the data and a classifier predicts every test record to be positive. What is the expected error rate of the classifier on the test data?

Since there are equal number of positive and negative, Every prediction made only have 50% chance to be correct.

(b) Repeat the previous analysis assuming that the classifier predicts each test record to be positive class with probability 0.8 and negative class with probability 0.2.

Since there are equal number of positive and negative, Every prediction made will have 50% chance to be correct.

(c) Suppose two-thirds of the data belong to the positive class and the remaining one-third belong to the negative class. What is the expected error of a classifier that predicts every test record to be positive?

The probability that a positive number is correct = ⅔
The probability that a negative number is correct = ⅓

The probability of our test results is correct is = 1 * (⅔) + 0 * (⅓)
Therefore the expected error is (1 - ⅔) = 33%

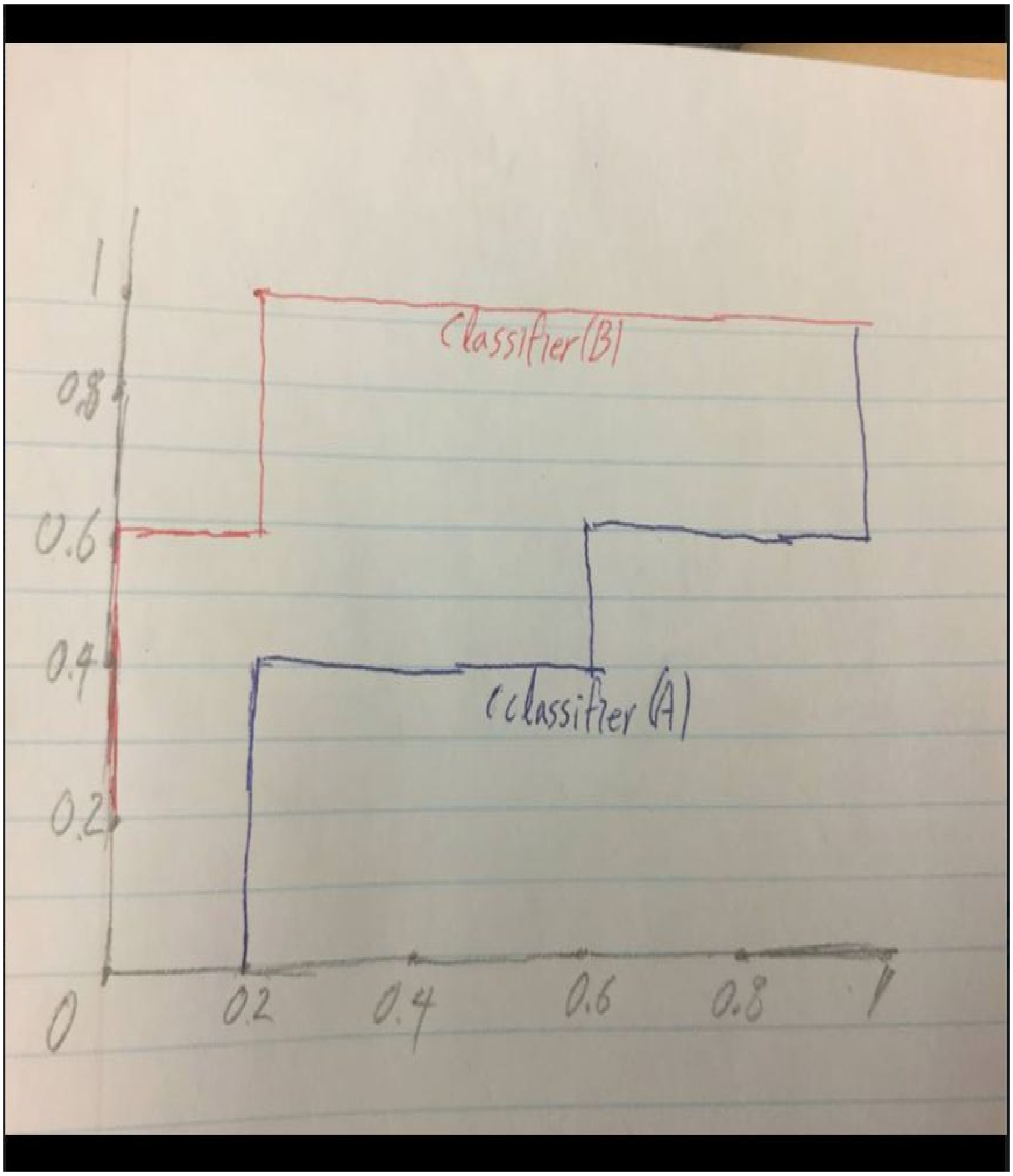
(d) Repeat the previous analysis assuming that the classifier predicts each test record to be positive class with probability 2/3 and negative class with probability 1/3

The probability that that a positive number is correct = ⅔
The probability that that a positive number is correct = ⅓
The probability that our test results is correct = ⅔ * ⅔ + ⅓ * ⅓ = 5/9
The probability that our test results is false (expected error) = 1 - 5/9 = 44.444%

.

3)

a) A has a bigger AUC and has a higher probability to rank a randomly chosen positive instance.

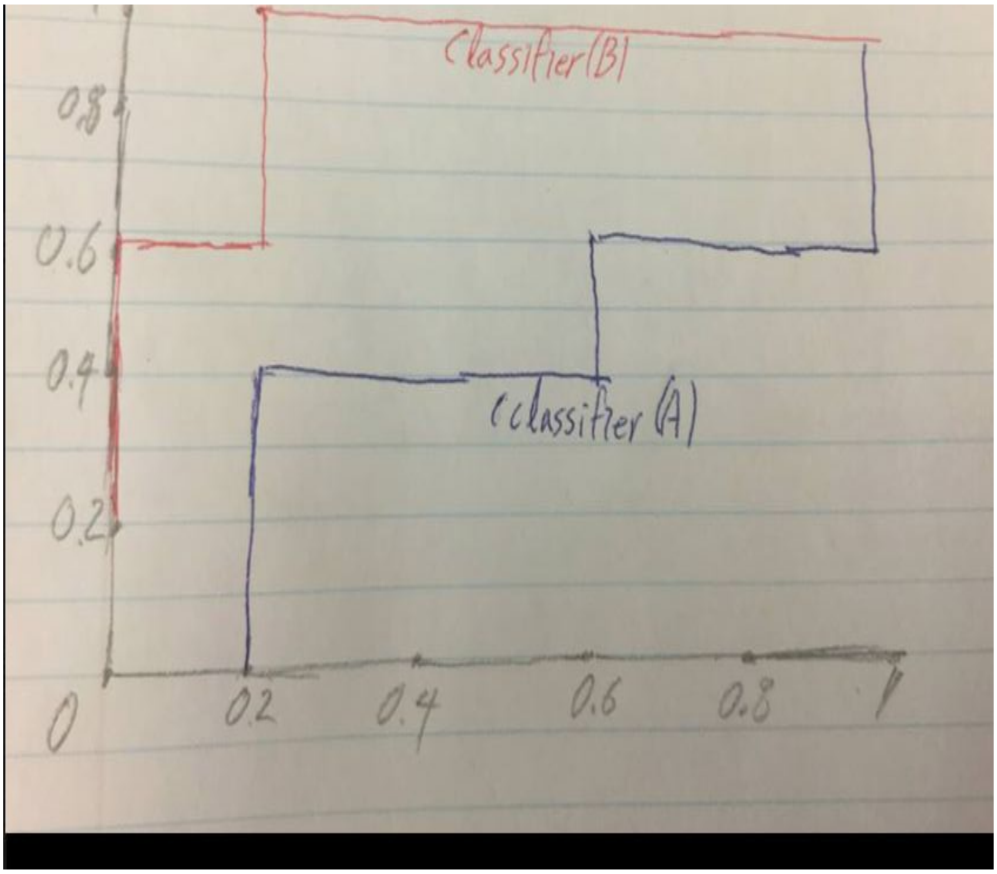


b) Precision: TP/(TP+FP) = 0.75
Recall: TP/P = 0.6
F-measure = 2/[(1/precision)+(1/TPR)] = 0.6667

c) Precision: TP/(TP+FP) = 0.5
Recall: TP/P = 0.2
F-measure = 2/[(1/precision)+(1/TPR)] = 0.2857

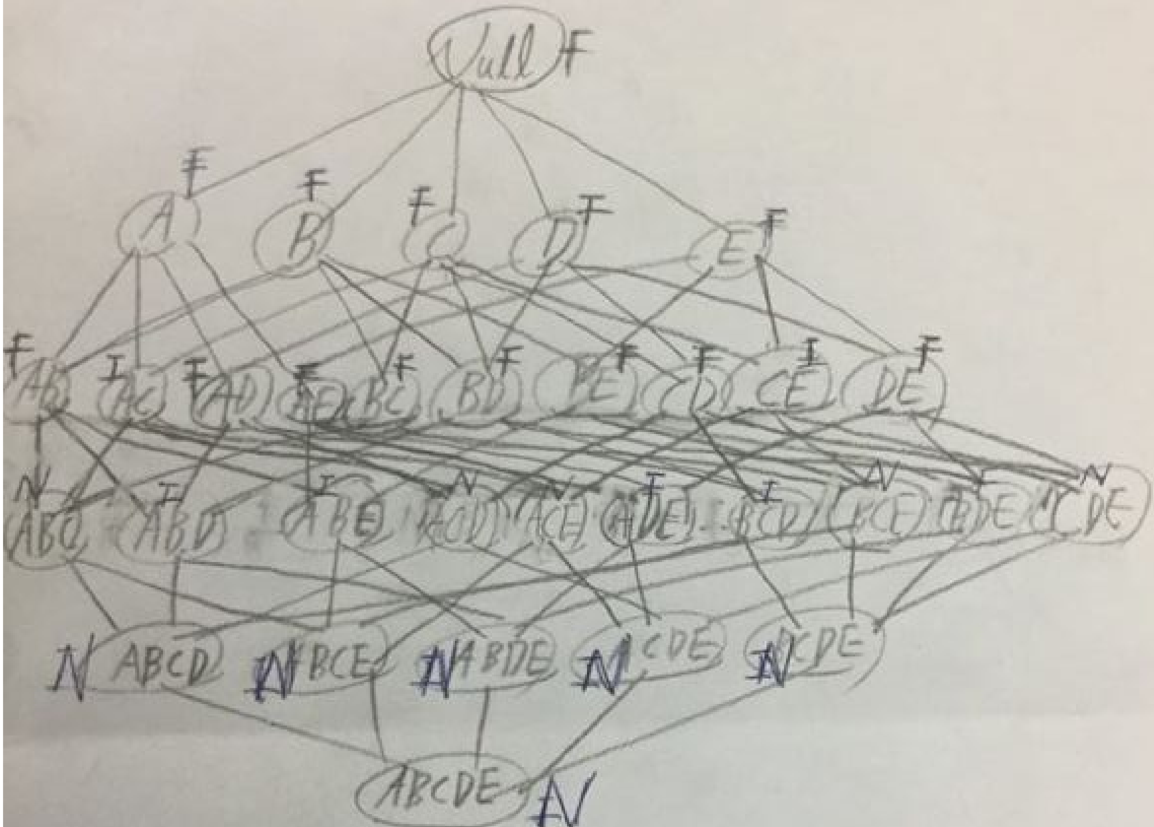
Since F-measure for A is higher than B, we conclude that A is better than B, and result is consistent with my ROC curve.

d)



we reduce the t under 0.5 for critical applications such as crime and health diagnosis because
It’s better to have safe result, and investigate more precisely.
On the other hand t = 0.5 is better when you don’t favor an answer over the other.

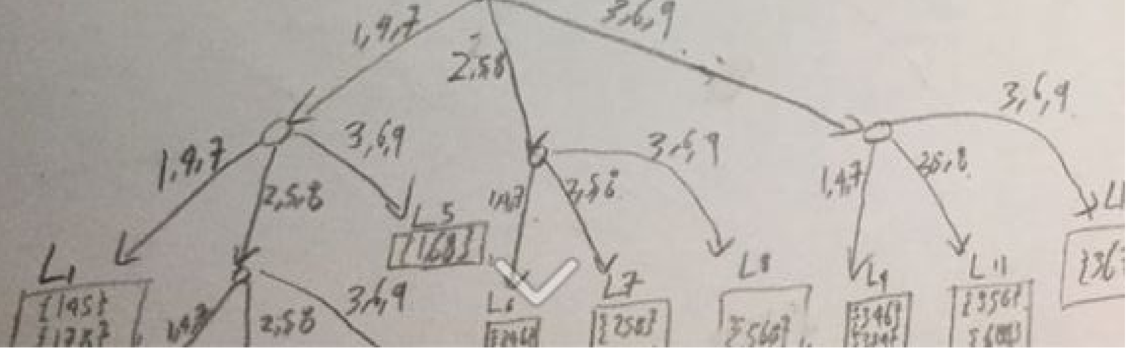
9) a) Draw on following lattice representing the data set, Label each Node with following letter.

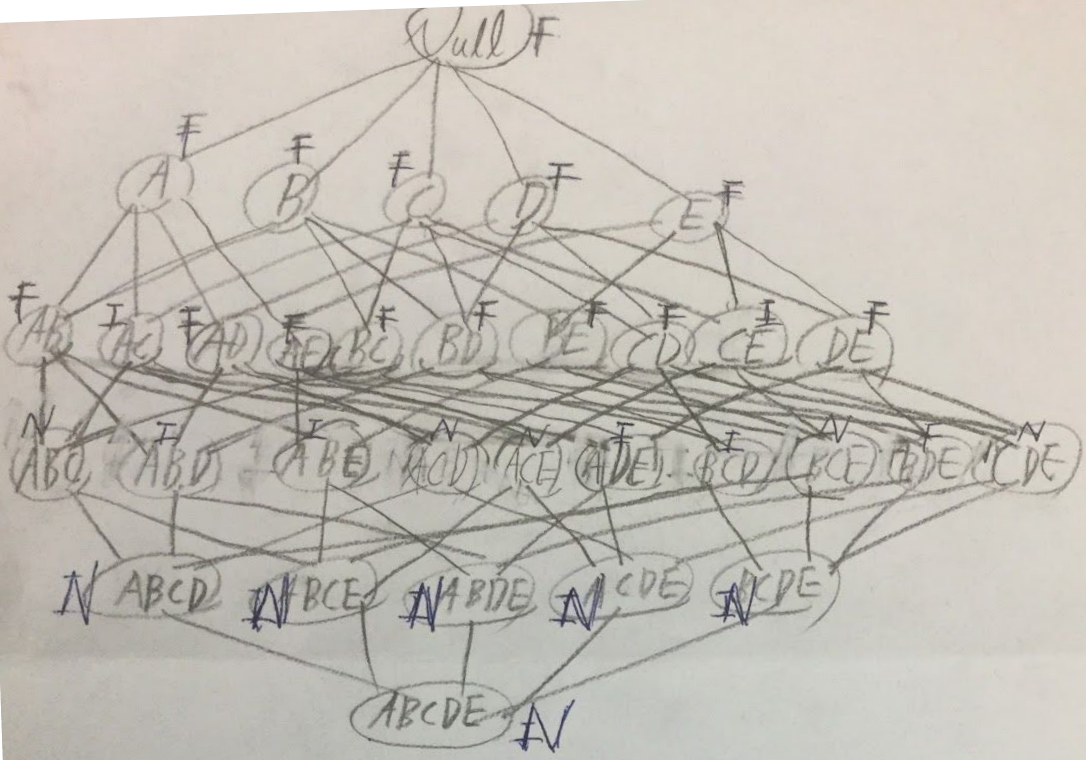


(b) what is the percentage of itemsets)

percentage of frequent itemsets = 16/32 = 50%

(c)

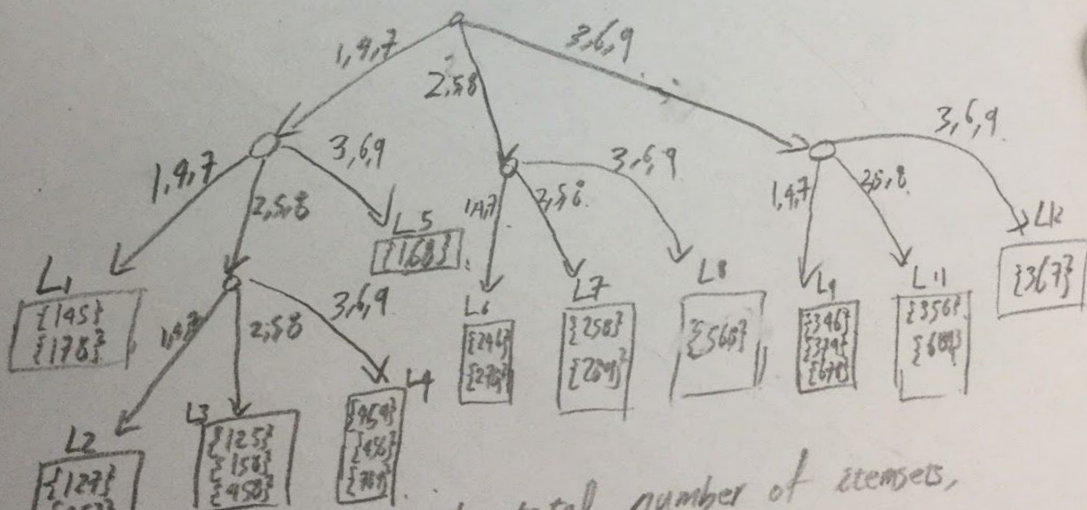




(b) What is the percentage of itemsets)

percentage of frequent itemsets = 16/32 = 50%.

(c)



Pruning ratio of N to the total number of itemsets,

N = 11, so the ratio is 11/32 = 34.4%

d)

False alarm rate is I/total number of sets

I = 5

false alarm is 5/32 = 15.6%

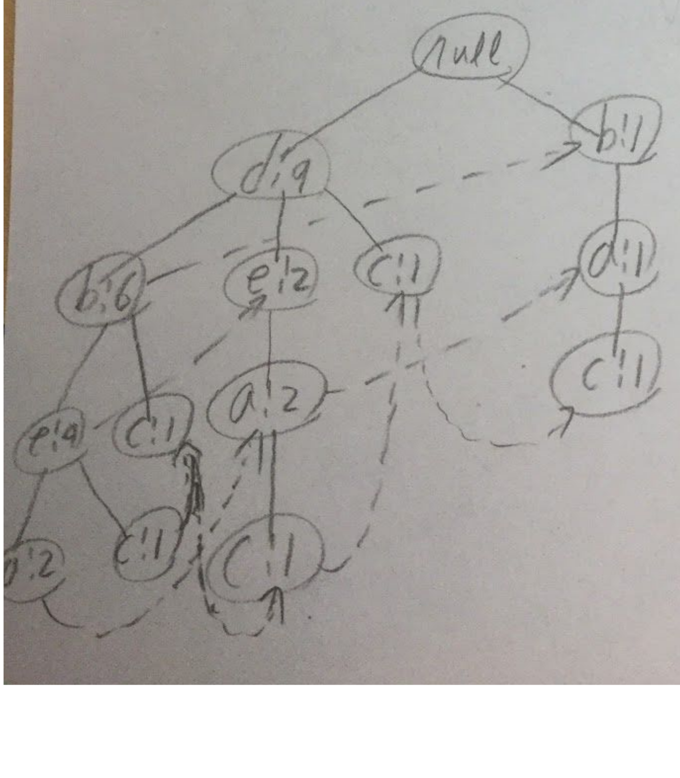
5)

First scan

TID	Items
1	{d,b,e,a}
2	{d,b,c}
3	{d,b,e,a}
4	{d,e,a,c}
5	{d,b,e,c}
6	{d,b,e}
7	{d,c}
8	{b,a,c}
9	{d,e,a}
10	{d,b}

d	9
A	7
C	7
D	5
E	3

Fp-Tree Construction

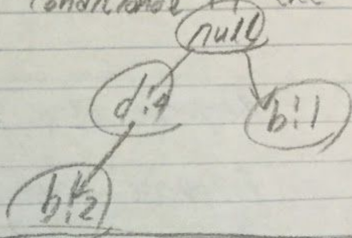


Suffix c: The basis of the condition pattern of "c", that is, the fragment of the tree containg the transaction for c (excluding graphics):

d,b,e : 1
d,b : 1
d,e,a : 1
d : 1
b,a : 1

Frequent items: d:4, b:3

Conditional FP-tree for "c"



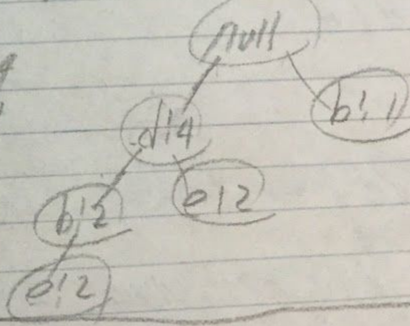
Suffix bc: Frequent itemset (FI) = {d,c} Conditional pattern base for "dc"
Nothing.
FIs so far = {c}: 5, {b,c}: 3, {d,c}: 4

Suffix a => Frequent itemset (FI) = {a}

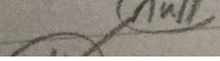
d,b,e : 2
d,e : 2
b : 1

Frequent items: d:4, e:4, b:3

Conditional FP-tree for "a":



Suffix ba: Frequent itemset (FI) = {b,a} Conditional pattern base for "ea": d:4, Frequent items: d:4
Conditional FP-tree for "ba":



Suffix ea: ⇒ frequent itemset (FI) = {e,a}
Conditional pattern base for "ea": d:4.
Frequent items: d:4.
Conditional FP-tree for "ea" null (one path tree)
d:4

Suffix "dea" ⇒ Frequent item set (FI) = {d,e,a}
Nothing.
Suffix "da" ⇒ Frequent item set (FI) = {d,a}
Nothing

Suffix "e" ⇒ Frequent itemset (FI) = {e}.
d, b : 4
d : 2

Frequent items, Conditional FP-tree for "e"
d:6
b:4

null
d:6
b:4

Suffix "be" ⇒ Frequent itemset (FI) = {be}
| Frequent items | Conditional FP-tree for "be":
d : 4

null
d:4

Suffix "de": ⇒ Frequent itemset (FI) = {de}.
Nothing
Suffix "bde" ⇒ Frequent itemset (FI) = {bde}.
Nothing
Suffix b: ⇒ Frequent itemset (FI) = {b}.
Frequent items: d-6 // Conditional FP-tree for "b":
Null
d:6

Suffix db ⇒ frequent itemset (FI) = {db}.
Nothing.
Suffix d ⇒ frequent itemset (FI) = {d}.
Nothing.

All frequent itemsets.

{c}:5, {b,c}:3, {d,c}:4
{a}:5, {b,a}:3, {e,a}:4, {d,e,a}:4, {d,a}:4.
{e}:6, {b,e}:4, {d,e}:6, {d,b,e}:4.
{b}:7, {d,b}:6.
{d}:9

6)

Choose the highest similarity, merge p2 and p5 together, use MIN to update the similarity matrix.

(P1,(P2,P5)) = MIN distance between a point in {P2,P5} and {P1} = MAX (sim(P2,P1),sim(P5,P1)) = MAX (0.1, 0.35) = 0.35 = sim (P5,P1)

	P1	P2,P5	P3	P4
P1	1	0.35	0.41	0.55
P2,P5		1	0.85	0.76
P3			1	0.44
P4				1

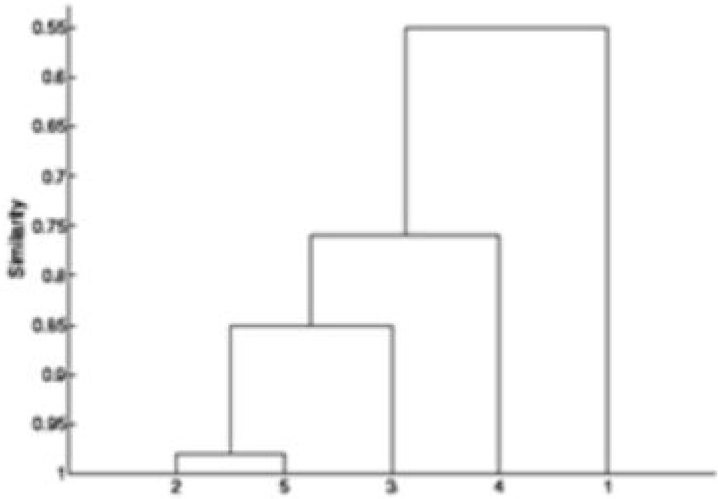
Choose the highest similarity, merge(P2,P3,P5) together, update the similarity matrix

	P1	P2,P5,P3	P4
P1	1	0.41	0.55
P2,P5,P3		1	0.76
P4			1

Choose the highest similarity, merge (P2,P5,P3, P4) together, update the similarity matrix

	P1	P2,P5,P3.P4
P1	1	0.41
P2,P5,P3,P4		1

Since only two clusters are left, merge everything and get the final dendrogram



Now we are using max,

Choose the highest similarity, merge(P2,P5).

(P1,(P2,P5)) = MAX distance between a point in {P2,P5} and {P1} =
Min(sim(p2,p1),sim(p5,p1)) = MIN(0.1,0.35) = 0.1

	P1	P2,P5	P3	P4
P1	1	0.1	0.41	0.55
P2,P5		1	0.64	0.47
P3			1	0.44
P4				1

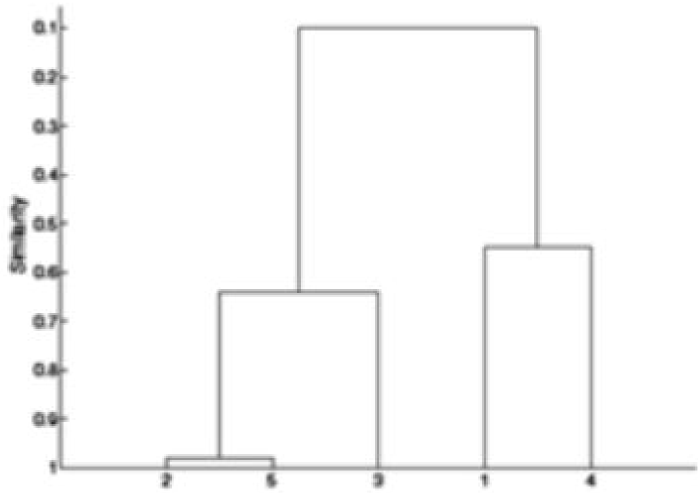
Choose the highest similarity, merge (P3,P2,P5) together, use Max to update the similarity matrix

	P1	P2,P5,P3	P4
P1	1	0.1	0.55
P2,P5,P3		1	0.44
P4			1

Choose the highest similarity, merge(P1,P4) use max to update the matrix

	P1,P4	P2,P5,P3
P1,P4	1	0.1
P2,P5,P3		1

Since only two clusters are left, we merge everything and get the final dendrogram.



7)
(a)

1) Lisa Rose,
X = [2.5,3.0,3.5,4.0], mean = 3.25, X' = [-0.75,-0.25,0.25,0.75]
Y = [2.5,3.5,3.5,3.0], mean = 3.125 Y' = [-0.625,0.375.0.375,-0.125]
sim = 0.375/(sqrt(1.25)*sqrt(0.6875)) = 0.4045

2) Mick Lasalle,

X = [2.5,3.0,3.5,4.0], mean = 3.25, X' = [-0.75,-0.25,0.25,0.75]
Y = [3.0,4.0,3.0,3.0], mean = 3.25 Y' = [-0.25,0.75.0.25,-0.25]
sim = -0.25/(sqrt(1.25)*sqrt(0.75)) = -0.2582

3) Toby,

X = [3.0,3.5], mean = 3.25, X’ = [-0.25,0.25]
Y = [4.5,4.0], mean = 3.25 Y’ = [0.25,-0.25]
sim = -0.125/(sqrt(0.125)*sqrt(0.125)) = -1

4) Gene Seymour,

X = [2.5,3.0,3.5,4.0], mean = 3.25, X’ = [-0.75,-0.25,0.25,0.75]
Y = [3.0,3.5,5.0,3.0], mean = 3.625 Y’ = [-0.625,-0.125,1.375,-0.625]
sim = 0.375/(sqrt(1.25)*sqrt(2.6875)) = 0.2046

5) Claudia Puig,

X = [3.0,3.5,4.0], mean = 3.5, X’ = [-0.5,0,0.5]
Y = [3.5,4.0,4.5], mean = 4 Y’ = [-0.5,0,0.5]
sim = 0.5/(sqrt(0.5)*sqrt(0.5)) = 1

6) Jack Matthews,

X = [2.5,3.0,3.5,4.0], mean = 3.25, X’ = [-0.75,-0.25,0.25,0.75]
Y = [3.0,4.0,5.0,3.0], mean = 3.75 Y’ = [-0.75,0.25.1.25,-0.75]
sim = 0.25/(sqrt(1.25)*sqrt(2.75)) = 0.1348

Claudia Puig has the strongest similarity with Michael

r = (2.5 * 0.4045 +3.5*0.2046+2.5*1 + 3.5 * 0.1348)/1.7439
= 2.7

b)
Population mean =
((2.5+3.5+3+3.5+2.5+3) + (2.5+3+3.5+4) + (3+4+2+3+3+2) + (4.5+4+ 1) +
(3+3.5+1.5+5+3+3.5)+(3.5+3+4.5+4+2.5)+(3+4+5+3+3.5))/35
= 3.23

1) Lady in the Water:
e^2 = (2.5 - 3.23 -0 -0)^2 = 0.047
bi = 0
bu = 0+0.1*(0.5329-0.1*00 = 0.0533

2) Snakes on a Plane
e^2 = (3.5-3.23-0.0533-0)^2 = 0.047
bu = 0.0533+0.1*(0.047 - 0.1 *0.0533) = 0.0575

3) Just my Luck
e^2 = (3 - 3.23 - 0.0575 - 0)^2 = 0.0827
bu = 0.0575 + 0.1*(0.0827 - 0.1* 0.0575) = 0.0652

4) Superman Returns

e^2 = (3.5 -3.23 - 0.0652 - 0)^2 = 0.0419
bu = 0.0652 + 0.1*(0.0419 - 0.1* 0.0652) = 0.0687

5) You Me and Dupree

e^2 = (2.5 - 3.23 - 0.0687 - 0)^2 = 0.6379
bu = 0.0687 + 0.1*(0.6379 - 0.1 *0.0687) = 0.1318

6) The night listener

e^2 = (3 - 3.23 -0.1318 - 0)^2 = 0.1309
bu = 0.1318 + 0.1*(0.1309 - 0.1 * 0.1318) = 0.1436