Problem 1

P(senior): 52/165=0.315
P(junior): 113/165=0.685

P(systems and senior) 8/52=0.154
P(26_30 and senior) 1/58 =0.017
P(46k_50k and senior) 40/52 =0.769

P(systems and junior) 23/113=0.204
P(26_30 and junior) 49/113=0.434
p(46k_50k and junior) 23/113=0.204

Senior:

.315*.154*.017*.769= 0.00063

Junior:

.685*.204*.434*.204= 0.124

Problem 3 a)

Centroid 1 h=(4,9)
Centroid 2 c=(8,4)

Point a
C1 |2-4| + |10-9| = 3
C2 |2-8| + |10-4| = 12
Closer to h

Point b
C1 |2-4| + |5-9| = 6
C2 |2-8| + |5-4| = 7
Closer to h

Point c = c

Point d
|5-4| + |8-9| = 2
|5-8| + |8-4| = 7
Closer to h

Point e
|7-4| + |5-9| = 7
|7-8| + |5-4| = 2

Closer to c

Point f
|6-4| + |4-9| = 7
|6-8| + |4-4| = 2
Closer to c

Point g
|1-4| + |2-9| = 10
|1-8| + |2-4| = 9
Closer to c

Point h = h

Cluster 1 = a,b,d,h
Cluser 2 = c,e,f,g

Cluster 1 mean
X1 (2+2+5+4)/4 = 3.25
X2 (10+5+8+9)/4 = 8
C1 = (3.25,8)

Cluster 2 mean
X1 (8+7+6+1)/4 = 5.5
X2 (4+5+4+2)/4 = 3.75
C2 = (5.5,3.75)

b)

Point a
|2-3.25| + |10-8| = 3.25
|2-5.5| + |10-3.75| = 9.75
Closer to new c1

Point b
|2-3.25| + |5-8| = 4.25
|2-5.5| + |5-3.75| =4.75
Closer to new c1

Point c
|8-3.25| + |4-8| = 8.75
|8-5.5| + |4-3.75 = 2.75
Closer to new c2 new

Point d

|5-3.25| + |8-8| = 1.75
|5-5.5| + |8-3.75| = 4.75
Closer to c1

Point e
|7-3.25| + |5-8| = 6.75
|7-5.5| + |5-3.75| = 2.75
Closer to c2 new

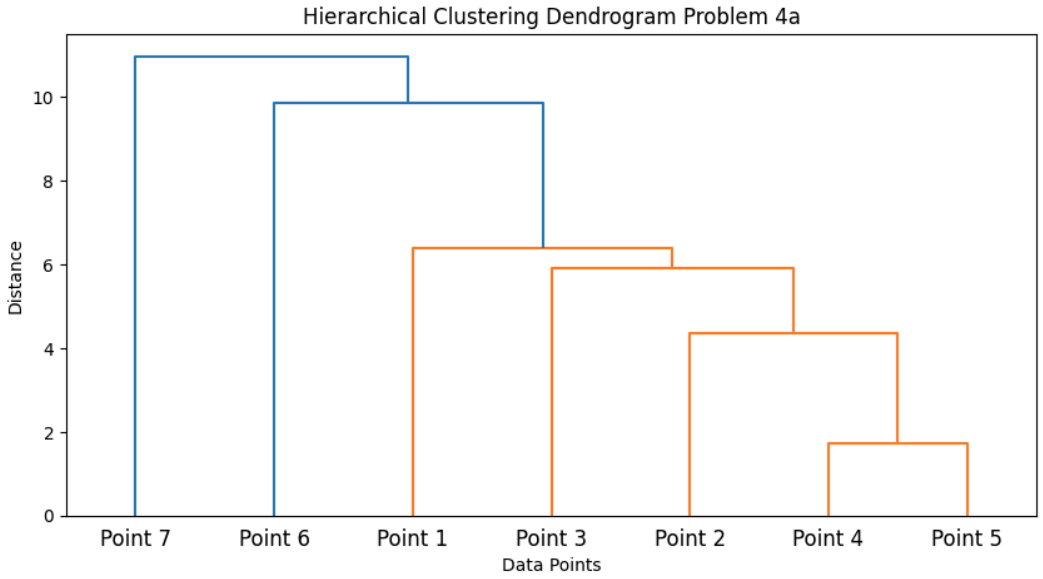
Point f
|6-3.25| + |4-8| = 6.75
|6-5.5| + |4-3.75| = .75
Closer to c2 new

Point g
|1-3.25| + |2-8| = 8.25
|1-5.5| + |2-3.75| = 6.25
Closer to c2

Point h
|4-3.25| + |9-8| = 1.75
|4 - 5.5| + | 9-3.75| = 6.75

Final cluset
C1 = a,b,d,h
C2 = c,e,f,g

Problem 4a)


Hierarchical Clustering Dendrogram Problem 4a
Distance
0
2
4
6
8
10
Point 7
Point 6
Point 1
Point 3
Point 2
Point 4
Point 5
Data Points