Homework 7 G

December 7, 2021

Jose Carlos Munoz

19.a)

19.b)

19.f)

24)

The possible pairs are the sets {P1,P2},{P1,P3},{P1,P4},{P2,P3},{P2,P4},{P3,P4}.
Based on the ideal similarity matrix we get the set x= {1,0,0,0,0,1}.
In the similarity matrix the we get the set y = {0.8,0.65,0.55,0.7,0.6,0.9}.
The σ_x is 0.5164 and σ_y is 0.1304.The cov(x,y) = 0.06.
To find the correlation its $\frac{cov(x,y)}{\sigma_x * \sigma_y}$.
So the correlation value is 0.08910.

25)

To find the F(i,j) value we first find the R(i,j) and P(i,j).R(i,j) is equal to $\frac{n_{ij}}{n_i}$. Where n_{ij} is the amount of class a in the cluster and n_i is how many class values over all.P(i,j) is equal to $\frac{n_{ij}}{n_j}$. Where n_{ij} is the amount of class a in the cluster and n_i is how many values in the cluster. F(i,j) is equal to $2 * R(i,j) * \frac{P(i,j)}{P(i,j)+R(i,j)}$ For Cluster 1

For Class A
R(A,1) = $\frac{3}{3} = 1$, P(A,1) = $\frac{3}{8}$
F(A,1) = $2 * 1\frac{1}{1+3/8} = 0.55$
For Class B
R(B,1) = $\frac{5}{5} = 1$, P(B,1) = $\frac{5}{8}$
F(B,1) = $2 * 1\frac{1}{1+5/8} = 0.77$

For Cluster 2

For Class A
R(A,2) = $\frac{2}{3}$, P(A,2) = $\frac{2}{4}$, F(A,2) = 0.57
For Class B
R(B,2) = $\frac{2}{5}$, P(B,2) = $\frac{2}{4}$, F(B,2) = 0.44

For Cluster 3

For Class A
R(A,3) = $\frac{1}{3}$, P(A,3) = $\frac{1}{4}$, F(A,3) = 0.29
For Class B
R(B,3) = $\frac{3}{5}$, P(B,3) = $\frac{3}{4}$, F(B,3) = 0.67

For Cluster 4

For Class A
R(A,4) = $\frac{2}{3}$, P(A,4) = $\frac{2}{2}$, F(A,4) = 0.80
For Class B
R(B,4) = $\frac{0}{5}$, P(B,4) = $\frac{0}{4}$, F(B,4) = 0.00

For Cluster 5

For Class A
R(A,5) = $\frac{0}{3}$, P(A,5) = $\frac{0}{2}$, F(A,5) = 0.00
For Class B
R(B,5) = $\frac{2}{5}$, P(B,5) = $\frac{2}{2}$, F(B,5) = 0.57

For Cluster 6

For Class A

R(A,6) = $\frac{1}{3}$, P(A,6) = $\frac{1}{2}$, F(A,6) = 0.40

For Class B

R(B,6) = $\frac{1}{5}$, P(B,6) = $\frac{1}{2}$, F(B,6) = 0.29

For Cluster 7

For Class A

R(A,7) = $\frac{0}{3}$, P(A,7) = $\frac{0}{2}$, F(A,7) = 0.00

For Class B

R(B,7) = $\frac{2}{5}$, P(B,7) = $\frac{2}{2}$, F(B,7) = 0.57

For Overall Clustering we have to use the Max F(A) and F(B) values which are 0.8 and 0.77 respectively.

The value is $\frac{3}{8} * 0.8 + \frac{5}{8} * 0.77$ which is 0.78

26)