

CS4445 A Term 2008 Homework 4 Solutions

Contents

1 Instance Based Learning	3
1.1 Original Data	3
1.1.1 Constant Weights	3
1.1.1.1 Instance i_1	3
1.1.1.2 Instance i_2	4
1.1.2 Prediction Summary	4
1.1.3 Error Measures	5
1.1.3.1 Root-Mean Squared Error	5
1.1.3.2 Mean Absolute Error	5
1.1.4 Error Summary	5
1.1.5 Inverse Distance Weights	5
1.1.5.1 Instance i_1	5
1.1.5.2 Instance i_2	6
1.1.6 Prediction Summary	6
1.1.7 Error Measures	6
1.1.7.1 Root-Mean Squared Error	6
1.1.7.2 Mean Absolute Error	7
1.1.8 Error Summary	7
1.2 Normalized Data	7
1.2.1 Normalization	7
1.2.2 Constant Weights	8
1.2.2.1 Instance i_1	8
1.2.2.2 Instance i_2	8
1.2.3 Prediction Summary	9
1.2.4 Error Measures	9
1.2.4.1 Root-Mean Squared Error	9
1.2.4.2 Mean Absolute Error	9
1.2.5 Error Summary	10
1.2.6 Inverse Distance Weights	10
1.2.6.1 Instance i_1	10
1.2.6.2 Instance i_2	10
1.2.7 Prediction Summary	11
1.2.8 Error Measures	11
1.2.8.1 Root-Mean Squared Error	11
1.2.8.2 Mean Absolute Error	11
1.2.9 Error Summary	12
1.3 Best Method	12

2 Clustering - Simple K-means	12
2.1 Original Data	12
2.1.1 Cluster Formation	12
2.1.1.1 Instance 1	12
2.1.1.2 Instance 2	12
2.1.1.3 Instance 3	13
2.1.1.4 Instance 4	13
2.1.1.5 Instance 5	13
2.1.1.6 Instance 6	13
2.1.1.7 Instance 7	14
2.1.1.8 Instance 8	14
2.1.1.9 Instance 9	14
2.1.1.10 Instance 10	14
2.1.2 Cluster Refinement	15
2.1.2.1 Cluster c_1	15
2.1.2.2 Cluster c_2	15
2.1.3 Cluster Formation	15
2.1.3.1 Instance 1	15
2.1.3.2 Instance 2	16
2.1.3.3 Instance 3	16
2.1.3.4 Instance 4	16
2.1.3.5 Instance 5	16
2.1.3.6 Instance 6	17
2.1.3.7 Instance 7	17
2.1.3.8 Instance 8	17
2.1.3.9 Instance 9	17
2.1.3.10 Instance 10	18
2.1.4 Cluster Refinement	18
2.1.4.1 Cluster c_1	18
2.1.4.2 Cluster c_2	18
2.1.5 Final Clusters	19
2.1.5.1 Cluster c_1	19
2.1.5.2 Cluster c_2	19
2.2 Normalized Data	19
2.2.1 Normalization	19
2.2.2 Cluster Formation	20
2.2.2.1 Instance 1	20
2.2.2.2 Instance 2	20
2.2.2.3 Instance 3	21
2.2.2.4 Instance 4	21
2.2.2.5 Instance 5	21
2.2.2.6 Instance 6	21
2.2.2.7 Instance 7	22
2.2.2.8 Instance 8	22
2.2.2.9 Instance 9	22

2.2.2.10 Instance 10 22

2.2.3 Cluster Refinement 23

2.2.3.1 Cluster c_1 23

2.2.3.2 Cluster c_2 23

2.2.4 Cluster Formation 23

2.2.4.1 Instance 1 23

2.2.4.2 Instance 2 24

2.2.4.3 Instance 3 24

2.2.4.4 Instance 4 24

2.2.4.5 Instance 5 24

2.2.4.6 Instance 6 25

2.2.4.7 Instance 7 25

2.2.4.8 Instance 8 25

2.2.4.9 Instance 9 25

2.2.4.10 Instance 10 26

2.2.5 Cluster Refinement 26

2.2.5.1 Cluster c_1 26

2.2.5.2 Cluster c_2 26

2.2.6 Final Clusters 27

2.2.6.1 Cluster c_1 27

2.2.6.2 Cluster c_2 27

3 Clustering - Hierarchical Clustering **27**

3.1 Number of Alternatives 27

3.2 Alternatives 27

3.3 Choice 29

1 Instance Based Learning

1.1 Original Data

1.1.1 Constant Weights

We have the following instances to evaluate:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ
i_1	6.48	female	1034	62.143	127
i_2	6.59	male	1100	88.452	114

1.1.1.1 Instance i_1

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ
i_1	6.48	female	1034	62.143	127

The training instances sorted by distance to the test instance are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
8	6.03	male	1029	81.648	97	20.166
1	6.08	female	1005	57.607	96	29.355
6	6.43	female	1070	83.009	126	41.610
5	6.84	female	1079	107.503	96	63.896
2	5.73	female	963	58.968	89	71.075
10	7.67	male	1160	72.576	124	126.441
9	7.52	male	1204	79.38	113	170.878
4	8.42	female	1272	61.69	103	238.008
3	7.99	female	1281	63.958	101	247.011
7	7.6	male	1347	97.524	94	314.997

We take the first 4 values for the **FIQ** attribute and average them (with proper weights):

$$\frac{97.000 * 1.000 + 96.000 * 1.000 + 126.000 * 1.000 + 96.000 * 1.000}{1.000 + 1.000 + 1.000 + 1.000} = \frac{415.000}{4.000} = 103.750$$

Thus our prediction for **FIQ** for this test instance is 103.750.

1.1.1.2 Instance i_2

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
i_2	6.59	male	1100	88.452	114

The training instances sorted by distance to the test instance are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
5	6.84	female	1079	107.503	96	28.373
6	6.43	female	1070	83.009	126	30.507
10	7.67	male	1160	72.576	124	62.074
8	6.03	male	1029	81.648	97	71.327
1	6.08	female	1005	57.607	96	99.888
9	7.52	male	1204	79.38	113	104.399
2	5.73	female	963	58.968	89	140.143
4	8.42	female	1272	61.69	103	174.082
3	7.99	female	1281	63.958	101	182.658
7	7.6	male	1347	97.524	94	247.169

We take the first 4 values for the **FIQ** attribute and average them (with proper weights):

$$\frac{96.000 * 1.000 + 126.000 * 1.000 + 124.000 * 1.000 + 97.000 * 1.000}{1.000 + 1.000 + 1.000 + 1.000} = \frac{443.000}{4.000} = 110.750$$

Thus our prediction for **FIQ** for this test instance is 110.750.

1.1.2 Prediction Summary

Our method made the following predictions:

#	FIQ	Constant Weights
i_1	127	103.750
i_2	114	110.750

1.1.3 Error Measures

1.1.3.1 Root-Mean Squared Error

For the root-mean squared error, we use the following formula from the class text:

$$\left(\frac{(p_1 - a_1)^2 + \dots + (p_n - a_n)^2}{n} \right)^{\frac{1}{2}}$$

$$\left(\frac{(103.750 - 127.000)^2 + (110.750 - 114.000)^2}{2} \right)^{\frac{1}{2}} = 16.600$$

1.1.3.2 Mean Absolute Error

For the mean absolute error, we use the following formula from the class text:

$$\frac{|p_1 - a_1| + \dots + |p_n - a_n|}{n}$$

Constant Weights

$$\frac{|103.750 - 127.000| + |110.750 - 114.000|}{2} = 13.250$$

1.1.4 Error Summary

And we have our accuracy measures:

measure	Constant Weights
root mean-squared error	16.600
mean absolute error	13.250

1.1.5 Inverse Distance Weights

We have the following instances to evaluate:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
i_1	6.48	female	1034	62.143	127
i_2	6.59	male	1100	88.452	114

1.1.5.1 Instance i_1

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
i_1	6.48	female	1034	62.143	127

The training instances sorted by distance to the test instance are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
8	6.03	male	1029	81.648	97	20.166
1	6.08	female	1005	57.607	96	29.355
6	6.43	female	1070	83.009	126	41.610
5	6.84	female	1079	107.503	96	63.896
2	5.73	female	963	58.968	89	71.075
10	7.67	male	1160	72.576	124	126.441
9	7.52	male	1204	79.38	113	170.878
4	8.42	female	1272	61.69	103	238.008
3	7.99	female	1281	63.958	101	247.011
7	7.6	male	1347	97.524	94	314.997

We take the first 4 values for the **FIQ** attribute and average them (with proper weights):

$$\frac{97.000 * \frac{1}{20.166} + 96.000 * \frac{1}{29.355} + 126.000 * \frac{1}{41.610} + 96.000 * \frac{1}{63.896}}{\frac{1}{20.166} + \frac{1}{29.355} + \frac{1}{41.610} + \frac{1}{63.896}} = \frac{12.611}{0.123} = 102.248$$

Thus our prediction for **FIQ** for this test instance is 102.248.

1.1.5.2 Instance i_2

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ
i_2	6.59	male	1100	88.452	114

The training instances sorted by distance to the test instance are:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
5	6.84	female	1079	107.503	96	28.373
6	6.43	female	1070	83.009	126	30.507
10	7.67	male	1160	72.576	124	62.074
8	6.03	male	1029	81.648	97	71.327
1	6.08	female	1005	57.607	96	99.888
9	7.52	male	1204	79.38	113	104.399
2	5.73	female	963	58.968	89	140.143
4	8.42	female	1272	61.69	103	174.082
3	7.99	female	1281	63.958	101	182.658
7	7.6	male	1347	97.524	94	247.169

We take the first 4 values for the **FIQ** attribute and average them (with proper weights):

$$\frac{96.000 * \frac{1}{28.373} + 126.000 * \frac{1}{30.507} + 124.000 * \frac{1}{62.074} + 97.000 * \frac{1}{71.327}}{\frac{1}{28.373} + \frac{1}{30.507} + \frac{1}{62.074} + \frac{1}{71.327}} = \frac{10.871}{0.098} = 110.757$$

Thus our prediction for **FIQ** for this test instance is 110.757.

1.1.6 Prediction Summary

Our method made the following predictions:

#	FIQ	Inverse Distance Weights
i_1	127	102.248
i_2	114	110.757

1.1.7 Error Measures

1.1.7.1 Root-Mean Squared Error

For the root-mean squared error, we use the following formula from the class text:

$$\left(\frac{(p_1 - a_1)^2 + \dots + (p_n - a_n)^2}{n} \right)^{\frac{1}{2}}$$

$$\left(\frac{(102.248 - 127.000)^2 + (110.757 - 114.000)^2}{2} \right)^{\frac{1}{2}} = 17.652$$

1.1.7.2 Mean Absolute Error

For the mean absolute error, we use the following formula from the class text:

$$\frac{|p_1 - a_1| + \dots + |p_n - a_n|}{n}$$

Inverse Distance Weights

$$\frac{|102.248 - 127.000| + |110.757 - 114.000|}{2} = 13.998$$

1.1.8 Error Summary

And we have our accuracy measures:

measure	Inverse Distance Weights
root mean-squared error	17.652
mean absolute error	13.998

1.2 Normalized Data

1.2.1 Normalization

We begin with normalizing training and test instances. We normalize both together and justify the use of the test instances in this process by the fact that the values derived (min and max of attributes) should have been derived from domain knowledge instead of instances to begin with. We find the minimum and maximum values for all of our non-target numeric attributes:

- **CCMIDSA**: minimum: 5.730, maximum: 8.420
- **GENDER**: This is not a numeric attribute so we leave it alone.
- **TOTVOL**: minimum: 963.000, maximum: 1347.000
- **WEIGHT**: minimum: 57.607, maximum: 107.503
- **FIQ**: This is the target attribute so we leave it alone.

Having the min and max values, we will apply the following transformation on each non-target numeric attribute:

$$new = \frac{old - min}{max - min}$$

Doing so produces the modified training set:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
1	0.130	female	0.109	0.000	96
2	0.000	female	0.000	0.027	89
3	0.840	female	0.828	0.127	101
4	1.000	female	0.805	0.082	103
5	0.413	female	0.302	1.000	96
6	0.260	female	0.279	0.509	126
7	0.695	male	1.000	0.800	94
8	0.112	male	0.172	0.482	97
9	0.665	male	0.628	0.436	113
10	0.721	male	0.513	0.300	124

We use the same min/max values we calculated to transform the test instances. This results in a modified set of test instances:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
i_1	0.279	female	0.185	0.091	127
i_2	0.320	male	0.357	0.618	114

We now repeat the testing process on these modified instances.

1.2.2 Constant Weights

We have the following instances to evaluate:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
i_1	0.279	female	0.185	0.091	127
i_2	0.320	male	0.357	0.618	114

1.2.2.1 Instance i_1

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
i_1	0.279	female	0.185	0.091	127

The training instances sorted by distance to the test instance are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
1	0.130	female	0.109	0.000	96	0.190
2	0.000	female	0.000	0.027	89	0.341
6	0.260	female	0.279	0.509	126	0.429
3	0.840	female	0.828	0.127	101	0.854
5	0.413	female	0.302	1.000	96	0.926
4	1.000	female	0.805	0.082	103	0.951
8	0.112	male	0.172	0.482	97	1.087
10	0.721	male	0.513	0.300	124	1.160
9	0.665	male	0.628	0.436	113	1.210
7	0.695	male	1.000	0.800	94	1.530

We take the first 4 values for the **FIQ** attribute and average them (with proper weights):

$$\frac{96.000 * 1.000 + 89.000 * 1.000 + 126.000 * 1.000 + 101.000 * 1.000}{1.000 + 1.000 + 1.000 + 1.000} = \frac{412.000}{4.000} = 103.000$$

Thus our prediction for **FIQ** for this test instance is 103.000.

1.2.2.2 Instance i_2

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
i_2	0.320	male	0.357	0.618	114

The training instances sorted by distance to the test instance are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
8	0.112	male	0.172	0.482	97	0.310
9	0.665	male	0.628	0.436	113	0.475
10	0.721	male	0.513	0.300	124	0.535
7	0.695	male	1.000	0.800	94	0.766
6	0.260	female	0.279	0.509	126	1.011
5	0.413	female	0.302	1.000	96	1.076
1	0.130	female	0.109	0.000	96	1.216
2	0.000	female	0.000	0.027	89	1.257
3	0.840	female	0.828	0.127	101	1.317
4	1.000	female	0.805	0.082	103	1.397

We take the first 4 values for the **FIQ** attribute and average them (with proper weights):

$$\frac{97.000 * 1.000 + 113.000 * 1.000 + 124.000 * 1.000 + 94.000 * 1.000}{1.000 + 1.000 + 1.000 + 1.000} = \frac{428.000}{4.000} = 107.000$$

Thus our prediction for **FIQ** for this test instance is 107.000.

1.2.3 Prediction Summary

Our method made the following predictions:

#	FIQ	Constant Weights
i_1	127	103.000
i_2	114	107.000

1.2.4 Error Measures

1.2.4.1 Root-Mean Squared Error

For the root-mean squared error, we use the following formula from the class text:

$$\left(\frac{(p_1 - a_1)^2 + \dots + (p_n - a_n)^2}{n} \right)^{\frac{1}{2}}$$

$$\left(\frac{(103.000 - 127.000)^2 + (107.000 - 114.000)^2}{2} \right)^{\frac{1}{2}} = 17.678$$

1.2.4.2 Mean Absolute Error

For the mean absolute error, we use the following formula from the class text:

$$\frac{|p_1 - a_1| + \dots + |p_n - a_n|}{n}$$

Constant Weights

$$\frac{|103.000 - 127.000| + |107.000 - 114.000|}{2} = 15.500$$

1.2.5 Error Summary

And we have our accuracy measures:

measure	Constant Weights
root mean-squared error	17.678
mean absolute error	15.500

1.2.6 Inverse Distance Weights

We have the following instances to evaluate:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ
i_1	0.279	female	0.185	0.091	127
i_2	0.320	male	0.357	0.618	114

1.2.6.1 Instance i_1

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ
i_1	0.279	female	0.185	0.091	127

The training instances sorted by distance to the test instance are:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
1	0.130	female	0.109	0.000	96	0.190
2	0.000	female	0.000	0.027	89	0.341
6	0.260	female	0.279	0.509	126	0.429
3	0.840	female	0.828	0.127	101	0.854
5	0.413	female	0.302	1.000	96	0.926
4	1.000	female	0.805	0.082	103	0.951
8	0.112	male	0.172	0.482	97	1.087
10	0.721	male	0.513	0.300	124	1.160
9	0.665	male	0.628	0.436	113	1.210
7	0.695	male	1.000	0.800	94	1.530

We take the first 4 values for the **FIQ** attribute and average them (with proper weights):

$$\frac{96.000 * \frac{1}{0.190} + 89.000 * \frac{1}{0.341} + 126.000 * \frac{1}{0.429} + 101.000 * \frac{1}{0.854}}{\frac{1}{0.190} + \frac{1}{0.341} + \frac{1}{0.429} + \frac{1}{0.854}} = \frac{1177.306}{11.688} = 100.729$$

Thus our prediction for **FIQ** for this test instance is 100.729.

1.2.6.2 Instance i_2

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ
i_2	0.320	male	0.357	0.618	114

The training instances sorted by distance to the test instance are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
8	0.112	male	0.172	0.482	97	0.310
9	0.665	male	0.628	0.436	113	0.475
10	0.721	male	0.513	0.300	124	0.535
7	0.695	male	1.000	0.800	94	0.766
6	0.260	female	0.279	0.509	126	1.011
5	0.413	female	0.302	1.000	96	1.076
1	0.130	female	0.109	0.000	96	1.216
2	0.000	female	0.000	0.027	89	1.257
3	0.840	female	0.828	0.127	101	1.317
4	1.000	female	0.805	0.082	103	1.397

We take the first 4 values for the **FIQ** attribute and average them (with proper weights):

$$\frac{97.000 * \frac{1}{0.310} + 113.000 * \frac{1}{0.475} + 124.000 * \frac{1}{0.535} + 94.000 * \frac{1}{0.766}}{\frac{1}{0.310} + \frac{1}{0.475} + \frac{1}{0.535} + \frac{1}{0.766}} = \frac{905.413}{8.507} = 106.432$$

Thus our prediction for **FIQ** for this test instance is 106.432.

1.2.7 Prediction Summary

Our method made the following predictions:

#	FIQ	Inverse Distance Weights
i_1	127	100.729
i_2	114	106.432

1.2.8 Error Measures

1.2.8.1 Root-Mean Squared Error

For the root-mean squared error, we use the following formula from the class text:

$$\left(\frac{(p_1 - a_1)^2 + \dots + (p_n - a_n)^2}{n} \right)^{\frac{1}{2}}$$

$$\left(\frac{(100.729 - 127.000)^2 + (106.432 - 114.000)^2}{2} \right)^{\frac{1}{2}} = 19.332$$

1.2.8.2 Mean Absolute Error

For the mean absolute error, we use the following formula from the class text:

$$\frac{|p_1 - a_1| + \dots + |p_n - a_n|}{n}$$

Inverse Distance Weights

$$\frac{|100.729 - 127.000| + |106.432 - 114.000|}{2} = 16.920$$

1.2.9 Error Summary

And we have our accuracy measures:

measure	Inverse Distance Weights
root mean-squared error	19.332
mean absolute error	16.920

1.3 Best Method

The error measures tell us that the regular method with constant weights did best.

2 Clustering - Simple K-means

2.1 Original Data

We begin clustering with our original (not normalized) data:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
1	6.08	female	1005	57.607	96
2	5.73	female	963	58.968	89
3	7.99	female	1281	63.958	101
4	8.42	female	1272	61.69	103
5	6.84	female	1079	107.503	96
6	6.43	female	1070	83.009	126
7	7.6	male	1347	97.524	94
8	6.03	male	1029	81.648	97
9	7.52	male	1204	79.38	113
10	7.67	male	1160	72.576	124

2.1.1 Cluster Formation

Let us assign each of our instances to a cluster.

2.1.1.1 Instance 1

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
1	6.08	female	1005	57.607	96

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_1	6.48	female	1034	62.143	127	42.694
c_2	6.59	male	1100	88.452	114	101.497

Thus we assign this instance to cluster c_1 .

2.1.1.2 Instance 2

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
2	5.73	female	963	58.968	89

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_1	6.48	female	1034	62.143	127	80.596
c_2	6.59	male	1100	88.452	114	142.355

Thus we assign this instance to cluster c_1 .

2.1.1.3 Instance 3

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
3	7.99	female	1281	63.958	101

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_2	6.59	male	1100	88.452	114	183.120
c_1	6.48	female	1034	62.143	127	248.376

Thus we assign this instance to cluster c_2 .

2.1.1.4 Instance 4

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
4	8.42	female	1272	61.69	103

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_2	6.59	male	1100	88.452	114	174.429
c_1	6.48	female	1034	62.143	127	239.215

Thus we assign this instance to cluster c_2 .

2.1.1.5 Instance 5

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
5	6.84	female	1079	107.503	96

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_2	6.59	male	1100	88.452	114	33.601
c_1	6.48	female	1034	62.143	127	71.019

Thus we assign this instance to cluster c_2 .

2.1.1.6 Instance 6

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
6	6.43	female	1070	83.009	126

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_2	6.59	male	1100	88.452	114	32.782
c_1	6.48	female	1034	62.143	127	41.622

Thus we assign this instance to cluster c_2 .

2.1.1.7 Instance 7

We are assigning the following instances to a cluster:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ
7	7.6	male	1347	97.524	94

The clusters, sorted by distance to the instance, are:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_2	6.59	male	1100	88.452	114	247.976
c_1	6.48	female	1034	62.143	127	316.721

Thus we assign this instance to cluster c_2 .

2.1.1.8 Instance 8

We are assigning the following instances to a cluster:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ
8	6.03	male	1029	81.648	97

The clusters, sorted by distance to the instance, are:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_1	6.48	female	1034	62.143	127	36.148
c_2	6.59	male	1100	88.452	114	73.325

Thus we assign this instance to cluster c_1 .

2.1.1.9 Instance 9

We are assigning the following instances to a cluster:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ
9	7.52	male	1204	79.38	113

The clusters, sorted by distance to the instance, are:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_2	6.59	male	1100	88.452	114	104.404
c_1	6.48	female	1034	62.143	127	171.450

Thus we assign this instance to cluster c_2 .

2.1.1.10 Instance 10

We are assigning the following instances to a cluster:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ
10	7.67	male	1160	72.576	124

The clusters, sorted by distance to the instance, are:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_2	6.59	male	1100	88.452	114	62.875
c_1	6.48	female	1034	62.143	127	126.476

Thus we assign this instance to cluster c_2 .

2.1.2 Cluster Refinement

2.1.2.1 Cluster c_1

Let us refine the center of cluster c_1 which is currently centered at:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
c_1	6.48	female	1034	62.143	127

The members of this cluster are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
1	6.08	female	1005	57.607	96
2	5.73	female	963	58.968	89
8	6.03	male	1029	81.648	97

Using the average of the numeric attributes and majority of nominal attributes over all the member instances we derive the new center for this cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
c_1	5.947	female	999.000	66.074	94.000

2.1.2.2 Cluster c_2

Let us refine the center of cluster c_2 which is currently centered at:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
c_2	6.59	male	1100	88.452	114

The members of this cluster are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
3	7.99	female	1281	63.958	101
4	8.42	female	1272	61.69	103
5	6.84	female	1079	107.503	96
6	6.43	female	1070	83.009	126
7	7.6	male	1347	97.524	94
9	7.52	male	1204	79.38	113
10	7.67	male	1160	72.576	124

Using the average of the numeric attributes and majority of nominal attributes over all the member instances we derive the new center for this cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
c_2	7.496	female	1201.857	80.806	108.143

2.1.3 Cluster Formation

Let us assign each of our instances to a cluster.

2.1.3.1 Instance 1

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
1	6.08	female	1005	57.607	96

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_1	5.947	female	999.000	66.074	94.000	10.569
c_2	7.496	female	1201.857	80.806	108.143	198.596

Thus we assign this instance to cluster c_1 .

2.1.3.2 Instance 2

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
2	5.73	female	963	58.968	89

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_1	5.947	female	999.000	66.074	94.000	37.034
c_2	7.496	female	1201.857	80.806	108.143	240.622

Thus we assign this instance to cluster c_1 .

2.1.3.3 Instance 3

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
3	7.99	female	1281	63.958	101

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_2	7.496	female	1201.857	80.806	108.143	81.233
c_1	5.947	female	999.000	66.074	94.000	282.102

Thus we assign this instance to cluster c_2 .

2.1.3.4 Instance 4

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
4	8.42	female	1272	61.69	103

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_2	7.496	female	1201.857	80.806	108.143	72.889
c_1	5.947	female	999.000	66.074	94.000	273.195

Thus we assign this instance to cluster c_2 .

2.1.3.5 Instance 5

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
5	6.84	female	1079	107.503	96

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_1	5.947	female	999.000	66.074	94.000	90.117
c_2	7.496	female	1201.857	80.806	108.143	126.311

Thus we assign this instance to cluster c_1 .

2.1.3.6 Instance 6

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
6	6.43	female	1070	83.009	126

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_1	5.947	female	999.000	66.074	94.000	79.700
c_2	7.496	female	1201.857	80.806	108.143	133.083

Thus we assign this instance to cluster c_1 .

2.1.3.7 Instance 7

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
7	7.6	male	1347	97.524	94

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_2	7.496	female	1201.857	80.806	108.143	146.789
c_1	5.947	female	999.000	66.074	94.000	349.424

Thus we assign this instance to cluster c_2 .

2.1.3.8 Instance 8

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
8	6.03	male	1029	81.648	97

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_1	5.947	female	999.000	66.074	94.000	33.949
c_2	7.496	female	1201.857	80.806	108.143	173.227

Thus we assign this instance to cluster c_1 .

2.1.3.9 Instance 9

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
9	7.52	male	1204	79.38	113

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_2	7.496	female	1201.857	80.806	108.143	5.587
c_1	5.947	female	999.000	66.074	94.000	206.317

Thus we assign this instance to cluster c_2 .

2.1.3.10 Instance 10

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
10	7.67	male	1160	72.576	124

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_2	7.496	female	1201.857	80.806	108.143	45.522
c_1	5.947	female	999.000	66.074	94.000	163.912

Thus we assign this instance to cluster c_2 .

2.1.4 Cluster Refinement**2.1.4.1 Cluster c_1**

Let us refine the center of cluster c_1 which is currently centered at:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
c_1	5.947	female	999.000	66.074	94.000

The members of this cluster are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
1	6.08	female	1005	57.607	96
2	5.73	female	963	58.968	89
5	6.84	female	1079	107.503	96
6	6.43	female	1070	83.009	126
8	6.03	male	1029	81.648	97

Using the average of the numeric attributes and majority of nominal attributes over all the member instances we derive the new center for this cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
c_1	6.222	female	1029.200	77.747	100.800

2.1.4.2 Cluster c_2

Let us refine the center of cluster c_2 which is currently centered at:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
c_2	7.496	female	1201.857	80.806	108.143

The members of this cluster are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
3	7.99	female	1281	63.958	101
4	8.42	female	1272	61.69	103
7	7.6	male	1347	97.524	94
9	7.52	male	1204	79.38	113
10	7.67	male	1160	72.576	124

Using the average of the numeric attributes and majority of nominal attributes over all the member instances we derive the new center for this cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
c_2	7.840	male	1252.800	75.026	107.000

2.1.5 Final Clusters

2.1.5.1 Cluster c_1

The center of this cluster is:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
c_1	6.222	female	1029.200	77.747	100.800

The members of this cluster are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
1	6.08	female	1005	57.607	96
2	5.73	female	963	58.968	89
5	6.84	female	1079	107.503	96
6	6.43	female	1070	83.009	126
8	6.03	male	1029	81.648	97

2.1.5.2 Cluster c_2

The center of this cluster is:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
c_2	7.840	male	1252.800	75.026	107.000

The members of this cluster are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
3	7.99	female	1281	63.958	101
4	8.42	female	1272	61.69	103
7	7.6	male	1347	97.524	94
9	7.52	male	1204	79.38	113
10	7.67	male	1160	72.576	124

2.2 Normalized Data

2.2.1 Normalization

We begin with normalizing training instances and the initial centroids. We normalize both together. We find the minimum and maximum values for all of our non-target numeric attributes:

- **CCMIDSA**: minimum: 5.730, maximum: 8.420
- **GENDER**: This is not a numeric attribute so we leave it alone.
- **TOTVOL**: minimum: 963.000, maximum: 1347.000
- **WEIGHT**: minimum: 57.607, maximum: 107.503
- **FIQ**: minimum: 89.000, maximum: 127.000

Having the min and max values, we will apply the following transformation on each non-target numeric attribute:

$$new = \frac{old - min}{max - min}$$

Doing so produces the modified training set:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
1	0.130	female	0.109	0.000	0.184
2	0.000	female	0.000	0.027	0.000
3	0.840	female	0.828	0.127	0.316
4	1.000	female	0.805	0.082	0.368
5	0.413	female	0.302	1.000	0.184
6	0.260	female	0.279	0.509	0.974
7	0.695	male	1.000	0.800	0.132
8	0.112	male	0.172	0.482	0.211
9	0.665	male	0.628	0.436	0.632
10	0.721	male	0.513	0.300	0.921

We use the same min/max values we calculated to transform the initial centroids. Thus our initial centroids are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
1	0.279	female	0.185	0.091	1.000
2	0.320	male	0.357	0.618	0.658

2.2.2 Cluster Formation

Let us assign each of our instances to a cluster.

2.2.2.1 Instance 1

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
1	0.130	female	0.109	0.000	0.184

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_1	0.279	female	0.185	0.091	1.000	0.838
c_2	0.320	male	0.357	0.618	0.658	1.305

Thus we assign this instance to cluster c_1 .

2.2.2.2 Instance 2

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
2	0.000	female	0.000	0.027	0.000

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_1	0.279	female	0.185	0.091	1.000	1.056
c_2	0.320	male	0.357	0.618	0.658	1.418

Thus we assign this instance to cluster c_1 .

2.2.2.3 Instance 3

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
3	0.840	female	0.828	0.127	0.316

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_1	0.279	female	0.185	0.091	1.000	1.094
c_2	0.320	male	0.357	0.618	0.658	1.360

Thus we assign this instance to cluster c_1 .

2.2.2.4 Instance 4

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
4	1.000	female	0.805	0.082	0.368

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_1	0.279	female	0.185	0.091	1.000	1.142
c_2	0.320	male	0.357	0.618	0.658	1.426

Thus we assign this instance to cluster c_1 .

2.2.2.5 Instance 5

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
5	0.413	female	0.302	1.000	0.184

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_2	0.320	male	0.357	0.618	0.658	1.176
c_1	0.279	female	0.185	0.091	1.000	1.234

Thus we assign this instance to cluster c_2 .

2.2.2.6 Instance 6

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
6	0.260	female	0.279	0.509	0.974

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_1	0.279	female	0.185	0.091	1.000	0.430
c_2	0.320	male	0.357	0.618	0.658	1.059

Thus we assign this instance to cluster c_1 .

2.2.2.7 Instance 7

We are assigning the following instances to a cluster:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ
7	0.695	male	1.000	0.800	0.132

The clusters, sorted by distance to the instance, are:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_2	0.320	male	0.357	0.618	0.658	0.929
c_1	0.279	female	0.185	0.091	1.000	1.759

Thus we assign this instance to cluster c_2 .

2.2.2.8 Instance 8

We are assigning the following instances to a cluster:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ
8	0.112	male	0.172	0.482	0.211

The clusters, sorted by distance to the instance, are:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_2	0.320	male	0.357	0.618	0.658	0.544
c_1	0.279	female	0.185	0.091	1.000	1.343

Thus we assign this instance to cluster c_2 .

2.2.2.9 Instance 9

We are assigning the following instances to a cluster:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ
9	0.665	male	0.628	0.436	0.632

The clusters, sorted by distance to the instance, are:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_2	0.320	male	0.357	0.618	0.658	0.476
c_1	0.279	female	0.185	0.091	1.000	1.265

Thus we assign this instance to cluster c_2 .

2.2.2.10 Instance 10

We are assigning the following instances to a cluster:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ
10	0.721	male	0.513	0.300	0.921

The clusters, sorted by distance to the instance, are:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_2	0.320	male	0.357	0.618	0.658	0.596
c_1	0.279	female	0.185	0.091	1.000	1.163

Thus we assign this instance to cluster c_2 .

2.2.3 Cluster Refinement

2.2.3.1 Cluster c_1

Let us refine the center of cluster c_1 which is currently centered at:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
c_1	0.279	female	0.185	0.091	1.000

The members of this cluster are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
1	0.130	female	0.109	0.000	0.184
2	0.000	female	0.000	0.027	0.000
3	0.840	female	0.828	0.127	0.316
4	1.000	female	0.805	0.082	0.368
6	0.260	female	0.279	0.509	0.974

Using the average of the numeric attributes and majority of nominal attributes over all the member instances we derive the new center for this cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
c_1	0.446	female	0.404	0.149	0.368

2.2.3.2 Cluster c_2

Let us refine the center of cluster c_2 which is currently centered at:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
c_2	0.320	male	0.357	0.618	0.658

The members of this cluster are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
5	0.413	female	0.302	1.000	0.184
7	0.695	male	1.000	0.800	0.132
8	0.112	male	0.172	0.482	0.211
9	0.665	male	0.628	0.436	0.632
10	0.721	male	0.513	0.300	0.921

Using the average of the numeric attributes and majority of nominal attributes over all the member instances we derive the new center for this cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
c_2	0.521	male	0.523	0.604	0.416

2.2.4 Cluster Formation

Let us assign each of our instances to a cluster.

2.2.4.1 Instance 1

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
1	0.130	female	0.109	0.000	0.184

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_1	0.446	female	0.404	0.149	0.368	0.493
c_2	0.521	male	0.523	0.604	0.416	1.320

Thus we assign this instance to cluster c_1 .

2.2.4.2 Instance 2

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
2	0.000	female	0.000	0.027	0.000

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_1	0.446	female	0.404	0.149	0.368	0.716
c_2	0.521	male	0.523	0.604	0.416	1.432

Thus we assign this instance to cluster c_1 .

2.2.4.3 Instance 3

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
3	0.840	female	0.828	0.127	0.316

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_1	0.446	female	0.404	0.149	0.368	0.582
c_2	0.521	male	0.523	0.604	0.416	1.197

Thus we assign this instance to cluster c_1 .

2.2.4.4 Instance 4

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
4	1.000	female	0.805	0.082	0.368

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_1	0.446	female	0.404	0.149	0.368	0.687
c_2	0.521	male	0.523	0.604	0.416	1.258

Thus we assign this instance to cluster c_1 .

2.2.4.5 Instance 5

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
5	0.413	female	0.302	1.000	0.184

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_1	0.446	female	0.404	0.149	0.368	0.877
c_2	0.521	male	0.523	0.604	0.416	1.127

Thus we assign this instance to cluster c_1 .

2.2.4.6 Instance 6

We are assigning the following instances to a cluster:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ
6	0.260	female	0.279	0.509	0.974

The clusters, sorted by distance to the instance, are:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_1	0.446	female	0.404	0.149	0.368	0.740
c_2	0.521	male	0.523	0.604	0.416	1.203

Thus we assign this instance to cluster c_1 .

2.2.4.7 Instance 7

We are assigning the following instances to a cluster:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ
7	0.695	male	1.000	0.800	0.132

The clusters, sorted by distance to the instance, are:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_2	0.521	male	0.523	0.604	0.416	0.614
c_1	0.446	female	0.404	0.149	0.368	1.377

Thus we assign this instance to cluster c_2 .

2.2.4.8 Instance 8

We are assigning the following instances to a cluster:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ
8	0.112	male	0.172	0.482	0.211

The clusters, sorted by distance to the instance, are:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_2	0.521	male	0.523	0.604	0.416	0.589
c_1	0.446	female	0.404	0.149	0.368	1.141

Thus we assign this instance to cluster c_2 .

2.2.4.9 Instance 9

We are assigning the following instances to a cluster:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ
9	0.665	male	0.628	0.436	0.632

The clusters, sorted by distance to the instance, are:

#	CCMDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_2	0.521	male	0.523	0.604	0.416	0.327
c_1	0.446	female	0.404	0.149	0.368	1.118

Thus we assign this instance to cluster c_2 .

2.2.4.10 Instance 10

We are assigning the following instances to a cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
10	0.721	male	0.513	0.300	0.921

The clusters, sorted by distance to the instance, are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ	distance
c_2	0.521	male	0.523	0.604	0.416	0.623
c_1	0.446	female	0.404	0.149	0.368	1.190

Thus we assign this instance to cluster c_2 .

2.2.5 Cluster Refinement**2.2.5.1 Cluster c_1**

Let us refine the center of cluster c_1 which is currently centered at:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
c_1	0.446	female	0.404	0.149	0.368

The members of this cluster are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
1	0.130	female	0.109	0.000	0.184
2	0.000	female	0.000	0.027	0.000
3	0.840	female	0.828	0.127	0.316
4	1.000	female	0.805	0.082	0.368
5	0.413	female	0.302	1.000	0.184
6	0.260	female	0.279	0.509	0.974

Using the average of the numeric attributes and majority of nominal attributes over all the member instances we derive the new center for this cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
c_1	0.440	female	0.387	0.291	0.338

2.2.5.2 Cluster c_2

Let us refine the center of cluster c_2 which is currently centered at:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
c_2	0.521	male	0.523	0.604	0.416

The members of this cluster are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
7	0.695	male	1.000	0.800	0.132
8	0.112	male	0.172	0.482	0.211
9	0.665	male	0.628	0.436	0.632
10	0.721	male	0.513	0.300	0.921

Using the average of the numeric attributes and majority of nominal attributes over all the member instances we derive the new center for this cluster:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
c_2	0.548	male	0.578	0.504	0.474

2.2.6 Final Clusters

2.2.6.1 Cluster c_1

The center of this cluster is:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
c_1	0.440	female	0.387	0.291	0.338

The members of this cluster are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
1	0.130	female	0.109	0.000	0.184
2	0.000	female	0.000	0.027	0.000
3	0.840	female	0.828	0.127	0.316
4	1.000	female	0.805	0.082	0.368
5	0.413	female	0.302	1.000	0.184
6	0.260	female	0.279	0.509	0.974

2.2.6.2 Cluster c_2

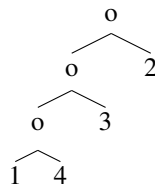
The center of this cluster is:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
c_2	0.548	male	0.578	0.504	0.474

The members of this cluster are:

#	CCMIDSA	GENDER	TOTVOL	WEIGHT	FIQ
7	0.695	male	1.000	0.800	0.132
8	0.112	male	0.172	0.482	0.211
9	0.665	male	0.628	0.436	0.632
10	0.721	male	0.513	0.300	0.921

3 Clustering - Hierarchical Clustering

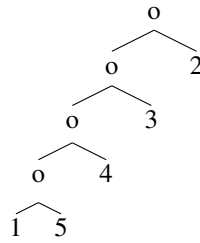


3.1 Number of Alternatives

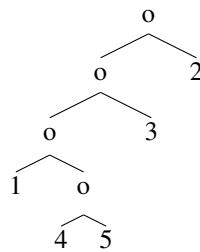
There are **nine** alternatives for the placement of the fifth instance.

3.2 Alternatives

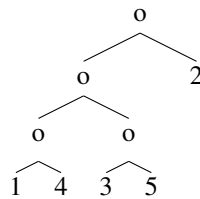
1. Place {5} in cluster with {1}:



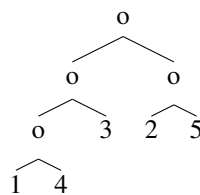
2. Place {5} in cluster with {4}:



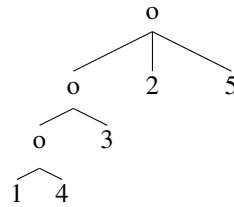
3. Place {5} in cluster with {3}:



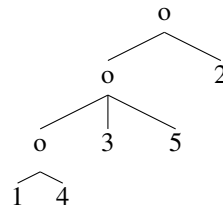
4. Place {5} in cluster with {2}:



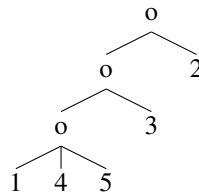
5. Place {5} in cluster with {2} and {1, 4, 3}:



6. Place {5} in cluster with {3} and {1, 4}:



7. Place {5} in cluster with {1} and {4}:



8. Merge the best two hosts for (5) (those two hosts will depend on the value produced by the category utility function, which we can't calculate here as no dataset is given.)
9. If merging is not the best option, splitting the best host should be considered.

3.3 Choice

See class text (pg. 260).