

# CS4445 B Term 2006

## Homework 4 Solutions

### Contents

<b>1 Instance Based Learning</b>	<b>4</b>
1.1 Original Data	4
1.1.1 Constant Weights	4
1.1.1.1 Instance $i_1$	4
1.1.1.2 Instance $i_2$	4
1.1.1.3 Instance $i_3$	5
1.1.2 Prediction Summary	5
1.1.3 Error Measures	6
1.1.3.1 Root-Mean Squared Error	6
1.1.3.2 Mean Absolute Error	6
1.1.4 Error Summary	6
1.1.5 Inverse Distance Weights	6
1.1.5.1 Instance $i_1$	6
1.1.5.2 Instance $i_2$	7
1.1.5.3 Instance $i_3$	7
1.1.6 Prediction Summary	8
1.1.7 Error Measures	8
1.1.7.1 Root-Mean Squared Error	8
1.1.7.2 Mean Absolute Error	8
1.1.8 Error Summary	9
1.2 Normalized Data	9
1.2.1 Normalization	9
1.2.2 Constant Weights	10
1.2.2.1 Instance $i_1$	10
1.2.2.2 Instance $i_2$	10
1.2.2.3 Instance $i_3$	11
1.2.3 Prediction Summary	11
1.2.4 Error Measures	11
1.2.4.1 Root-Mean Squared Error	11
1.2.4.2 Mean Absolute Error	12
1.2.5 Error Summary	12
1.2.6 Inverse Distance Weights	12
1.2.6.1 Instance $i_1$	12
1.2.6.2 Instance $i_2$	13
1.2.6.3 Instance $i_3$	13

1.2.7	Prediction Summary	14
1.2.8	Error Measures	14
1.2.8.1	Root-Mean Squared Error	14
1.2.8.2	Mean Absolute Error	14
1.2.9	Error Summary	14
1.3	Best Method	14
<b>2</b>	<b>Clustering - Simple K-means</b>	<b>14</b>
2.1	Original Data	15
2.1.1	Cluster Formation	15
2.1.1.1	Instance 1	15
2.1.1.2	Instance 2	15
2.1.1.3	Instance 3	16
2.1.1.4	Instance 4	16
2.1.1.5	Instance 5	16
2.1.1.6	Instance 6	16
2.1.1.7	Instance 7	17
2.1.1.8	Instance 8	17
2.1.1.9	Instance 9	17
2.1.1.10	Instance 10	17
2.1.1.11	Instance 11	18
2.1.1.12	Instance 12	18
2.1.2	Cluster Refinement	18
2.1.2.1	Cluster $c_1$	18
2.1.2.2	Cluster $c_2$	19
2.1.2.3	Cluster $c_3$	19
2.1.3	Cluster Formation	19
2.1.3.1	Instance 1	19
2.1.3.2	Instance 2	20
2.1.3.3	Instance 3	20
2.1.3.4	Instance 4	20
2.1.3.5	Instance 5	20
2.1.3.6	Instance 6	21
2.1.3.7	Instance 7	21
2.1.3.8	Instance 8	21
2.1.3.9	Instance 9	21
2.1.3.10	Instance 10	22
2.1.3.11	Instance 11	22
2.1.3.12	Instance 12	22
2.1.4	Cluster Refinement	22
2.1.4.1	Cluster $c_1$	22
2.1.4.2	Cluster $c_2$	23
2.1.4.3	Cluster $c_3$	23
2.1.5	Final Clusters	23
2.1.5.1	Cluster $c_1$	23
2.1.5.2	Cluster $c_2$	24
2.1.5.3	Cluster $c_3$	24

2.2	Normalized Data . . . . .	24
2.2.1	Normalization . . . . .	24
2.2.2	Cluster Formation . . . . .	25
2.2.2.1	Instance 1 . . . . .	25
2.2.2.2	Instance 2 . . . . .	25
2.2.2.3	Instance 3 . . . . .	26
2.2.2.4	Instance 4 . . . . .	26
2.2.2.5	Instance 5 . . . . .	26
2.2.2.6	Instance 6 . . . . .	26
2.2.2.7	Instance 7 . . . . .	27
2.2.2.8	Instance 8 . . . . .	27
2.2.2.9	Instance 9 . . . . .	27
2.2.2.10	Instance 10 . . . . .	27
2.2.2.11	Instance 11 . . . . .	28
2.2.2.12	Instance 12 . . . . .	28
2.2.3	Cluster Refinement . . . . .	28
2.2.3.1	Cluster $c_1$ . . . . .	28
2.2.3.2	Cluster $c_2$ . . . . .	29
2.2.3.3	Cluster $c_3$ . . . . .	29
2.2.4	Cluster Formation . . . . .	29
2.2.4.1	Instance 1 . . . . .	29
2.2.4.2	Instance 2 . . . . .	30
2.2.4.3	Instance 3 . . . . .	30
2.2.4.4	Instance 4 . . . . .	30
2.2.4.5	Instance 5 . . . . .	30
2.2.4.6	Instance 6 . . . . .	31
2.2.4.7	Instance 7 . . . . .	31
2.2.4.8	Instance 8 . . . . .	31
2.2.4.9	Instance 9 . . . . .	31
2.2.4.10	Instance 10 . . . . .	32
2.2.4.11	Instance 11 . . . . .	32
2.2.4.12	Instance 12 . . . . .	32
2.2.5	Cluster Refinement . . . . .	32
2.2.5.1	Cluster $c_1$ . . . . .	32
2.2.5.2	Cluster $c_2$ . . . . .	33
2.2.5.3	Cluster $c_3$ . . . . .	33
2.2.6	Final Clusters . . . . .	33
2.2.6.1	Cluster $c_1$ . . . . .	33
2.2.6.2	Cluster $c_2$ . . . . .	34
2.2.6.3	Cluster $c_3$ . . . . .	34
<b>3</b>	<b>Clustering - Hierarchical Clustering</b>	<b>34</b>
3.1	Number of Alternatives . . . . .	34
3.2	Alternatives . . . . .	34
3.3	Choice . . . . .	36

# 1 Instance Based Learning

**Note:** Throughout the following computation we will display the *country* attribute on instances, the attribute will be completely ignored.

## 1.1 Original Data

### 1.1.1 Constant Weights

We have the following instances to evaluate:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$i_1$	Costa-Rica	Americas	78.2	11.1	50.9	250
$i_2$	United-Kingdom	Europe	78.4	30.3	157.2	236.67
$i_3$	South-Africa	Africa	48.4	12	90.2	190

#### 1.1.1.1 Instance $i_1$

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$i_1$	Costa-Rica	Americas	78.2	11.1	50.9	250

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
10	India	Asia	63.3	3.3	49.9	180	16.877
5	Mexico	Americas	75.1	10	73.4	230	22.739
7	Thailand	Asia	70	8.3	79	216.67	29.423
12	Russia	Asia	65.3	11.1	81.9	143.3	33.592
4	Germany	Europe	78.7	30.4	99	240	51.840
8	Brazil	Americas	70.5	8.4	103.2	210	52.933
3	Usa	Americas	77.4	41.8	94.6	246.67	53.412
1	Switzerland	Europe	80.5	32.3	99.9	273.33	53.448
9	Japan	Asia	82	31.5	102.1	206.67	55.254
11	Ethiopia	Africa	47.6	0.9	5.2	156.67	55.945
2	Canada	Americas	80	34	102.6	253.33	56.573
6	France	Europe	79.5	29.9	108.7	220	60.803

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

$$\frac{180.000 * 1.000 + 230.000 * 1.000 + 216.670 * 1.000 + 143.300 * 1.000}{1.000 + 1.000 + 1.000 + 1.000} = \frac{769.970}{4.000} = 192.493$$

Thus our prediction for **SWL-index** for this test instance is 192.493.

#### 1.1.1.2 Instance $i_2$

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$i_2$	United-Kingdom	Europe	78.4	30.3	157.2	236.67

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
6	France	Europe	79.5	29.9	108.7	220	48.514
2	Canada	Americas	80	34	102.6	253.33	54.758
9	Japan	Asia	82	31.5	102.1	206.67	55.240
1	Switzerland	Europe	80.5	32.3	99.9	273.33	57.373
4	Germany	Europe	78.7	30.4	99	240	58.201
8	Brazil	Americas	70.5	8.4	103.2	210	58.813
3	Usa	Americas	77.4	41.8	94.6	246.67	63.663
12	Russia	Asia	65.3	11.1	81.9	143.3	78.812
7	Thailand	Asia	70	8.3	79	216.67	81.675
5	Mexico	Americas	75.1	10	73.4	230	86.293
10	India	Asia	63.3	3.3	49.9	180	111.675
11	Ethiopia	Africa	47.6	0.9	5.2	156.67	157.854

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

$$\frac{220.000 * 1.000 + 253.330 * 1.000 + 206.670 * 1.000 + 273.330 * 1.000}{1.000 + 1.000 + 1.000 + 1.000} = \frac{953.330}{4.000} = 238.332$$

Thus our prediction for **SWL-index** for this test instance is 238.332.

### 1.1.1.3 Instance $i_3$

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$i_3$	South-Africa	Africa	48.4	12	90.2	190

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
12	Russia	Asia	65.3	11.1	81.9	143.3	18.876
7	Thailand	Asia	70	8.3	79	216.67	24.631
8	Brazil	Americas	70.5	8.4	103.2	210	25.911
5	Mexico	Americas	75.1	10	73.4	230	31.625
4	Germany	Europe	78.7	30.4	99	240	36.539
1	Switzerland	Europe	80.5	32.3	99.9	273.33	39.212
6	France	Europe	79.5	29.9	108.7	220	40.384
2	Canada	Americas	80	34	102.6	253.33	40.464
9	Japan	Asia	82	31.5	102.1	206.67	40.643
3	Usa	Americas	77.4	41.8	94.6	246.67	41.826
10	India	Asia	63.3	3.3	49.9	180	43.850
11	Ethiopia	Africa	47.6	0.9	5.2	156.67	85.725

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

$$\frac{143.300 * 1.000 + 216.670 * 1.000 + 210.000 * 1.000 + 230.000 * 1.000}{1.000 + 1.000 + 1.000 + 1.000} = \frac{799.970}{4.000} = 199.993$$

Thus our prediction for **SWL-index** for this test instance is 199.993.

## 1.1.2 Prediction Summary

Our method made the following predictions:

#	SWL-index	Constant Weights
$i_1$	250	192.493
$i_2$	236.67	238.332
$i_3$	190	199.993

### 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:

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

$$\left( \frac{(192.493 - 250.000)^2 + (238.332 - 236.670)^2 + (199.993 - 190.000)^2}{3} \right)^{\frac{1}{2}} = 33.713$$

#### 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

$$\frac{|192.493 - 250.000| + |238.332 - 236.670| + |199.993 - 190.000|}{3} = 23.054$$

### 1.1.4 Error Summary

And we have our accuracy measures:

measure	Constant Weights
root mean-squared error	33.713
mean absolute error	23.054

### 1.1.5 Inverse Distance Weights

We have the following instances to evaluate:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$i_1$	Costa-Rica	Americas	78.2	11.1	50.9	250
$i_2$	United-Kingdom	Europe	78.4	30.3	157.2	236.67
$i_3$	South-Africa	Africa	48.4	12	90.2	190

#### 1.1.5.1 Instance $i_1$

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$i_1$	Costa-Rica	Americas	78.2	11.1	50.9	250

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
10	India	Asia	63.3	3.3	49.9	180	16.877
5	Mexico	Americas	75.1	10	73.4	230	22.739
7	Thailand	Asia	70	8.3	79	216.67	29.423
12	Russia	Asia	65.3	11.1	81.9	143.3	33.592
4	Germany	Europe	78.7	30.4	99	240	51.840
8	Brazil	Americas	70.5	8.4	103.2	210	52.933
3	Usa	Americas	77.4	41.8	94.6	246.67	53.412
1	Switzerland	Europe	80.5	32.3	99.9	273.33	53.448
9	Japan	Asia	82	31.5	102.1	206.67	55.254
11	Ethiopia	Africa	47.6	0.9	5.2	156.67	55.945
2	Canada	Americas	80	34	102.6	253.33	56.573
6	France	Europe	79.5	29.9	108.7	220	60.803

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

$$\frac{180.000 * \frac{1}{16.877} + 230.000 * \frac{1}{22.739} + 216.670 * \frac{1}{29.423} + 143.300 * \frac{1}{33.592}}{\frac{1}{16.877} + \frac{1}{22.739} + \frac{1}{29.423} + \frac{1}{33.592}} = \frac{32.410}{0.167} = 194.089$$

Thus our prediction for **SWL-index** for this test instance is 194.089.

### 1.1.5.2 Instance $i_2$

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$i_2$	United-Kingdom	Europe	78.4	30.3	157.2	236.67

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
6	France	Europe	79.5	29.9	108.7	220	48.514
2	Canada	Americas	80	34	102.6	253.33	54.758
9	Japan	Asia	82	31.5	102.1	206.67	55.240
1	Switzerland	Europe	80.5	32.3	99.9	273.33	57.373
4	Germany	Europe	78.7	30.4	99	240	58.201
8	Brazil	Americas	70.5	8.4	103.2	210	58.813
3	Usa	Americas	77.4	41.8	94.6	246.67	63.663
12	Russia	Asia	65.3	11.1	81.9	143.3	78.812
7	Thailand	Asia	70	8.3	79	216.67	81.675
5	Mexico	Americas	75.1	10	73.4	230	86.293
10	India	Asia	63.3	3.3	49.9	180	111.675
11	Ethiopia	Africa	47.6	0.9	5.2	156.67	157.854

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

$$\frac{220.000 * \frac{1}{48.514} + 253.330 * \frac{1}{54.758} + 206.670 * \frac{1}{55.240} + 273.330 * \frac{1}{57.373}}{\frac{1}{48.514} + \frac{1}{54.758} + \frac{1}{55.240} + \frac{1}{57.373}} = \frac{17.667}{0.074} = 237.430$$

Thus our prediction for **SWL-index** for this test instance is 237.430.

### 1.1.5.3 Instance $i_3$

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$i_3$	South-Africa	Africa	48.4	12	90.2	190

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
12	Russia	Asia	65.3	11.1	81.9	143.3	18.876
7	Thailand	Asia	70	8.3	79	216.67	24.631
8	Brazil	Americas	70.5	8.4	103.2	210	25.911
5	Mexico	Americas	75.1	10	73.4	230	31.625
4	Germany	Europe	78.7	30.4	99	240	36.539
1	Switzerland	Europe	80.5	32.3	99.9	273.33	39.212
6	France	Europe	79.5	29.9	108.7	220	40.384
2	Canada	Americas	80	34	102.6	253.33	40.464
9	Japan	Asia	82	31.5	102.1	206.67	40.643
3	Usa	Americas	77.4	41.8	94.6	246.67	41.826
10	India	Asia	63.3	3.3	49.9	180	43.850
11	Ethiopia	Africa	47.6	0.9	5.2	156.67	85.725

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

$$\frac{143.300 * \frac{1}{18.876} + 216.670 * \frac{1}{24.631} + 210.000 * \frac{1}{25.911} + 230.000 * \frac{1}{31.625}}{\frac{1}{18.876} + \frac{1}{24.631} + \frac{1}{25.911} + \frac{1}{31.625}} = \frac{31.766}{0.164} = 193.941$$

Thus our prediction for **SWL-index** for this test instance is 193.941.

### 1.1.6 Prediction Summary

Our method made the following predictions:

#	SWL-index	Inverse Distance Weights
$i_1$	250	194.089
$i_2$	236.67	237.430
$i_3$	190	193.941

### 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:

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

$$\left( \frac{(194.089 - 250.000)^2 + (237.430 - 236.670)^2 + (193.941 - 190.000)^2}{3} \right)^{\frac{1}{2}} = 32.363$$

#### 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

$$\frac{|194.089 - 250.000| + |237.430 - 236.670| + |193.941 - 190.000|}{3} = 20.204$$



### 1.1.8 Error Summary

And we have our accuracy measures:

measure	Inverse Distance Weights
root mean-squared error	32.363
mean absolute error	20.204

## 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:

- **country:** We are ignoring this attribute.
- **continent:** This is not a numeric attribute so we leave it alone.
- **life-expectancy:** minimum: 47.600, maximum: 82.000
- **GDP-per-capita:** minimum: 0.900, maximum: 41.800
- **access-to-education-score:** minimum: 5.200, maximum: 157.200
- **SWL-index:** 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:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
1	Switzerland	Europe	0.956	0.768	0.623	273.33
2	Canada	Americas	0.942	0.809	0.641	253.33
3	Usa	Americas	0.866	1.000	0.588	246.67
4	Germany	Europe	0.904	0.721	0.617	240
5	Mexico	Americas	0.799	0.222	0.449	230
6	France	Europe	0.927	0.709	0.681	220
7	Thailand	Asia	0.651	0.181	0.486	216.67
8	Brazil	Americas	0.666	0.183	0.645	210
9	Japan	Asia	1.000	0.748	0.637	206.67
10	India	Asia	0.456	0.059	0.294	180
11	Ethiopia	Africa	0.000	0.000	0.000	156.67
12	Russia	Asia	0.515	0.249	0.505	143.3

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$i_1$	Costa-Rica	Americas	0.890	0.249	0.301	250
$i_2$	United-Kingdom	Europe	0.895	0.719	1.000	236.67
$i_3$	South-Africa	Africa	0.023	0.271	0.559	190

We now repeat the testing process on these modified instances.

## 1.2.2 Constant Weights

We have the following instances to evaluate:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$i_1$	Costa-Rica	Americas	0.890	0.249	0.301	250
$i_2$	United-Kingdom	Europe	0.895	0.719	1.000	236.67
$i_3$	South-Africa	Africa	0.023	0.271	0.559	190

### 1.2.2.1 Instance $i_1$

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$i_1$	Costa-Rica	Americas	0.890	0.249	0.301	250

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
5	Mexico	Americas	0.799	0.222	0.449	230	0.176
8	Brazil	Americas	0.666	0.183	0.645	210	0.416
2	Canada	Americas	0.942	0.809	0.641	253.33	0.657
3	Usa	Americas	0.866	1.000	0.588	246.67	0.804
7	Thailand	Asia	0.651	0.181	0.486	216.67	1.047
12	Russia	Asia	0.515	0.249	0.505	143.3	1.087
10	India	Asia	0.456	0.059	0.294	180	1.107
4	Germany	Europe	0.904	0.721	0.617	240	1.150
6	France	Europe	0.927	0.709	0.681	220	1.165
9	Japan	Asia	1.000	0.748	0.637	206.67	1.172
1	Switzerland	Europe	0.956	0.768	0.623	273.33	1.174
11	Ethiopia	Africa	0.000	0.000	0.000	156.67	1.395

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

$$\frac{230.000 * 1.000 + 210.000 * 1.000 + 253.330 * 1.000 + 246.670 * 1.000}{1.000 + 1.000 + 1.000 + 1.000} = \frac{940.000}{4.000} = 235.000$$

Thus our prediction for **SWL-index** for this test instance is 235.000.

### 1.2.2.2 Instance $i_2$

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$i_2$	United-Kingdom	Europe	0.895	0.719	1.000	236.67

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
6	France	Europe	0.927	0.709	0.681	220	0.321
4	Germany	Europe	0.904	0.721	0.617	240	0.383
1	Switzerland	Europe	0.956	0.768	0.623	273.33	0.385
2	Canada	Americas	0.942	0.809	0.641	253.33	1.067
9	Japan	Asia	1.000	0.748	0.637	206.67	1.069
3	Usa	Americas	0.866	1.000	0.588	246.67	1.118
8	Brazil	Americas	0.666	0.183	0.645	210	1.211
5	Mexico	Americas	0.799	0.222	0.449	230	1.249
12	Russia	Asia	0.515	0.249	0.505	143.3	1.269
7	Thailand	Asia	0.651	0.181	0.486	216.67	1.270
10	India	Asia	0.456	0.059	0.294	180	1.458
11	Ethiopia	Africa	0.000	0.000	0.000	156.67	1.822

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

$$\frac{220.000 * 1.000 + 240.000 * 1.000 + 273.330 * 1.000 + 253.330 * 1.000}{1.000 + 1.000 + 1.000 + 1.000} = \frac{986.660}{4.000} = 246.665$$

Thus our prediction for **SWL-index** for this test instance is 246.665.

### 1.2.2.3 Instance $i_3$

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$i_3$	South-Africa	Africa	0.023	0.271	0.559	190

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
11	Ethiopia	Africa	0.000	0.000	0.000	156.67	0.622
12	Russia	Asia	0.515	0.249	0.505	143.3	1.116
10	India	Asia	0.456	0.059	0.294	180	1.141
7	Thailand	Asia	0.651	0.181	0.486	216.67	1.187
8	Brazil	Americas	0.666	0.183	0.645	210	1.195
5	Mexico	Americas	0.799	0.222	0.449	230	1.271
4	Germany	Europe	0.904	0.721	0.617	240	1.408
6	France	Europe	0.927	0.709	0.681	220	1.423
1	Switzerland	Europe	0.956	0.768	0.623	273.33	1.457
2	Canada	Americas	0.942	0.809	0.641	253.33	1.463
9	Japan	Asia	1.000	0.748	0.637	206.67	1.479
3	Usa	Americas	0.866	1.000	0.588	246.67	1.498

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

$$\frac{156.670 * 1.000 + 143.300 * 1.000 + 180.000 * 1.000 + 216.670 * 1.000}{1.000 + 1.000 + 1.000 + 1.000} = \frac{696.640}{4.000} = 174.160$$

Thus our prediction for **SWL-index** for this test instance is 174.160.

### 1.2.3 Prediction Summary

Our method made the following predictions:

#	SWL-index	Constant Weights
$i_1$	250	235.000
$i_2$	236.67	246.665
$i_3$	190	174.160

### 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:

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

$$\left( \frac{(235.000 - 250.000)^2 + (246.665 - 236.670)^2 + (174.160 - 190.000)^2}{3} \right)^{\frac{1}{2}} = 13.854$$

### 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

$$\frac{|235.000 - 250.000| + |246.665 - 236.670| + |174.160 - 190.000|}{3} = 13.612$$

### 1.2.5 Error Summary

And we have our accuracy measures:

measure	Constant Weights
root mean-squared error	13.854
mean absolute error	13.612

### 1.2.6 Inverse Distance Weights

We have the following instances to evaluate:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$i_1$	Costa-Rica	Americas	0.890	0.249	0.301	250
$i_2$	United-Kingdom	Europe	0.895	0.719	1.000	236.67
$i_3$	South-Africa	Africa	0.023	0.271	0.559	190

#### 1.2.6.1 Instance $i_1$

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$i_1$	Costa-Rica	Americas	0.890	0.249	0.301	250

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
5	Mexico	Americas	0.799	0.222	0.449	230	0.176
8	Brazil	Americas	0.666	0.183	0.645	210	0.416
2	Canada	Americas	0.942	0.809	0.641	253.33	0.657
3	Usa	Americas	0.866	1.000	0.588	246.67	0.804
7	Thailand	Asia	0.651	0.181	0.486	216.67	1.047
12	Russia	Asia	0.515	0.249	0.505	143.3	1.087
10	India	Asia	0.456	0.059	0.294	180	1.107
4	Germany	Europe	0.904	0.721	0.617	240	1.150
6	France	Europe	0.927	0.709	0.681	220	1.165
9	Japan	Asia	1.000	0.748	0.637	206.67	1.172
1	Switzerland	Europe	0.956	0.768	0.623	273.33	1.174
11	Ethiopia	Africa	0.000	0.000	0.000	156.67	1.395

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

$$\frac{230.000 * \frac{1}{0.176} + 210.000 * \frac{1}{0.416} + 253.330 * \frac{1}{0.657} + 246.670 * \frac{1}{0.804}}{\frac{1}{0.176} + \frac{1}{0.416} + \frac{1}{0.657} + \frac{1}{0.804}} = \frac{2505.264}{10.857} = 230.748$$

Thus our prediction for **SWL-index** for this test instance is 230.748.

1.2.6.2 Instance  $i_2$ 

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$i_2$	United-Kingdom	Europe	0.895	0.719	1.000	236.67

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
6	France	Europe	0.927	0.709	0.681	220	0.321
4	Germany	Europe	0.904	0.721	0.617	240	0.383
1	Switzerland	Europe	0.956	0.768	0.623	273.33	0.385
2	Canada	Americas	0.942	0.809	0.641	253.33	1.067
9	Japan	Asia	1.000	0.748	0.637	206.67	1.069
3	Usa	Americas	0.866	1.000	0.588	246.67	1.118
8	Brazil	Americas	0.666	0.183	0.645	210	1.211
5	Mexico	Americas	0.799	0.222	0.449	230	1.249
12	Russia	Asia	0.515	0.249	0.505	143.3	1.269
7	Thailand	Asia	0.651	0.181	0.486	216.67	1.270
10	India	Asia	0.456	0.059	0.294	180	1.458
11	Ethiopia	Africa	0.000	0.000	0.000	156.67	1.822

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

$$\frac{220.000 * \frac{1}{0.321} + 240.000 * \frac{1}{0.383} + 273.330 * \frac{1}{0.385} + 253.330 * \frac{1}{1.067}}{\frac{1}{0.321} + \frac{1}{0.383} + \frac{1}{0.385} + \frac{1}{1.067}} = \frac{2259.505}{9.262} = 243.963$$

Thus our prediction for **SWL-index** for this test instance is 243.963.

1.2.6.3 Instance  $i_3$ 

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$i_3$	South-Africa	Africa	0.023	0.271	0.559	190

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
11	Ethiopia	Africa	0.000	0.000	0.000	156.67	0.622
12	Russia	Asia	0.515	0.249	0.505	143.3	1.116
10	India	Asia	0.456	0.059	0.294	180	1.141
7	Thailand	Asia	0.651	0.181	0.486	216.67	1.187
8	Brazil	Americas	0.666	0.183	0.645	210	1.195
5	Mexico	Americas	0.799	0.222	0.449	230	1.271
4	Germany	Europe	0.904	0.721	0.617	240	1.408
6	France	Europe	0.927	0.709	0.681	220	1.423
1	Switzerland	Europe	0.956	0.768	0.623	273.33	1.457
2	Canada	Americas	0.942	0.809	0.641	253.33	1.463
9	Japan	Asia	1.000	0.748	0.637	206.67	1.479
3	Usa	Americas	0.866	1.000	0.588	246.67	1.498

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

$$\frac{156.670 * \frac{1}{0.622} + 143.300 * \frac{1}{1.116} + 180.000 * \frac{1}{1.141} + 216.670 * \frac{1}{1.187}}{\frac{1}{0.622} + \frac{1}{1.116} + \frac{1}{1.141} + \frac{1}{1.187}} = \frac{720.739}{4.224} = 170.646$$

Thus our prediction for **SWL-index** for this test instance is 170.646.

### 1.2.7 Prediction Summary

Our method made the following predictions:

#	SWL-index	Inverse Distance Weights
$i_1$	250	230.748
$i_2$	236.67	243.963
$i_3$	190	170.646

### 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:

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

$$\left( \frac{(230.748 - 250.000)^2 + (243.963 - 236.670)^2 + (170.646 - 190.000)^2}{3} \right)^{\frac{1}{2}} = 16.314$$

#### 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

$$\frac{|230.748 - 250.000| + |243.963 - 236.670| + |170.646 - 190.000|}{3} = 15.300$$

### 1.2.9 Error Summary

And we have our accuracy measures:

measure	Inverse Distance Weights
root mean-squared error	16.314
mean absolute error	15.300

## 1.3 Best Method

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

## 2 Clustering - Simple K-means

**Note:** Throughout the following computation we will display the **country** attribute on instances and clusters, the attribute will be completely ignored.

## 2.1 Original Data

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
1	Switzerland	Europe	80.5	32.3	99.9	273.33
2	Canada	Americas	80	34	102.6	253.33
3	Usa	Americas	77.4	41.8	94.6	246.67
4	Germany	Europe	78.7	30.4	99	240
5	Mexico	Americas	75.1	10	73.4	230
6	France	Europe	79.5	29.9	108.7	220
7	Thailand	Asia	70	8.3	79	216.67
8	Brazil	Americas	70.5	8.4	103.2	210
9	Japan	Asia	82	31.5	102.1	206.67
10	India	Asia	63.3	3.3	49.9	180
11	Ethiopia	Africa	47.6	0.9	5.2	156.67
12	Russia	Asia	65.3	11.1	81.9	143.3

### 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:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
1	Switzerland	Europe	80.5	32.3	99.9	273.33

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_1$	Costa-Rica	Americas	78.2	11.1	50.9	250	58.318
$c_2$	United-Kingdom	Europe	78.4	30.3	157.2	236.67	68.086
$c_3$	South-Africa	Africa	48.4	12	90.2	190	92.095

Thus we assign this instance to cluster  $c_1$ .

#### 2.1.1.2 Instance 2

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
2	Canada	Americas	80	34	102.6	253.33

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_1$	Costa-Rica	Americas	78.2	11.1	50.9	250	56.671
$c_2$	United-Kingdom	Europe	78.4	30.3	157.2	236.67	57.236
$c_3$	South-Africa	Africa	48.4	12	90.2	190	75.153

Thus we assign this instance to cluster  $c_1$ .

**2.1.1.3 Instance 3**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
3	Usa	Americas	77.4	41.8	94.6	246.67

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_1$	Costa-Rica	Americas	78.2	11.1	50.9	250	53.516
$c_2$	United-Kingdom	Europe	78.4	30.3	157.2	236.67	64.444
$c_3$	South-Africa	Africa	48.4	12	90.2	190	70.434

Thus we assign this instance to cluster  $c_1$ .

**2.1.1.4 Instance 4**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
4	Germany	Europe	78.7	30.4	99	240

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_1$	Costa-Rica	Americas	78.2	11.1	50.9	250	52.795
$c_2$	United-Kingdom	Europe	78.4	30.3	157.2	236.67	58.296
$c_3$	South-Africa	Africa	48.4	12	90.2	190	61.928

Thus we assign this instance to cluster  $c_1$ .

**2.1.1.5 Instance 5**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
5	Mexico	Americas	75.1	10	73.4	230

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_1$	Costa-Rica	Americas	78.2	11.1	50.9	250	30.283
$c_3$	South-Africa	Africa	48.4	12	90.2	190	50.991
$c_2$	United-Kingdom	Europe	78.4	30.3	157.2	236.67	86.550

Thus we assign this instance to cluster  $c_1$ .

**2.1.1.6 Instance 6**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
6	France	Europe	79.5	29.9	108.7	220

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_3$	South-Africa	Africa	48.4	12	90.2	190	50.308
$c_2$	United-Kingdom	Europe	78.4	30.3	157.2	236.67	51.298
$c_1$	Costa-Rica	Americas	78.2	11.1	50.9	250	67.801

Thus we assign this instance to cluster  $c_3$ .



**2.1.1.7 Instance 7**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
7	Thailand	Asia	70	8.3	79	216.67

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_3$	South-Africa	Africa	48.4	12	90.2	190	36.304
$c_1$	Costa-Rica	Americas	78.2	11.1	50.9	250	44.459
$c_2$	United-Kingdom	Europe	78.4	30.3	157.2	236.67	84.088

Thus we assign this instance to cluster  $c_3$ .

**2.1.1.8 Instance 8**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
8	Brazil	Americas	70.5	8.4	103.2	210

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_3$	South-Africa	Africa	48.4	12	90.2	190	32.732
$c_2$	United-Kingdom	Europe	78.4	30.3	157.2	236.67	64.578
$c_1$	Costa-Rica	Americas	78.2	11.1	50.9	250	66.347

Thus we assign this instance to cluster  $c_3$ .

**2.1.1.9 Instance 9**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
9	Japan	Asia	82	31.5	102.1	206.67

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_3$	South-Africa	Africa	48.4	12	90.2	190	43.928
$c_2$	United-Kingdom	Europe	78.4	30.3	157.2	236.67	62.860
$c_1$	Costa-Rica	Americas	78.2	11.1	50.9	250	70.218

Thus we assign this instance to cluster  $c_3$ .

**2.1.1.10 Instance 10**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
10	India	Asia	63.3	3.3	49.9	180

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_3$	South-Africa	Africa	48.4	12	90.2	190	44.975
$c_1$	Costa-Rica	Americas	78.2	11.1	50.9	250	72.006
$c_2$	United-Kingdom	Europe	78.4	30.3	157.2	236.67	125.231

Thus we assign this instance to cluster  $c_3$ .

**2.1.1.11 Instance 11**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
11	Ethiopia	Africa	47.6	0.9	5.2	156.67

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_3$	South-Africa	Africa	48.4	12	90.2	190	91.977
$c_1$	Costa-Rica	Americas	78.2	11.1	50.9	250	108.814
$c_2$	United-Kingdom	Europe	78.4	30.3	157.2	236.67	176.969

Thus we assign this instance to cluster  $c_3$ .

**2.1.1.12 Instance 12**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
12	Russia	Asia	65.3	11.1	81.9	143.3

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_3$	South-Africa	Africa	48.4	12	90.2	190	50.371
$c_1$	Costa-Rica	Americas	78.2	11.1	50.9	250	111.863
$c_2$	United-Kingdom	Europe	78.4	30.3	157.2	236.67	122.186

Thus we assign this instance to cluster  $c_3$ .

**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:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$c_1$	Costa-Rica	Americas	78.2	11.1	50.9	250

The members of this cluster are:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
1	Switzerland	Europe	80.5	32.3	99.9	273.33
2	Canada	Americas	80	34	102.6	253.33
3	Usa	Americas	77.4	41.8	94.6	246.67
4	Germany	Europe	78.7	30.4	99	240
5	Mexico	Americas	75.1	10	73.4	230

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:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$c_1$	Costa-Rica	Americas	78.340	29.700	93.900	248.666

**2.1.2.2 Cluster  $c_2$** 

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$c_2$	United-Kingdom	Europe	78.4	30.3	157.2	236.67

The members of this cluster are:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
---	---------	-----------	-----------------	----------------	---------------------------	-----------

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:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$c_2$	United-Kingdom	Europe	78.4	30.3	157.2	236.67

**2.1.2.3 Cluster  $c_3$** 

Let us refine the center of cluster  $c_3$  which is currently centered at:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$c_3$	South-Africa	Africa	48.4	12	90.2	190

The members of this cluster are:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
6	France	Europe	79.5	29.9	108.7	220
7	Thailand	Asia	70	8.3	79	216.67
8	Brazil	Americas	70.5	8.4	103.2	210
9	Japan	Asia	82	31.5	102.1	206.67
10	India	Asia	63.3	3.3	49.9	180
11	Ethiopia	Africa	47.6	0.9	5.2	156.67
12	Russia	Asia	65.3	11.1	81.9	143.3

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:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$c_3$	South-Africa	Asia	68.314	13.343	75.714	190.473

**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:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
1	Switzerland	Europe	80.5	32.3	99.9	273.33

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_1$	Costa-Rica	Americas	78.340	29.700	93.900	248.666	25.627
$c_2$	United-Kingdom	Europe	78.4	30.3	157.2	236.67	68.086
$c_3$	South-Africa	Asia	68.314	13.343	75.714	190.473	89.214

Thus we assign this instance to cluster  $c_1$ .

**2.1.3.2 Instance 2**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
2	Canada	Americas	80	34	102.6	253.33

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_1$	Costa-Rica	Americas	78.340	29.700	93.900	248.666	10.894
$c_2$	United-Kingdom	Europe	78.4	30.3	157.2	236.67	57.236
$c_3$	South-Africa	Asia	68.314	13.343	75.714	190.473	72.375

Thus we assign this instance to cluster  $c_1$ .

**2.1.3.3 Instance 3**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
3	Usa	Americas	77.4	41.8	94.6	246.67

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_1$	Costa-Rica	Americas	78.340	29.700	93.900	248.666	12.319
$c_2$	United-Kingdom	Europe	78.4	30.3	157.2	236.67	64.444
$c_3$	South-Africa	Asia	68.314	13.343	75.714	190.473	66.394

Thus we assign this instance to cluster  $c_1$ .

**2.1.3.4 Instance 4**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
4	Germany	Europe	78.7	30.4	99	240

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_1$	Costa-Rica	Americas	78.340	29.700	93.900	248.666	10.136
$c_3$	South-Africa	Asia	68.314	13.343	75.714	190.473	58.266
$c_2$	United-Kingdom	Europe	78.4	30.3	157.2	236.67	58.296

Thus we assign this instance to cluster  $c_1$ .

**2.1.3.5 Instance 5**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
5	Mexico	Americas	75.1	10	73.4	230

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_1$	Costa-Rica	Americas	78.340	29.700	93.900	248.666	34.165
$c_3$	South-Africa	Asia	68.314	13.343	75.714	190.473	40.323
$c_2$	United-Kingdom	Europe	78.4	30.3	157.2	236.67	86.550

Thus we assign this instance to cluster  $c_1$ .

**2.1.3.6 Instance 6**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
6	France	Europe	79.5	29.9	108.7	220

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
c <sub>1</sub>	Costa-Rica	Americas	78.340	29.700	93.900	248.666	32.298
c <sub>3</sub>	South-Africa	Asia	68.314	13.343	75.714	190.473	48.582
c <sub>2</sub>	United-Kingdom	Europe	78.4	30.3	157.2	236.67	51.298

Thus we assign this instance to cluster  $c_1$ .

**2.1.3.7 Instance 7**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
7	Thailand	Asia	70	8.3	79	216.67

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
c <sub>3</sub>	South-Africa	Asia	68.314	13.343	75.714	190.473	26.932
c <sub>1</sub>	Costa-Rica	Americas	78.340	29.700	93.900	248.666	42.122
c <sub>2</sub>	United-Kingdom	Europe	78.4	30.3	157.2	236.67	84.088

Thus we assign this instance to cluster  $c_3$ .

**2.1.3.8 Instance 8**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
8	Brazil	Americas	70.5	8.4	103.2	210

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
c <sub>3</sub>	South-Africa	Asia	68.314	13.343	75.714	190.473	34.161
c <sub>1</sub>	Costa-Rica	Americas	78.340	29.700	93.900	248.666	45.790
c <sub>2</sub>	United-Kingdom	Europe	78.4	30.3	157.2	236.67	64.578

Thus we assign this instance to cluster  $c_3$ .

**2.1.3.9 Instance 9**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
9	Japan	Asia	82	31.5	102.1	206.67

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
c <sub>3</sub>	South-Africa	Asia	68.314	13.343	75.714	190.473	38.413
c <sub>1</sub>	Costa-Rica	Americas	78.340	29.700	93.900	248.666	42.995
c <sub>2</sub>	United-Kingdom	Europe	78.4	30.3	157.2	236.67	62.860

Thus we assign this instance to cluster  $c_3$ .

**2.1.3.10 Instance 10**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
10	India	Asia	63.3	3.3	49.9	180

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_3$	South-Africa	Asia	68.314	13.343	75.714	190.473	30.034
$c_1$	Costa-Rica	Americas	78.340	29.700	93.900	248.666	87.036
$c_2$	United-Kingdom	Europe	78.4	30.3	157.2	236.67	125.231

Thus we assign this instance to cluster  $c_3$ .

**2.1.3.11 Instance 11**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
11	Ethiopia	Africa	47.6	0.9	5.2	156.67

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_3$	South-Africa	Asia	68.314	13.343	75.714	190.473	81.852
$c_1$	Costa-Rica	Americas	78.340	29.700	93.900	248.666	134.560
$c_2$	United-Kingdom	Europe	78.4	30.3	157.2	236.67	176.969

Thus we assign this instance to cluster  $c_3$ .

**2.1.3.12 Instance 12**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
12	Russia	Asia	65.3	11.1	81.9	143.3

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_3$	South-Africa	Asia	68.314	13.343	75.714	190.473	47.725
$c_1$	Costa-Rica	Americas	78.340	29.700	93.900	248.666	108.457
$c_2$	United-Kingdom	Europe	78.4	30.3	157.2	236.67	122.186

Thus we assign this instance to cluster  $c_3$ .

**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:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$c_1$	Costa-Rica	Americas	78.340	29.700	93.900	248.666

The members of this cluster are:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
1	Switzerland	Europe	80.5	32.3	99.9	273.33
2	Canada	Americas	80	34	102.6	253.33
3	Usa	Americas	77.4	41.8	94.6	246.67
4	Germany	Europe	78.7	30.4	99	240
5	Mexico	Americas	75.1	10	73.4	230
6	France	Europe	79.5	29.9	108.7	220

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:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
c <sub>1</sub>	Costa-Rica	Europe	78.533	29.733	96.367	243.888

### 2.1.4.2 Cluster $c_2$

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
c <sub>2</sub>	United-Kingdom	Europe	78.4	30.3	157.2	236.67

The members of this cluster are:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
c <sub>2</sub>	United-Kingdom	Europe	78.4	30.3	157.2	236.67

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:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
c <sub>2</sub>	United-Kingdom	Europe	78.4	30.3	157.2	236.67

### 2.1.4.3 Cluster $c_3$

Let us refine the center of cluster  $c_3$  which is currently centered at:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
c <sub>3</sub>	South-Africa	Asia	68.314	13.343	75.714	190.473

The members of this cluster are:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
7	Thailand	Asia	70	8.3	79	216.67
8	Brazil	Americas	70.5	8.4	103.2	210
9	Japan	Asia	82	31.5	102.1	206.67
10	India	Asia	63.3	3.3	49.9	180
11	Ethiopia	Africa	47.6	0.9	5.2	156.67
12	Russia	Asia	65.3	11.1	81.9	143.3

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:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
c <sub>3</sub>	South-Africa	Asia	66.450	10.583	70.217	185.552

## 2.1.5 Final Clusters

### 2.1.5.1 Cluster $c_1$

The center of this cluster is:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
c <sub>1</sub>	Costa-Rica	Europe	78.533	29.733	96.367	243.888

The members of this cluster are:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
1	Switzerland	Europe	80.5	32.3	99.9	273.33
2	Canada	Americas	80	34	102.6	253.33
3	Usa	Americas	77.4	41.8	94.6	246.67
4	Germany	Europe	78.7	30.4	99	240
5	Mexico	Americas	75.1	10	73.4	230
6	France	Europe	79.5	29.9	108.7	220

### 2.1.5.2 Cluster $c_2$

The center of this cluster is:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$c_2$	United-Kingdom	Europe	78.4	30.3	157.2	236.67

The members of this cluster are:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
---	---------	-----------	-----------------	----------------	---------------------------	-----------

### 2.1.5.3 Cluster $c_3$

The center of this cluster is:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$c_3$	South-Africa	Asia	66.450	10.583	70.217	185.552

The members of this cluster are:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
7	Thailand	Asia	70	8.3	79	216.67
8	Brazil	Americas	70.5	8.4	103.2	210
9	Japan	Asia	82	31.5	102.1	206.67
10	India	Asia	63.3	3.3	49.9	180
11	Ethiopia	Africa	47.6	0.9	5.2	156.67
12	Russia	Asia	65.3	11.1	81.9	143.3

## 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:

- **country:** We are ignoring this attribute.
- **continent:** This is not a numeric attribute so we leave it alone.
- **life-expectancy:** minimum: 47.600, maximum: 82.000
- **GDP-per-capita:** minimum: 0.900, maximum: 41.800
- **access-to-education-score:** minimum: 5.200, maximum: 157.200
- **SWL-index:** minimum: 143.300, maximum: 273.330

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:



#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
1	Switzerland	Europe	0.956	0.768	0.623	1.000
2	Canada	Americas	0.942	0.809	0.641	0.846
3	Usa	Americas	0.866	1.000	0.588	0.795
4	Germany	Europe	0.904	0.721	0.617	0.744
5	Mexico	Americas	0.799	0.222	0.449	0.667
6	France	Europe	0.927	0.709	0.681	0.590
7	Thailand	Asia	0.651	0.181	0.486	0.564
8	Brazil	Americas	0.666	0.183	0.645	0.513
9	Japan	Asia	1.000	0.748	0.637	0.487
10	India	Asia	0.456	0.059	0.294	0.282
11	Ethiopia	Africa	0.000	0.000	0.000	0.103
12	Russia	Asia	0.515	0.249	0.505	0.000

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
1	Costa-Rica	Americas	0.890	0.249	0.301	0.821
2	United-Kingdom	Europe	0.895	0.719	1.000	0.718
3	South-Africa	Africa	0.023	0.271	0.559	0.359

## 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:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
1	Switzerland	Europe	0.956	0.768	0.623	1.000

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
c <sub>2</sub>	United-Kingdom	Europe	0.895	0.719	1.000	0.718	0.477
c <sub>1</sub>	Costa-Rica	Americas	0.890	0.249	0.301	0.821	1.187
c <sub>3</sub>	South-Africa	Africa	0.023	0.271	0.559	0.359	1.591

Thus we assign this instance to cluster  $c_2$ .

### 2.2.2.2 Instance 2

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
2	Canada	Americas	0.942	0.809	0.641	0.846

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
c <sub>1</sub>	Costa-Rica	Americas	0.890	0.249	0.301	0.821	0.658
c <sub>2</sub>	United-Kingdom	Europe	0.895	0.719	1.000	0.718	1.075
c <sub>3</sub>	South-Africa	Africa	0.023	0.271	0.559	0.359	1.542

Thus we assign this instance to cluster  $c_1$ .

**2.2.2.3 Instance 3**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
3	Usa	Americas	0.866	1.000	0.588	0.795

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_1$	Costa-Rica	Americas	0.890	0.249	0.301	0.821	0.805
$c_2$	United-Kingdom	Europe	0.895	0.719	1.000	0.718	1.120
$c_3$	South-Africa	Africa	0.023	0.271	0.559	0.359	1.560

Thus we assign this instance to cluster  $c_1$ .

**2.2.2.4 Instance 4**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
4	Germany	Europe	0.904	0.721	0.617	0.744

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_2$	United-Kingdom	Europe	0.895	0.719	1.000	0.718	0.384
$c_1$	Costa-Rica	Americas	0.890	0.249	0.301	0.821	1.153
$c_3$	South-Africa	Africa	0.023	0.271	0.559	0.359	1.460

Thus we assign this instance to cluster  $c_2$ .

**2.2.2.5 Instance 5**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
5	Mexico	Americas	0.799	0.222	0.449	0.667

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_1$	Costa-Rica	Americas	0.890	0.249	0.301	0.821	0.234
$c_2$	United-Kingdom	Europe	0.895	0.719	1.000	0.718	1.250
$c_3$	South-Africa	Africa	0.023	0.271	0.559	0.359	1.308

Thus we assign this instance to cluster  $c_1$ .

**2.2.2.6 Instance 6**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
6	France	Europe	0.927	0.709	0.681	0.590

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_2$	United-Kingdom	Europe	0.895	0.719	1.000	0.718	0.345
$c_1$	Costa-Rica	Americas	0.890	0.249	0.301	0.821	1.188
$c_3$	South-Africa	Africa	0.023	0.271	0.559	0.359	1.441

Thus we assign this instance to cluster  $c_2$ .

**2.2.2.7 Instance 7**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
7	Thailand	Asia	0.651	0.181	0.486	0.564

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_1$	Costa-Rica	Americas	0.890	0.249	0.301	0.821	1.078
$c_3$	South-Africa	Africa	0.023	0.271	0.559	0.359	1.204
$c_2$	United-Kingdom	Europe	0.895	0.719	1.000	0.718	1.279

Thus we assign this instance to cluster  $c_1$ .

**2.2.2.8 Instance 8**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
8	Brazil	Americas	0.666	0.183	0.645	0.513

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_1$	Costa-Rica	Americas	0.890	0.249	0.301	0.821	0.517
$c_3$	South-Africa	Africa	0.023	0.271	0.559	0.359	1.205
$c_2$	United-Kingdom	Europe	0.895	0.719	1.000	0.718	1.228

Thus we assign this instance to cluster  $c_1$ .

**2.2.2.9 Instance 9**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
9	Japan	Asia	1.000	0.748	0.637	0.487

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_2$	United-Kingdom	Europe	0.895	0.719	1.000	0.718	1.094
$c_1$	Costa-Rica	Americas	0.890	0.249	0.301	0.821	1.219
$c_3$	South-Africa	Africa	0.023	0.271	0.559	0.359	1.485

Thus we assign this instance to cluster  $c_2$ .

**2.2.2.10 Instance 10**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
10	India	Asia	0.456	0.059	0.294	0.282

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_3$	South-Africa	Africa	0.023	0.271	0.559	0.359	1.144
$c_1$	Costa-Rica	Americas	0.890	0.249	0.301	0.821	1.231
$c_2$	United-Kingdom	Europe	0.895	0.719	1.000	0.718	1.522

Thus we assign this instance to cluster  $c_3$ .

**2.2.2.11 Instance 11**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
11	Ethiopia	Africa	0.000	0.000	0.000	0.103

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_3$	South-Africa	Africa	0.023	0.271	0.559	0.359	0.672
$c_1$	Costa-Rica	Americas	0.890	0.249	0.301	0.821	1.569
$c_2$	United-Kingdom	Europe	0.895	0.719	1.000	0.718	1.923

Thus we assign this instance to cluster  $c_3$ .

**2.2.2.12 Instance 12**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
12	Russia	Asia	0.515	0.249	0.505	0.000

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_3$	South-Africa	Africa	0.023	0.271	0.559	0.359	1.172
$c_1$	Costa-Rica	Americas	0.890	0.249	0.301	0.821	1.362
$c_2$	United-Kingdom	Europe	0.895	0.719	1.000	0.718	1.458

Thus we assign this instance to cluster  $c_3$ .

**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:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$c_1$	Costa-Rica	Americas	0.890	0.249	0.301	0.821

The members of this cluster are:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
2	Canada	Americas	0.942	0.809	0.641	0.846
3	Usa	Americas	0.866	1.000	0.588	0.795
5	Mexico	Americas	0.799	0.222	0.449	0.667
7	Thailand	Asia	0.651	0.181	0.486	0.564
8	Brazil	Americas	0.666	0.183	0.645	0.513

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:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$c_1$	Costa-Rica	Americas	0.785	0.479	0.562	0.677

### 2.2.3.2 Cluster $c_2$

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$c_2$	United-Kingdom	Europe	0.895	0.719	1.000	0.718

The members of this cluster are:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
1	Switzerland	Europe	0.956	0.768	0.623	1.000
4	Germany	Europe	0.904	0.721	0.617	0.744
6	France	Europe	0.927	0.709	0.681	0.590
9	Japan	Asia	1.000	0.748	0.637	0.487

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:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$c_2$	United-Kingdom	Europe	0.947	0.736	0.639	0.705

### 2.2.3.3 Cluster $c_3$

Let us refine the center of cluster  $c_3$  which is currently centered at:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$c_3$	South-Africa	Africa	0.023	0.271	0.559	0.359

The members of this cluster are:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
10	India	Asia	0.456	0.059	0.294	0.282
11	Ethiopia	Africa	0.000	0.000	0.000	0.103
12	Russia	Asia	0.515	0.249	0.505	0.000

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:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$c_3$	South-Africa	Asia	0.324	0.103	0.266	0.128

## 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:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
1	Switzerland	Europe	0.956	0.768	0.623	1.000

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_2$	United-Kingdom	Europe	0.947	0.736	0.639	0.705	0.297
$c_1$	Costa-Rica	Americas	0.785	0.479	0.562	0.677	1.105
$c_3$	South-Africa	Asia	0.324	0.103	0.266	0.128	1.652

Thus we assign this instance to cluster  $c_2$ .

**2.2.4.2 Instance 2**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
2	Canada	Americas	0.942	0.809	0.641	0.846

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_1$	Costa-Rica	Americas	0.785	0.479	0.562	0.677	0.410
$c_2$	United-Kingdom	Europe	0.947	0.736	0.639	0.705	1.013
$c_3$	South-Africa	Asia	0.324	0.103	0.266	0.128	1.593

Thus we assign this instance to cluster  $c_1$ .

**2.2.4.3 Instance 3**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
3	Usa	Americas	0.866	1.000	0.588	0.795

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_1$	Costa-Rica	Americas	0.785	0.479	0.562	0.677	0.541
$c_2$	United-Kingdom	Europe	0.947	0.736	0.639	0.705	1.043
$c_3$	South-Africa	Asia	0.324	0.103	0.266	0.128	1.627

Thus we assign this instance to cluster  $c_1$ .

**2.2.4.4 Instance 4**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
4	Germany	Europe	0.904	0.721	0.617	0.744

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_2$	United-Kingdom	Europe	0.947	0.736	0.639	0.705	0.064
$c_1$	Costa-Rica	Americas	0.785	0.479	0.562	0.677	1.039
$c_3$	South-Africa	Asia	0.324	0.103	0.266	0.128	1.490

Thus we assign this instance to cluster  $c_2$ .

**2.2.4.5 Instance 5**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
5	Mexico	Americas	0.799	0.222	0.449	0.667

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_1$	Costa-Rica	Americas	0.785	0.479	0.562	0.677	0.281
$c_2$	United-Kingdom	Europe	0.947	0.736	0.639	0.705	1.150
$c_3$	South-Africa	Asia	0.324	0.103	0.266	0.128	1.251

Thus we assign this instance to cluster  $c_1$ .

**2.2.4.6 Instance 6**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
6	France	Europe	0.927	0.709	0.681	0.590

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_2$	United-Kingdom	Europe	0.947	0.736	0.639	0.705	0.127
$c_1$	Costa-Rica	Americas	0.785	0.479	0.562	0.677	1.046
$c_3$	South-Africa	Asia	0.324	0.103	0.266	0.128	1.455

Thus we assign this instance to cluster  $c_2$ .

**2.2.4.7 Instance 7**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
7	Thailand	Asia	0.651	0.181	0.486	0.564

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_3$	South-Africa	Asia	0.324	0.103	0.266	0.128	0.593
$c_1$	Costa-Rica	Americas	0.785	0.479	0.562	0.677	1.061
$c_2$	United-Kingdom	Europe	0.947	0.736	0.639	0.705	1.200

Thus we assign this instance to cluster  $c_3$ .

**2.2.4.8 Instance 8**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
8	Brazil	Americas	0.666	0.183	0.645	0.513

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_1$	Costa-Rica	Americas	0.785	0.479	0.562	0.677	0.368
$c_3$	South-Africa	Asia	0.324	0.103	0.266	0.128	1.190
$c_2$	United-Kingdom	Europe	0.947	0.736	0.639	0.705	1.192

Thus we assign this instance to cluster  $c_1$ .

**2.2.4.9 Instance 9**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
9	Japan	Asia	1.000	0.748	0.637	0.487

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_2$	United-Kingdom	Europe	0.947	0.736	0.639	0.705	1.025
$c_3$	South-Africa	Asia	0.324	0.103	0.266	0.128	1.067
$c_1$	Costa-Rica	Americas	0.785	0.479	0.562	0.677	1.077

Thus we assign this instance to cluster  $c_2$ .

**2.2.4.10 Instance 10**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
10	India	Asia	0.456	0.059	0.294	0.282

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_3$	South-Africa	Asia	0.324	0.103	0.266	0.128	0.209
$c_1$	Costa-Rica	Americas	0.785	0.479	0.562	0.677	1.230
$c_2$	United-Kingdom	Europe	0.947	0.736	0.639	0.705	1.413

Thus we assign this instance to cluster  $c_3$ .

**2.2.4.11 Instance 11**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
11	Ethiopia	Africa	0.000	0.000	0.000	0.103

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_3$	South-Africa	Asia	0.324	0.103	0.266	0.128	1.089
$c_1$	Costa-Rica	Americas	0.785	0.479	0.562	0.677	1.578
$c_2$	United-Kingdom	Europe	0.947	0.736	0.639	0.705	1.791

Thus we assign this instance to cluster  $c_3$ .

**2.2.4.12 Instance 12**

We are assigning the following instances to a cluster:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
12	Russia	Asia	0.515	0.249	0.505	0.000

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index	distance
$c_3$	South-Africa	Asia	0.324	0.103	0.266	0.128	0.362
$c_1$	Costa-Rica	Americas	0.785	0.479	0.562	0.677	1.260
$c_2$	United-Kingdom	Europe	0.947	0.736	0.639	0.705	1.392

Thus we assign this instance to cluster  $c_3$ .

**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:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$c_1$	Costa-Rica	Americas	0.785	0.479	0.562	0.677

The members of this cluster are:



#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
2	Canada	Americas	0.942	0.809	0.641	0.846
3	Usa	Americas	0.866	1.000	0.588	0.795
5	Mexico	Americas	0.799	0.222	0.449	0.667
8	Brazil	Americas	0.666	0.183	0.645	0.513

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:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$c_1$	Costa-Rica	Americas	0.818	0.553	0.581	0.705

### 2.2.5.2 Cluster $c_2$

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

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$c_2$	United-Kingdom	Europe	0.947	0.736	0.639	0.705

The members of this cluster are:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
1	Switzerland	Europe	0.956	0.768	0.623	1.000
4	Germany	Europe	0.904	0.721	0.617	0.744
6	France	Europe	0.927	0.709	0.681	0.590
9	Japan	Asia	1.000	0.748	0.637	0.487

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:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$c_2$	United-Kingdom	Europe	0.947	0.736	0.639	0.705

### 2.2.5.3 Cluster $c_3$

Let us refine the center of cluster  $c_3$  which is currently centered at:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$c_3$	South-Africa	Asia	0.324	0.103	0.266	0.128

The members of this cluster are:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
7	Thailand	Asia	0.651	0.181	0.486	0.564
10	India	Asia	0.456	0.059	0.294	0.282
11	Ethiopia	Africa	0.000	0.000	0.000	0.103
12	Russia	Asia	0.515	0.249	0.505	0.000

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:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$c_3$	South-Africa	Asia	0.405	0.122	0.321	0.237

## 2.2.6 Final Clusters

### 2.2.6.1 Cluster $c_1$

The center of this cluster is:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$c_1$	Costa-Rica	Americas	0.818	0.553	0.581	0.705

The members of this cluster are:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
2	Canada	Americas	0.942	0.809	0.641	0.846
3	Usa	Americas	0.866	1.000	0.588	0.795
5	Mexico	Americas	0.799	0.222	0.449	0.667
8	Brazil	Americas	0.666	0.183	0.645	0.513

### 2.2.6.2 Cluster $c_2$

The center of this cluster is:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$c_2$	United-Kingdom	Europe	0.947	0.736	0.639	0.705

The members of this cluster are:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
1	Switzerland	Europe	0.956	0.768	0.623	1.000
4	Germany	Europe	0.904	0.721	0.617	0.744
6	France	Europe	0.927	0.709	0.681	0.590
9	Japan	Asia	1.000	0.748	0.637	0.487

### 2.2.6.3 Cluster $c_3$

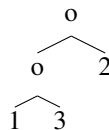
The center of this cluster is:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
$c_3$	South-Africa	Asia	0.405	0.122	0.321	0.237

The members of this cluster are:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
7	Thailand	Asia	0.651	0.181	0.486	0.564
10	India	Asia	0.456	0.059	0.294	0.282
11	Ethiopia	Africa	0.000	0.000	0.000	0.103
12	Russia	Asia	0.515	0.249	0.505	0.000

## 3 Clustering - Hierarchical Clustering

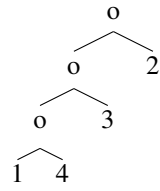


### 3.1 Number of Alternatives

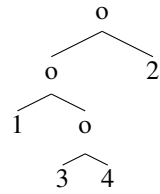
There are **nine** alternatives for the placement of the fourth instance.

### 3.2 Alternatives

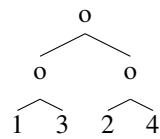
1. Place in cluster with {1}:



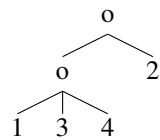
2. Place in cluster with {3}:



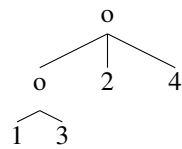
3. Place in cluster with {2}:



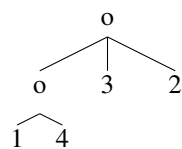
4. Place in cluster with {1} and {3}:



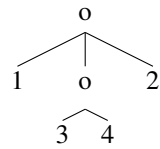
5. Place in cluster with {2} and {1,3}:



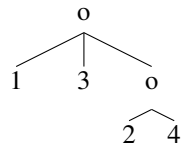
6. Split {1, 3} and place with {1}:



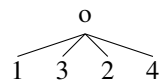
7. Split {1, 3} and place with {3}:



8. Split {1, 3} and place with {2}:



9. Split {1, 3} and place with {1, 2, 3}:



Note that merging is not possible here as there are only two (sub)clusters at each level. Normally we would consider the best and the second best “hosts” on each level of the hierarchy as candidates for a join but since there are exactly two in each level, such a join would result in a hierarchy that is equivalent to the one without the join.

### 3.3 Choice

See class text (pg. 260).