

CS4445 B Term 2006 Homework 3 Solutions

Contents

1	Linear Regression	2
2	Regression Trees and Model Trees	2
2.1	Continent Attribute Conversion	2
2.1.1	continent=Europe	3
2.1.2	continent=Africa	3
2.1.3	continent=Asia	3
2.1.4	continent=Americas	3
2.1.5	New Attributes	4
2.1.6	Converted Dataset	4
2.2	Tree Construction	4
2.2.1	Split	5
2.2.1.1	Split Points for Asia-Americas-Europe	5
2.2.1.2	Split Points for Americas-Europe	5
2.2.1.3	Split Points for Europe	5
2.2.1.4	Split Points for life-expectancy	6
2.2.1.5	Split Points for GDP-per-capita	6
2.2.1.6	Split Points for access-to-education-score	7
2.2.1.7	All Split Points	7
2.2.2	Split	9
2.2.2.1	Split Points for Asia-Americas-Europe	9
2.2.2.2	Split Points for Americas-Europe	10
2.2.2.3	Split Points for Europe	10
2.2.2.4	Split Points for life-expectancy	10
2.2.2.5	Split Points for GDP-per-capita	10
2.2.2.6	Split Points for access-to-education-score	11
2.2.2.7	All Split Points	12
2.2.3	Split	13
2.2.3.1	Split Points for Asia-Americas-Europe	13
2.2.3.2	Split Points for Americas-Europe	13
2.2.3.3	Split Points for Europe	13
2.2.3.4	Split Points for life-expectancy	13
2.2.3.5	Split Points for GDP-per-capita	14
2.2.3.6	Split Points for access-to-education-score	14
2.2.3.7	All Split Points	14
2.2.4	Final Tree	17
2.2.5	Final Models	18
2.2.6	M0	18

2.2.7	M1	18
2.2.8	M2	19
2.2.9	M3	19
3	Testing	20
3.1	Instance 1	20
3.1.1	Linear Regression	20
3.1.2	Model Tree	21
3.1.3	Regression Tree	21
3.2	Instance 2	21
3.2.1	Linear Regression	22
3.2.2	Model Tree	22
3.2.3	Regression Tree	23
3.3	Instance 3	23
3.3.1	Linear Regression	23
3.3.2	Model Tree	23
3.3.3	Regression Tree	24
3.4	Instance 4	24
3.4.1	Linear Regression	24
3.4.2	Model Tree	24
3.4.3	Regression Tree	25
3.5	Testing Results	25
3.5.1	Predicted Values	25
3.5.2	Error Measures	26
3.5.2.1	Root-Mean Squared Error	26
3.5.2.2	Mean Absolute Error	26
3.5.3	Error Summary	26

1 Linear Regression

See text for the descriptions of the algorithms and procedures.

$$\begin{aligned}
 \text{SWL-index} &= \\
 &-4.039 * \text{Asia-Americas-Europe} \\
 &+36.553 * \text{Americas-Europe} \\
 &+0.281 * \text{Europe} \\
 &+4.083 * \text{life-expectancy} \\
 &+0.139 * \text{GDP-per-capita} \\
 &-0.819 * \text{access-to-education-score} \\
 &+ - 33.569
 \end{aligned}$$

2 Regression Trees and Model Trees

2.1 Continent Attribute Conversion

Let us start by converting the nominal **continent** attribute into several numeric boolean attributes. We have, to begin with, the instances in this form:

#	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

We remove the **country** attribute and proceed to convert the **continent** to binary values. We several values for the attribute: *Europe,Africa,Asia,Americas*

2.1.1 continent=*Europe*

#	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
1	Europe	80.5	32.3	99.9	273.33
4	Europe	78.7	30.4	99	240
6	Europe	79.5	29.9	108.7	220

We see the average class over these instances if 244.443.

2.1.2 continent=*Africa*

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

We see the average class over these instances if 156.670.

2.1.3 continent=*Asia*

#	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
7	Asia	70	8.3	79	216.67
9	Asia	82	31.5	102.1	206.67
10	Asia	63.3	3.3	49.9	180
12	Asia	65.3	11.1	81.9	143.3

We see the average class over these instances if 186.660.

2.1.4 continent=*Americas*

#	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
2	Americas	80	34	102.6	253.33
3	Americas	77.4	41.8	94.6	246.67
5	Americas	75.1	10	73.4	230
8	Americas	70.5	8.4	103.2	210

We see the average class over these instances if 235.000.

2.1.5 New Attributes

The attribute values sorted by average class value are thus: *Europe, Americas, Asia, Africa*. We create new binary attributes accordingly:

- *Asia-Americas-Europe*
- *Americas-Europe*
- *Europe*

2.1.6 Converted Dataset

Our final dataset, in its converted form, is:

#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
1	1	1	1	80.5	32.3	99.9	273.33
2	1	1	0	80	34	102.6	253.33
3	1	1	0	77.4	41.8	94.6	246.67
4	1	1	1	78.7	30.4	99	240
5	1	1	0	75.1	10	73.4	230
6	1	1	1	79.5	29.9	108.7	220
7	1	0	0	70	8.3	79	216.67
8	1	1	0	70.5	8.4	103.2	210
9	1	0	0	82	31.5	102.1	206.67
10	1	0	0	63.3	3.3	49.9	180
11	0	0	0	47.6	0.9	5.2	156.67
12	1	0	0	65.3	11.1	81.9	143.3

2.2 Tree Construction

Next we will construct a model tree. First we compute the standard deviation over all the instances arriving with 37.209. We will therefore use $0.05 * 37.209 = 1.860$ as one of our stopping criteria. Our tree so far:

?

The instances applicable to this node are:

#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
1	1	1	1	80.5	32.3	99.9	273.33
2	1	1	0	80	34	102.6	253.33
3	1	1	0	77.4	41.8	94.6	246.67
4	1	1	1	78.7	30.4	99	240
5	1	1	0	75.1	10	73.4	230
6	1	1	1	79.5	29.9	108.7	220
7	1	0	0	70	8.3	79	216.67
8	1	1	0	70.5	8.4	103.2	210
9	1	0	0	82	31.5	102.1	206.67
10	1	0	0	63.3	3.3	49.9	180
11	0	0	0	47.6	0.9	5.2	156.67
12	1	0	0	65.3	11.1	81.9	143.3

Let us first check stop conditions. The standard deviation over these instances is 37.209 and there are/is 12 of them. The conditions for stopping are not met so we will look for the best split point.

2.2.1 Split

2.2.1.1 Split Points for Asia-Americas-Europe

Sorting our instances by **Asia-Americas-Europe** gives us:

#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
11	0	0	0	47.6	0.9	5.2	156.67
1	1	1	1	80.5	32.3	99.9	273.33
2	1	1	0	80	34	102.6	253.33
3	1	1	0	77.4	41.8	94.6	246.67
4	1	1	1	78.7	30.4	99	240
5	1	1	0	75.1	10	73.4	230
6	1	1	1	79.5	29.9	108.7	220
7	1	0	0	70	8.3	79	216.67
8	1	1	0	70.5	8.4	103.2	210
9	1	0	0	82	31.5	102.1	206.67
10	1	0	0	63.3	3.3	49.9	180
12	1	0	0	65.3	11.1	81.9	143.3

This results in these split points: *0.500*.

2.2.1.2 Split Points for Americas-Europe

Sorting our instances by **Americas-Europe** gives us:

#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
7	1	0	0	70	8.3	79	216.67
9	1	0	0	82	31.5	102.1	206.67
10	1	0	0	63.3	3.3	49.9	180
11	0	0	0	47.6	0.9	5.2	156.67
12	1	0	0	65.3	11.1	81.9	143.3
1	1	1	1	80.5	32.3	99.9	273.33
2	1	1	0	80	34	102.6	253.33
3	1	1	0	77.4	41.8	94.6	246.67
4	1	1	1	78.7	30.4	99	240
5	1	1	0	75.1	10	73.4	230
6	1	1	1	79.5	29.9	108.7	220
8	1	1	0	70.5	8.4	103.2	210

This results in these split points: *0.500*.

2.2.1.3 Split Points for Europe

Sorting our instances by **Europe** gives us:

#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
2	1	1	0	80	34	102.6	253.33
3	1	1	0	77.4	41.8	94.6	246.67
5	1	1	0	75.1	10	73.4	230
7	1	0	0	70	8.3	79	216.67
8	1	1	0	70.5	8.4	103.2	210
9	1	0	0	82	31.5	102.1	206.67
10	1	0	0	63.3	3.3	49.9	180
11	0	0	0	47.6	0.9	5.2	156.67
12	1	0	0	65.3	11.1	81.9	143.3
1	1	1	1	80.5	32.3	99.9	273.33
4	1	1	1	78.7	30.4	99	240
6	1	1	1	79.5	29.9	108.7	220

This results in these split points: *0.500*.

2.2.1.4 Split Points for life-expectancy

Sorting our instances by **life-expectancy** gives us:

#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
11	0	0	0	47.6	0.9	5.2	156.67
10	1	0	0	63.3	3.3	49.9	180
12	1	0	0	65.3	11.1	81.9	143.3
7	1	0	0	70	8.3	79	216.67
8	1	1	0	70.5	8.4	103.2	210
5	1	1	0	75.1	10	73.4	230
3	1	1	0	77.4	41.8	94.6	246.67
4	1	1	1	78.7	30.4	99	240
6	1	1	1	79.5	29.9	108.7	220
2	1	1	0	80	34	102.6	253.33
1	1	1	1	80.5	32.3	99.9	273.33
9	1	0	0	82	31.5	102.1	206.67

This results in these split points: *55.450, 64.300, 67.650, 70.250, 72.800, 76.250, 78.050, 79.100, 79.750, 80.250, 81.250*.

2.2.1.5 Split Points for GDP-per-capita

Sorting our instances by **GDP-per-capita** gives us:

#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
11	0	0	0	47.6	0.9	5.2	156.67
10	1	0	0	63.3	3.3	49.9	180
7	1	0	0	70	8.3	79	216.67
8	1	1	0	70.5	8.4	103.2	210
5	1	1	0	75.1	10	73.4	230
12	1	0	0	65.3	11.1	81.9	143.3
6	1	1	1	79.5	29.9	108.7	220
4	1	1	1	78.7	30.4	99	240
9	1	0	0	82	31.5	102.1	206.67
1	1	1	1	80.5	32.3	99.9	273.33
2	1	1	0	80	34	102.6	253.33
3	1	1	0	77.4	41.8	94.6	246.67

This results in these split points: *2.100, 5.800, 8.350, 9.200, 10.550, 20.500, 30.150, 30.950, 31.900, 33.150, 37.900.*

2.2.1.6 Split Points for access-to-education-score

Sorting our instances by **access-to-education-score** gives us:

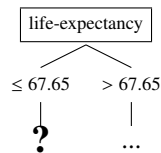
#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
11	0	0	0	47.6	0.9	5.2	156.67
10	1	0	0	63.3	3.3	49.9	180
5	1	1	0	75.1	10	73.4	230
7	1	0	0	70	8.3	79	216.67
12	1	0	0	65.3	11.1	81.9	143.3
3	1	1	0	77.4	41.8	94.6	246.67
4	1	1	1	78.7	30.4	99	240
1	1	1	1	80.5	32.3	99.9	273.33
9	1	0	0	82	31.5	102.1	206.67
2	1	1	0	80	34	102.6	253.33
8	1	1	0	70.5	8.4	103.2	210
6	1	1	1	79.5	29.9	108.7	220

This results in these split points: *27.550, 61.650, 76.200, 80.450, 88.250, 96.800, 99.450, 101.000, 102.350, 102.900, 105.950.*

2.2.1.7 All Split Points

attribute	split	instances \leq	instances $>$	std. dev. \leq	std. dev. $>$	SDR
Asia-Americas-Europe	0.500	11	1 2 3 4 5 6 7 8 9 10 12	0.000	34.295	5.771
Americas-Europe	0.500	7 9 10 11 12	1 2 3 4 5 6 8	28.089	19.737	13.992
Europe	0.500	2 3 5 7 8 9 10 11 12	1 4 6	35.945	21.997	4.751
life-expectancy	55.450	11	1 2 3 4 5 6 7 8 9 10 12	0.000	34.295	5.771
life-expectancy	64.300	10 11	1 2 3 4 5 6 7 8 9 12	11.665	33.434	7.403
life-expectancy	67.650	10 11 12	1 2 3 4 5 6 7 8 9	15.166	20.930	17.720
life-expectancy	70.250	7 10 11 12	1 2 3 4 5 6 8 9	27.836	21.342	13.702
life-expectancy	72.800	7 8 10 11 12	1 2 3 4 5 6 9	28.730	20.458	13.304
life-expectancy	76.250	5 7 8 10 11 12	1 2 3 4 6 9	31.888	21.771	10.379
life-expectancy	78.050	3 5 7 8 10 11 12	1 2 4 6 9	35.674	23.624	6.555
life-expectancy	79.100	3 4 5 7 8 10 11 12	1 2 6 9	36.195	26.402	4.278
life-expectancy	79.750	3 4 5 6 7 8 10 11 12	1 2 9	34.545	27.930	4.318
life-expectancy	80.250	2 3 4 5 6 7 8 10 11 12	1 9	35.859	33.330	1.772
life-expectancy	81.250	1 2 3 4 5 6 7 8 9 10 11 12	9	38.781	0.000	1.660
GDP-per-capita	2.100	11	1 2 3 4 5 6 7 8 9 10 12	0.000	34.295	5.771
GDP-per-capita	5.800	10 11	1 2 3 4 5 6 7 8 9 12	11.665	33.434	7.403
GDP-per-capita	8.350	7 10 11	1 2 3 4 5 6 8 9 12	24.696	35.148	4.674
GDP-per-capita	9.200	7 8 10 11	1 2 3 4 5 6 9 12	24.080	36.864	4.606
GDP-per-capita	10.550	5 7 8 10 11	1 2 3 4 6 9 12	26.633	39.386	3.137
GDP-per-capita	20.500	5 7 8 10 11 12	1 2 3 4 6 9	31.888	21.771	10.379
GDP-per-capita	30.150	5 6 7 8 10 11 12	1 2 3 4 9	31.400	21.744	9.832
GDP-per-capita	30.950	4 5 6 7 8 10 11 12	1 2 3 9	33.108	24.207	7.068
GDP-per-capita	31.900	4 5 6 7 8 9 10 11 12	1 2 3	31.293	11.329	10.907
GDP-per-capita	33.150	1 4 5 6 7 8 9 10 11 12	2 3	36.885	3.330	5.917
GDP-per-capita	37.900	1 2 4 5 6 7 8 9 10 11 12	3	37.538	0.000	2.799
access-to-education-score	27.550	11	1 2 3 4 5 6 7 8 9 10 12	0.000	34.295	5.771
access-to-education-score	61.650	10 11	1 2 3 4 5 6 7 8 9 12	11.665	33.434	7.403
access-to-education-score	76.200	5 10 11	1 2 3 4 6 7 8 9 12	30.590	35.179	3.177
access-to-education-score	80.450	5 7 10 11	1 2 3 4 6 8 9 12	29.095	37.229	2.691
access-to-education-score	88.250	5 7 10 11 12	1 2 3 4 6 8 9	33.448	22.726	10.015
access-to-education-score	96.800	3 5 7 10 11 12	1 2 4 6 8 9	38.144	24.067	6.104
access-to-education-score	99.450	3 4 5 7 10 11 12	1 2 6 8 9	38.588	26.194	3.785
access-to-education-score	101.000	1 3 4 5 7 10 11 12	2 6 8 9	43.138	18.463	2.296
access-to-education-score	102.350	1 3 4 5 7 9 10 11 12	2 6 8	40.692	18.524	2.059
access-to-education-score	102.900	1 2 3 4 5 7 9 10 11 12	6 8	40.699	5.000	2.460
access-to-education-score	105.950	1 2 3 4 5 7 8 9 10 11 12	6	38.828	0.000	1.617

We see that the best split is by life-expectancy around 67.650. We create nodes based on this split and continue. Our tree so far:



The instances applicable to this node are:

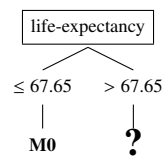
#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
10	1	0	0	63.3	3.3	49.9	180
11	0	0	0	47.6	0.9	5.2	156.67
12	1	0	0	65.3	11.1	81.9	143.3

Let us first check stop conditions. The standard deviation over these instances is 15.166 and there are/is 3 of them. The number of instances here is below our minimum of 4. We label this node with a model. The average class over

these instances is 159.990. We will use this as our RegressionTree model at the node. We also run a linearRegression fit on these instances and note the coefficients produced. The non-regressionTree model here will therefore be:

$$\begin{aligned}
 \text{SWL-index} &= \\
 &+24.828 * \text{Asia-Americas-Europe} \\
 &+0.991 * \text{life-expectancy} \\
 &-4.353 * \text{GDP-per-capita} \\
 &-0.148 * \text{access-to-education-score} \\
 &+114.193
 \end{aligned}$$

Our tree so far:



The instances applicable to this node are:

#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
1	1	1	1	80.5	32.3	99.9	273.33
2	1	1	0	80	34	102.6	253.33
3	1	1	0	77.4	41.8	94.6	246.67
4	1	1	1	78.7	30.4	99	240
5	1	1	0	75.1	10	73.4	230
6	1	1	1	79.5	29.9	108.7	220
7	1	0	0	70	8.3	79	216.67
8	1	1	0	70.5	8.4	103.2	210
9	1	0	0	82	31.5	102.1	206.67

Let us first check stop conditions. The standard deviation over these instances is 20.930 and there are/is 9 of them. The conditions for stopping are not met so we will look for the best split point.

2.2.2 Split

2.2.2.1 Split Points for Asia-Americas-Europe

Sorting our instances by Asia-Americas-Europe gives us:

#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
1	1	1	1	80.5	32.3	99.9	273.33
2	1	1	0	80	34	102.6	253.33
3	1	1	0	77.4	41.8	94.6	246.67
4	1	1	1	78.7	30.4	99	240
5	1	1	0	75.1	10	73.4	230
6	1	1	1	79.5	29.9	108.7	220
7	1	0	0	70	8.3	79	216.67
8	1	1	0	70.5	8.4	103.2	210
9	1	0	0	82	31.5	102.1	206.67

This results in these split points: .

2.2.2.2 Split Points for Americas-Europe

Sorting our instances by **Americas-Europe** gives us:

#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
7	1	0	0	70	8.3	79	216.67
9	1	0	0	82	31.5	102.1	206.67
1	1	1	1	80.5	32.3	99.9	273.33
2	1	1	0	80	34	102.6	253.33
3	1	1	0	77.4	41.8	94.6	246.67
4	1	1	1	78.7	30.4	99	240
5	1	1	0	75.1	10	73.4	230
6	1	1	1	79.5	29.9	108.7	220
8	1	1	0	70.5	8.4	103.2	210

This results in these split points: *0.500*.

2.2.2.3 Split Points for Europe

Sorting our instances by **Europe** gives us:

#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
2	1	1	0	80	34	102.6	253.33
3	1	1	0	77.4	41.8	94.6	246.67
5	1	1	0	75.1	10	73.4	230
7	1	0	0	70	8.3	79	216.67
8	1	1	0	70.5	8.4	103.2	210
9	1	0	0	82	31.5	102.1	206.67
1	1	1	1	80.5	32.3	99.9	273.33
4	1	1	1	78.7	30.4	99	240
6	1	1	1	79.5	29.9	108.7	220

This results in these split points: *0.500*.

2.2.2.4 Split Points for life-expectancy

Sorting our instances by **life-expectancy** gives us:

#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
7	1	0	0	70	8.3	79	216.67
8	1	1	0	70.5	8.4	103.2	210
5	1	1	0	75.1	10	73.4	230
3	1	1	0	77.4	41.8	94.6	246.67
4	1	1	1	78.7	30.4	99	240
6	1	1	1	79.5	29.9	108.7	220
2	1	1	0	80	34	102.6	253.33
1	1	1	1	80.5	32.3	99.9	273.33
9	1	0	0	82	31.5	102.1	206.67

This results in these split points: *70.250, 72.800, 76.250, 78.050, 79.100, 79.750, 80.250, 81.250*.

2.2.2.5 Split Points for GDP-per-capita

Sorting our instances by **GDP-per-capita** gives us:

#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
7	1	0	0	70	8.3	79	216.67
8	1	1	0	70.5	8.4	103.2	210
5	1	1	0	75.1	10	73.4	230
6	1	1	1	79.5	29.9	108.7	220
4	1	1	1	78.7	30.4	99	240
9	1	0	0	82	31.5	102.1	206.67
1	1	1	1	80.5	32.3	99.9	273.33
2	1	1	0	80	34	102.6	253.33
3	1	1	0	77.4	41.8	94.6	246.67

This results in these split points: 8.350, 9.200, 19.950, 30.150, 30.950, 31.900, 33.150, 37.900.

2.2.2.6 Split Points for access-to-education-score

Sorting our instances by **access-to-education-score** gives us:

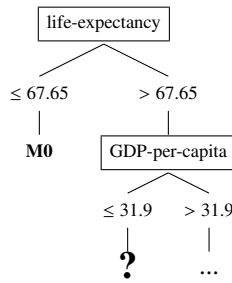
#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
5	1	1	0	75.1	10	73.4	230
7	1	0	0	70	8.3	79	216.67
3	1	1	0	77.4	41.8	94.6	246.67
4	1	1	1	78.7	30.4	99	240
1	1	1	1	80.5	32.3	99.9	273.33
9	1	0	0	82	31.5	102.1	206.67
2	1	1	0	80	34	102.6	253.33
8	1	1	0	70.5	8.4	103.2	210
6	1	1	1	79.5	29.9	108.7	220

This results in these split points: 76.200, 86.800, 96.800, 99.450, 101.000, 102.350, 102.900, 105.950.

2.2.2.7 All Split Points

attribute	split	instances \leq	instances $>$	std. dev. \leq	std. dev. $>$	SDR
Americas-Europe	0.500	7 9	1 2 3 4 5 6 8	5.000	19.737	4.468
Europe	0.500	2 3 5 7 8 9	1 4 6	17.785	21.997	1.741
life-expectancy	70.250	7	1 2 3 4 5 6 8 9	0.000	21.342	1.959
life-expectancy	72.800	7 8	1 2 3 4 5 6 9	3.335	20.458	4.277
life-expectancy	76.250	5 7 8	1 2 3 4 6 9	8.314	21.771	3.644
life-expectancy	78.050	3 5 7 8	1 2 4 6 9	14.020	23.624	1.574
life-expectancy	79.100	3 4 5 7 8	1 2 6 9	13.760	26.402	1.551
life-expectancy	79.750	3 4 5 6 7 8	1 2 9	12.970	27.930	2.973
life-expectancy	80.250	2 3 4 5 6 7 8	1 9	15.088	33.330	1.788
life-expectancy	81.250	1 2 3 4 5 6 7 8	9	19.890	0.000	3.250
GDP-per-capita	8.350	7	1 2 3 4 5 6 8 9	0.000	21.342	1.959
GDP-per-capita	9.200	7 8	1 2 3 4 5 6 9	3.335	20.458	4.277
GDP-per-capita	19.950	5 7 8	1 2 3 4 6 9	8.314	21.771	3.644
GDP-per-capita	30.150	5 6 7 8	1 2 3 4 9	7.217	21.744	5.643
GDP-per-capita	30.950	4 5 6 7 8	1 2 3 9	10.541	24.207	4.315
GDP-per-capita	31.900	4 5 6 7 8 9	1 2 3	11.452	11.329	9.519
GDP-per-capita	33.150	1 4 5 6 7 8 9	2 3	21.294	3.330	3.628
GDP-per-capita	37.900	1 2 4 5 6 7 8 9	3	21.596	0.000	1.733
access-to-education-score	76.200	5	1 2 3 4 6 7 8 9	0.000	22.172	1.222
access-to-education-score	86.800	5 7	1 2 3 4 6 8 9	6.665	22.726	1.773
access-to-education-score	96.800	3 5 7	1 2 4 6 8 9	12.273	24.067	0.794
access-to-education-score	99.450	3 4 5 7	1 2 6 8 9	11.304	26.194	1.354
access-to-education-score	101.000	1 3 4 5 7	2 6 8 9	18.925	18.463	2.210
access-to-education-score	102.350	1 3 4 5 7 9	2 6 8	21.572	18.524	0.374
access-to-education-score	102.900	1 2 3 4 5 7 9	6 8	20.918	5.000	3.550
access-to-education-score	105.950	1 2 3 4 5 7 8 9	6	21.661	0.000	1.676

We see that the best split is by GDP-per-capita around 31.900. We create nodes based on this split and continue. Our tree so far:



The instances applicable to this node are:

#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
4	1	1	1	78.7	30.4	99	240
5	1	1	0	75.1	10	73.4	230
6	1	1	1	79.5	29.9	108.7	220
7	1	0	0	70	8.3	79	216.67
8	1	1	0	70.5	8.4	103.2	210
9	1	0	0	82	31.5	102.1	206.67

Let us first check stop conditions. The standard deviation over these instances is 11.452 and there are/is 6 of them. The conditions for stopping are not met so we will look for the best split point.

2.2.3 Split

2.2.3.1 Split Points for Asia-Americas-Europe

Sorting our instances by **Asia-Americas-Europe** gives us:

#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
4	1	1	1	78.7	30.4	99	240
5	1	1	0	75.1	10	73.4	230
6	1	1	1	79.5	29.9	108.7	220
7	1	0	0	70	8.3	79	216.67
8	1	1	0	70.5	8.4	103.2	210
9	1	0	0	82	31.5	102.1	206.67

This results in these split points: .

2.2.3.2 Split Points for Americas-Europe

Sorting our instances by **Americas-Europe** gives us:

#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
7	1	0	0	70	8.3	79	216.67
9	1	0	0	82	31.5	102.1	206.67
4	1	1	1	78.7	30.4	99	240
5	1	1	0	75.1	10	73.4	230
6	1	1	1	79.5	29.9	108.7	220
8	1	1	0	70.5	8.4	103.2	210

This results in these split points: *0.500*.

2.2.3.3 Split Points for Europe

Sorting our instances by **Europe** gives us:

#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
5	1	1	0	75.1	10	73.4	230
7	1	0	0	70	8.3	79	216.67
8	1	1	0	70.5	8.4	103.2	210
9	1	0	0	82	31.5	102.1	206.67
4	1	1	1	78.7	30.4	99	240
6	1	1	1	79.5	29.9	108.7	220

This results in these split points: *0.500*.

2.2.3.4 Split Points for life-expectancy

Sorting our instances by **life-expectancy** gives us:

#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
7	1	0	0	70	8.3	79	216.67
8	1	1	0	70.5	8.4	103.2	210
5	1	1	0	75.1	10	73.4	230
4	1	1	1	78.7	30.4	99	240
6	1	1	1	79.5	29.9	108.7	220
9	1	0	0	82	31.5	102.1	206.67

This results in these split points: *70.250, 72.800, 76.900, 79.100, 80.750*.

2.2.3.5 Split Points for GDP-per-capita

Sorting our instances by **GDP-per-capita** gives us:

#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
7	1	0	0	70	8.3	79	216.67
8	1	1	0	70.5	8.4	103.2	210
5	1	1	0	75.1	10	73.4	230
6	1	1	1	79.5	29.9	108.7	220
4	1	1	1	78.7	30.4	99	240
9	1	0	0	82	31.5	102.1	206.67

This results in these split points: 8.350, 9.200, 19.950, 30.150, 30.950.

2.2.3.6 Split Points for access-to-education-score

Sorting our instances by **access-to-education-score** gives us:

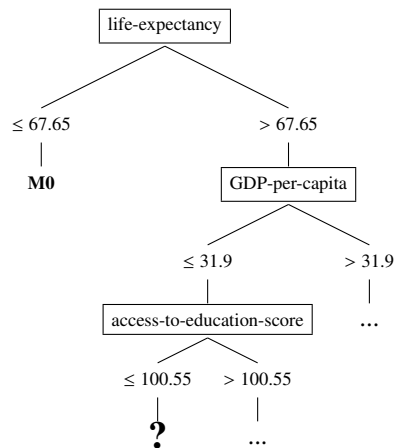
#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
5	1	1	0	75.1	10	73.4	230
7	1	0	0	70	8.3	79	216.67
4	1	1	1	78.7	30.4	99	240
9	1	0	0	82	31.5	102.1	206.67
8	1	1	0	70.5	8.4	103.2	210
6	1	1	1	79.5	29.9	108.7	220

This results in these split points: 76.200, 89.000, 100.550, 102.650, 105.950.

2.2.3.7 All Split Points

attribute	split	instaces ≤	instaces >	std. dev. ≤	std. dev. >	SDR
Americas-Europe	0.500	7 9	4 5 6 8	5.000	11.180	2.332
Europe	0.500	5 7 8 9	4 6	8.936	10.000	2.162
life-expectancy	70.250	7	4 5 6 8 9	0.000	12.400	1.119
life-expectancy	72.800	7 8	4 5 6 9	3.335	12.331	2.120
life-expectancy	76.900	5 7 8	4 6 9	8.314	13.697	0.446
life-expectancy	79.100	4 5 7 8	6 9	11.636	6.665	1.473
life-expectancy	80.750	4 5 6 7 8	9	10.541	0.000	2.668
GDP-per-capita	8.350	7	4 5 6 8 9	0.000	12.400	1.119
GDP-per-capita	9.200	7 8	4 5 6 9	3.335	12.331	2.120
GDP-per-capita	19.950	5 7 8	4 6 9	8.314	13.697	0.446
GDP-per-capita	30.150	5 6 7 8	4 9	7.217	16.665	1.086
GDP-per-capita	30.950	4 5 6 7 8	9	10.541	0.000	2.668
access-to-education-score	76.200	5	4 6 7 8 9	0.000	11.661	1.735
access-to-education-score	89.000	5 7	4 6 8 9	6.665	12.990	0.571
access-to-education-score	100.550	4 5 7	6 8 9	9.557	5.664	3.842
access-to-education-score	102.650	4 5 7 9	6 8	12.691	5.000	1.325
access-to-education-score	105.950	4 5 7 8 9	6	12.542	0.000	1.000

We see that the best split is by access-to-education-score around 100.550. We create nodes based on this split and continue. Our tree so far:



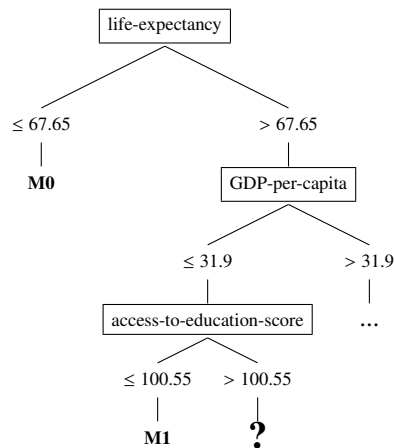
The instances applicable to this node are:

#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
4	1	1	1	78.7	30.4	99	240
5	1	1	0	75.1	10	73.4	230
7	1	0	0	70	8.3	79	216.67

Let us first check stop conditions. The standard deviation over these instances is 9.557 and there are/is 3 of them. The number of instances here is below our minimum of 4. We label this node with a model. The average class over these instances is 228.890. We will use this as our RegressionTree model at the node. We also run a linearRegression fit on these instances and note the coefficients produced. The non-regressionTree model here will therefore be:

$$\begin{aligned}
 \text{SWL-index} &= \\
 &+8.551 * \text{Americas-Europe} \\
 &+2.646 * \text{Europe} \\
 &+0.927 * \text{life-expectancy} \\
 &+0.151 * \text{GDP-per-capita} \\
 &+0.037 * \text{access-to-education-score} \\
 &+147.648
 \end{aligned}$$

Our tree so far:



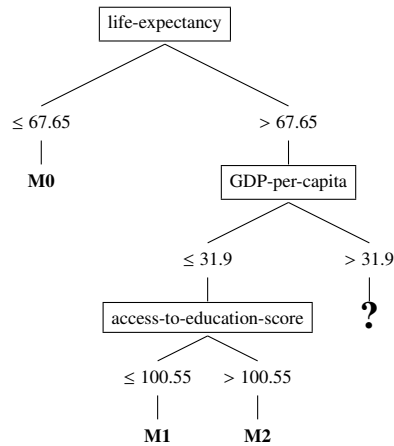
The instances applicable to this node are:

#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
6	1	1	1	79.5	29.9	108.7	220
8	1	1	0	70.5	8.4	103.2	210
9	1	0	0	82	31.5	102.1	206.67

Let us first check stop conditions. The standard deviation over these instances is 5.664 and there are/is 3 of them. The number of instances here is below our minimum of 4. We label this node with a model. The average class over these instances is 212.223. We will use this as our RegressionTree model at the node. We also run a linearRegression fit on these instances and note the coefficients produced. The non-regressionTree model here will therefore be:

$$\begin{aligned}
 \text{SWL-index} &= \\
 &+3.525 * \text{Americas-Europe} \\
 &+4.686 * \text{Europe} \\
 &+0.015 * \text{life-expectancy} \\
 &+0.039 * \text{GDP-per-capita} \\
 &+0.791 * \text{access-to-education-score} \\
 &+123.519
 \end{aligned}$$

Our tree so far:



The instances applicable to this node are:

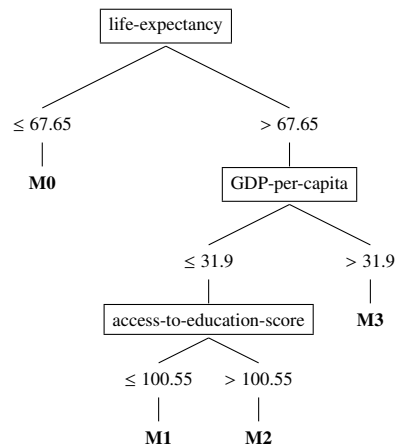
#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
1	1	1	1	80.5	32.3	99.9	273.33
2	1	1	0	80	34	102.6	253.33
3	1	1	0	77.4	41.8	94.6	246.67

Let us first check stop conditions. The standard deviation over these instances is 11.329 and there are/is 3 of them. The number of instances here is below our minimum of 4. We label this node with a model. The average class over these instances is 257.777. We will use this as our RegressionTree model at the node. We also run a linearRegression fit on these instances and note the coefficients produced. The non-regressionTree model here will therefore be:

$$\begin{aligned}
 \text{SWL-index} &= \\
 &+16.725 * \text{Europe} \\
 &+1.919 * \text{life-expectancy} \\
 &-0.665 * \text{GDP-per-capita} \\
 &-0.439 * \text{access-to-education-score} \\
 &+167.496
 \end{aligned}$$

2.2.4 Final Tree

Our final tree now is:



2.2.5 Final Models

The various sub models produced are summarized below:

2.2.6 M0

M0 for RegressionTree

$$\text{SWL-index} = 159.990$$

M0 for ModelTree

$$\begin{aligned}
 \text{SWL-index} = & \\
 & +24.828 * \text{Asia-Americas-Europe} \\
 & +0.991 * \text{life-expectancy} \\
 & -4.353 * \text{GDP-per-capita} \\
 & -0.148 * \text{access-to-education-score} \\
 & +114.193
 \end{aligned}$$

2.2.7 M1

M1 for RegressionTree

$$\text{SWL-index} = 228.890$$

M1 for ModelTree

$$\begin{aligned}\text{SWL-index} &= \\ &+8.551 * \text{Americas-Europe} \\ &+2.646 * \text{Europe} \\ &+0.927 * \text{life-expectancy} \\ &+0.151 * \text{GDP-per-capita} \\ &+0.037 * \text{access-to-education-score} \\ &+147.648\end{aligned}$$

2.2.8 M2**M2 for RegressionTree**

$$\text{SWL-index} = 212.223$$

M2 for ModelTree

$$\begin{aligned}\text{SWL-index} &= \\ &+3.525 * \text{Americas-Europe} \\ &+4.686 * \text{Europe} \\ &+0.015 * \text{life-expectancy} \\ &+0.039 * \text{GDP-per-capita} \\ &+0.791 * \text{access-to-education-score} \\ &+123.519\end{aligned}$$

2.2.9 M3**M3 for RegressionTree**

$$\text{SWL-index} = 257.777$$

M3 for ModelTree

$$\begin{aligned}
 \text{SWL-index} &= \\
 &+16.725 * \text{Europe} \\
 &+1.919 * \text{life-expectancy} \\
 &-0.665 * \text{GDP-per-capita} \\
 &-0.439 * \text{access-to-education-score} \\
 &+167.496
 \end{aligned}$$

3 Testing

We test our three models on four sample instances. We have the following instances to evaluate:

#	country	continent	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
1	Costa-Rica	Americas	78.2	11.1	50.9	250
2	United-Kingdom	Europe	78.4	30.3	157.2	236.67
3	South-Africa	Africa	48.4	12	90.2	190
4	Lithuania	Europe	72.3	3.7	93.4	156.67

We begin by converting them to our format (with country removed and continent binarized):

#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
1	1	1	0	78.2	11.1	50.9	250
2	1	1	1	78.4	30.3	157.2	236.67
3	0	0	0	48.4	12	90.2	190
4	1	1	1	72.3	3.7	93.4	156.67

We can now use our three methods to predict the target value for each instance.

3.1 Instance 1

Let us classify:

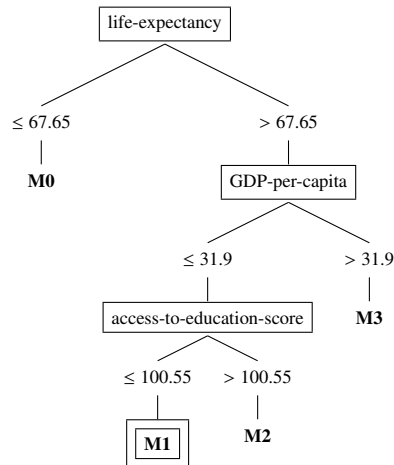
#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
1	1	1	0	78.2	11.1	50.9	250

3.1.1 Linear Regression

$$\begin{aligned}
 \text{SWL-index} &= \\
 &-4.039 * 1.000 \quad [\text{Asia-Americas-Europe}] \\
 &+36.553 * 1.000 \quad [\text{Americas-Europe}] \\
 &+0.281 * 0.000 \quad [\text{Europe}] \\
 &+4.083 * 78.200 \quad [\text{life-expectancy}] \\
 &+0.139 * 11.100 \quad [\text{GDP-per-capita}] \\
 &-0.819 * 50.900 \quad [\text{access-to-education-score}] \\
 &-33.569 = \boxed{278.125}
 \end{aligned}$$

3.1.2 Model Tree

We see that this instance trickled down to this (marked) model in our tree:



We will use this for our ModelTree and RegressionTree predictions.

$$\begin{aligned}
 \text{SWL-index} &= \\
 &+8.551 * 1.000 \quad [\text{Americas-Europe}] \\
 &+2.646 * 0.000 \quad [\text{Europe}] \\
 &+0.927 * 78.200 \quad [\text{life-expectancy}] \\
 &+0.151 * 11.100 \quad [\text{GDP-per-capita}] \\
 &+0.037 * 50.900 \quad [\text{access-to-education-score}] \\
 &+147.648 = \boxed{232.217}
 \end{aligned}$$

3.1.3 Regression Tree

$$\text{SWL-index} = 228.890$$

3.2 Instance 2

Let us classify:

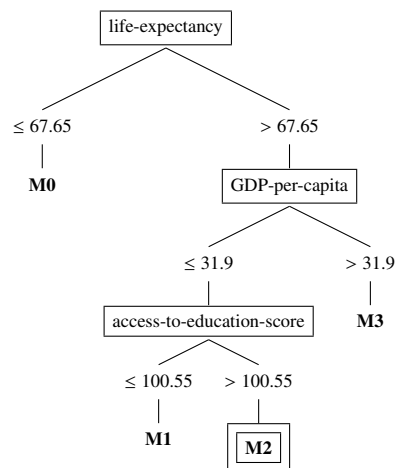
#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
2	1	1	1	78.4	30.3	157.2	236.67

3.2.1 Linear Regression

$$\begin{aligned}
 \text{SWL-index} &= \\
 &-4.039 * 1.000 \quad [\text{Asia-Americas-Europe}] \\
 &+36.553 * 1.000 \quad [\text{Americas-Europe}] \\
 &+0.281 * 1.000 \quad [\text{Europe}] \\
 &+4.083 * 78.400 \quad [\text{life-expectancy}] \\
 &+0.139 * 30.300 \quad [\text{GDP-per-capita}] \\
 &-0.819 * 157.200 \quad [\text{access-to-education-score}] \\
 -33.569 &= \boxed{194.826}
 \end{aligned}$$

3.2.2 Model Tree

We see that this instance trickled down to this (marked) model in our tree:



We will use this for our ModelTree and RegressionTree predictions.

$$\begin{aligned}
 \text{SWL-index} &= \\
 &+3.525 * 1.000 \quad [\text{Americas-Europe}] \\
 &+4.686 * 1.000 \quad [\text{Europe}] \\
 &+0.015 * 78.400 \quad [\text{life-expectancy}] \\
 &+0.039 * 30.300 \quad [\text{GDP-per-capita}] \\
 &+0.791 * 157.200 \quad [\text{access-to-education-score}] \\
 +123.519 &= \boxed{258.339}
 \end{aligned}$$

3.2.3 Regression Tree

$$\text{SWL-index} = 212.223$$

3.3 Instance 3

Let us classify:

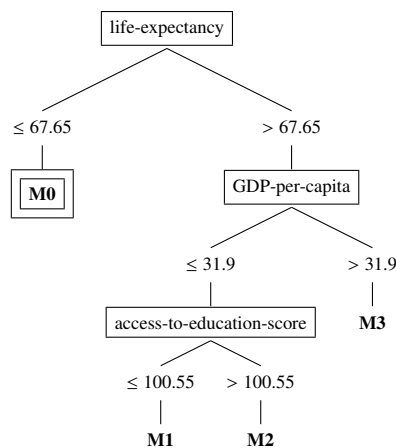
#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
3	0	0	0	48.4	12	90.2	190

3.3.1 Linear Regression

$$\begin{aligned}
 \text{SWL-index} &= \\
 &-4.039 * 0.000 \quad [\text{Asia-Americas-Europe}] \\
 &+36.553 * 0.000 \quad [\text{Americas-Europe}] \\
 &+0.281 * 0.000 \quad [\text{Europe}] \\
 &+4.083 * 48.400 \quad [\text{life-expectancy}] \\
 &+0.139 * 12.000 \quad [\text{GDP-per-capita}] \\
 &-0.819 * 90.200 \quad [\text{access-to-education-score}] \\
 &-33.569 = \boxed{91.860}
 \end{aligned}$$

3.3.2 Model Tree

We see that this instance trickled down to this (marked) model in our tree:



We will use this for our ModelTree and RegressionTree predictions.

$$\begin{aligned}
 \text{SWL-index} &= \\
 &+24.828 * 0.000 \quad [\text{Asia-Americas-Europe}] \\
 &+0.991 * 48.400 \quad [\text{life-expectancy}] \\
 &-4.353 * 12.000 \quad [\text{GDP-per-capita}] \\
 &-0.148 * 90.200 \quad [\text{access-to-education-score}] \\
 &+114.193 = \boxed{96.582}
 \end{aligned}$$

3.3.3 Regression Tree

$$\text{SWL-index} = 159.990$$

3.4 Instance 4

Let us classify:

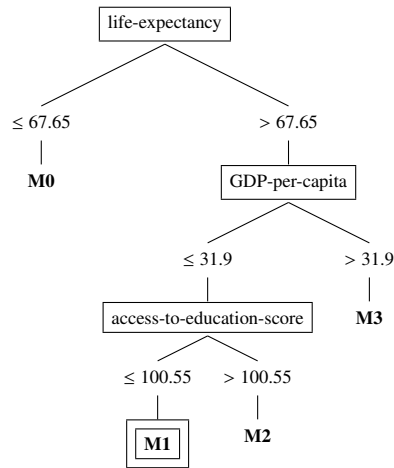
#	Asia-Americas-Europe	Americas-Europe	Europe	life-expectancy	GDP-per-capita	access-to-education-score	SWL-index
4	1	1	1	72.3	3.7	93.4	156.67

3.4.1 Linear Regression

$$\begin{aligned}
 \text{SWL-index} &= \\
 &-4.039 * 1.000 \quad [\text{Asia-Americas-Europe}] \\
 &+36.553 * 1.000 \quad [\text{Americas-Europe}] \\
 &+0.281 * 1.000 \quad [\text{Europe}] \\
 &+4.083 * 72.300 \quad [\text{life-expectancy}] \\
 &+0.139 * 3.700 \quad [\text{GDP-per-capita}] \\
 &-0.819 * 93.400 \quad [\text{access-to-education-score}] \\
 &-33.569 = \boxed{218.477}
 \end{aligned}$$

3.4.2 Model Tree

We see that this instance trickled down to this (marked) model in our tree:



We will use this for our ModelTree and RegressionTree predictions.

$$\begin{aligned}
 \text{SWL-index} &= \\
 &+8.551 * 1.000 \quad [Americas-Europe] \\
 &+2.646 * 1.000 \quad [Europe] \\
 &+0.927 * 72.300 \quad [life-expectancy] \\
 &+0.151 * 3.700 \quad [GDP-per-capita] \\
 &+0.037 * 93.400 \quad [access-to-education-score] \\
 +147.648 &= \boxed{229.831}
 \end{aligned}$$

3.4.3 Regression Tree

$$\text{SWL-index} = 228.890$$

3.5 Testing Results

3.5.1 Predicted Values

Our three methods made the following predictions:

#	SWL-index	Linear Regression	Model Tree	Regression Tree
1	250	278.125	232.217	228.890
2	236.67	194.826	258.339	212.223
3	190	91.860	96.582	159.990
4	156.67	218.477	229.831	228.890

3.5.2 Error Measures

3.5.2.1 Root-Mean Squared Error

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

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

Linear Regression

$$\left(\frac{(278.125 - 250.000)^2 + (194.826 - 236.670)^2 + (91.860 - 190.000)^2 + (218.477 - 156.670)^2}{4} \right)^{\frac{1}{2}} = 63.233$$

Model Tree

$$\left(\frac{(232.217 - 250.000)^2 + (258.339 - 236.670)^2 + (96.582 - 190.000)^2 + (229.831 - 156.670)^2}{4} \right)^{\frac{1}{2}} = 60.962$$

Regression Tree

$$\left(\frac{(228.890 - 250.000)^2 + (212.223 - 236.670)^2 + (159.990 - 190.000)^2 + (228.890 - 156.670)^2}{4} \right)^{\frac{1}{2}} = 42.307$$

3.5.2.2 Mean Absolute Error

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

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

Linear Regression

$$\frac{|278.125 - 250.000| + |194.826 - 236.670| + |91.860 - 190.000| + |218.477 - 156.670|}{4} = 57.479$$

Model Tree

$$\frac{|232.217 - 250.000| + |258.339 - 236.670| + |96.582 - 190.000| + |229.831 - 156.670|}{4} = 51.508$$

Regression Tree

$$\frac{|228.890 - 250.000| + |212.223 - 236.670| + |159.990 - 190.000| + |228.890 - 156.670|}{4} = 36.947$$

3.5.3 Error Summary

And we have our accuracy measures:

measure	Linear Regression	Model Tree	Regression Tree
root mean-squared error	63.233	60.962	42.307
mean absolute error	57.479	51.508	36.947