

HW4 Rubrics

Total: 5 points

Late Penalty: 0.5/day, after 26th April 2020 11:59pm

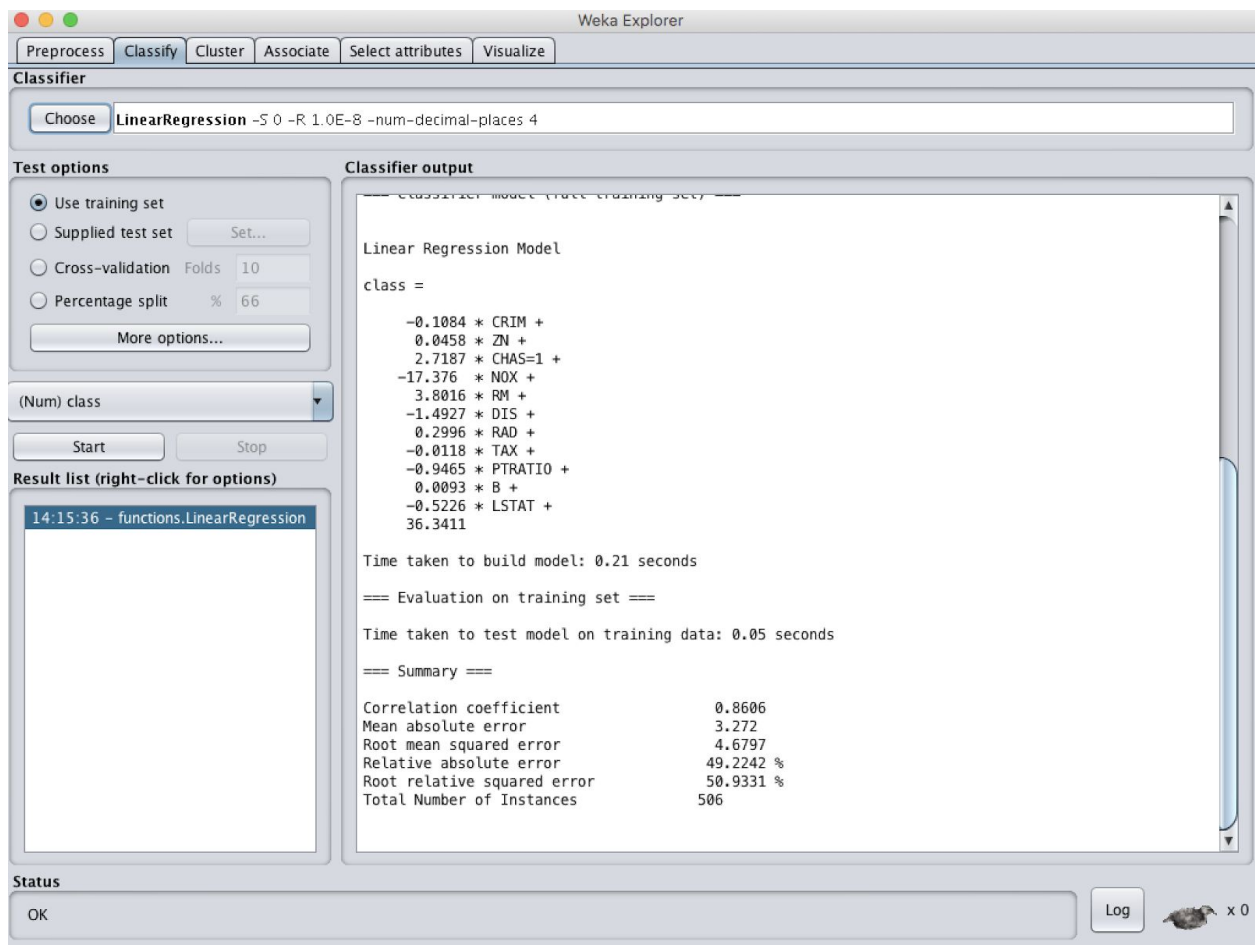
Files not zipped penalty: 0.5 point

Q1: (0.5 points)

2 possible solutions - both accepted.

Method1:

If attributeSelectionMethod = M5 method



README contents:

There are 12 terms in this equation because each term affects the MEDV. A negative value would lower the median home price whereas a positive value raises it.

(Also explaining about the M5 algorithm is fine - includes answers about Age/Indus)

Screenshot - 0.25

Number of variables + explanation of 'why' - 0.25

(If the constant is not included as a term and **is mentioned separately**, that is okay)

Explanations that mean the same thing are accepted (e.g. terms like proportional and inversely proportional, explanation with examples etc.)

For partially correct explanations, -0.1

Note:

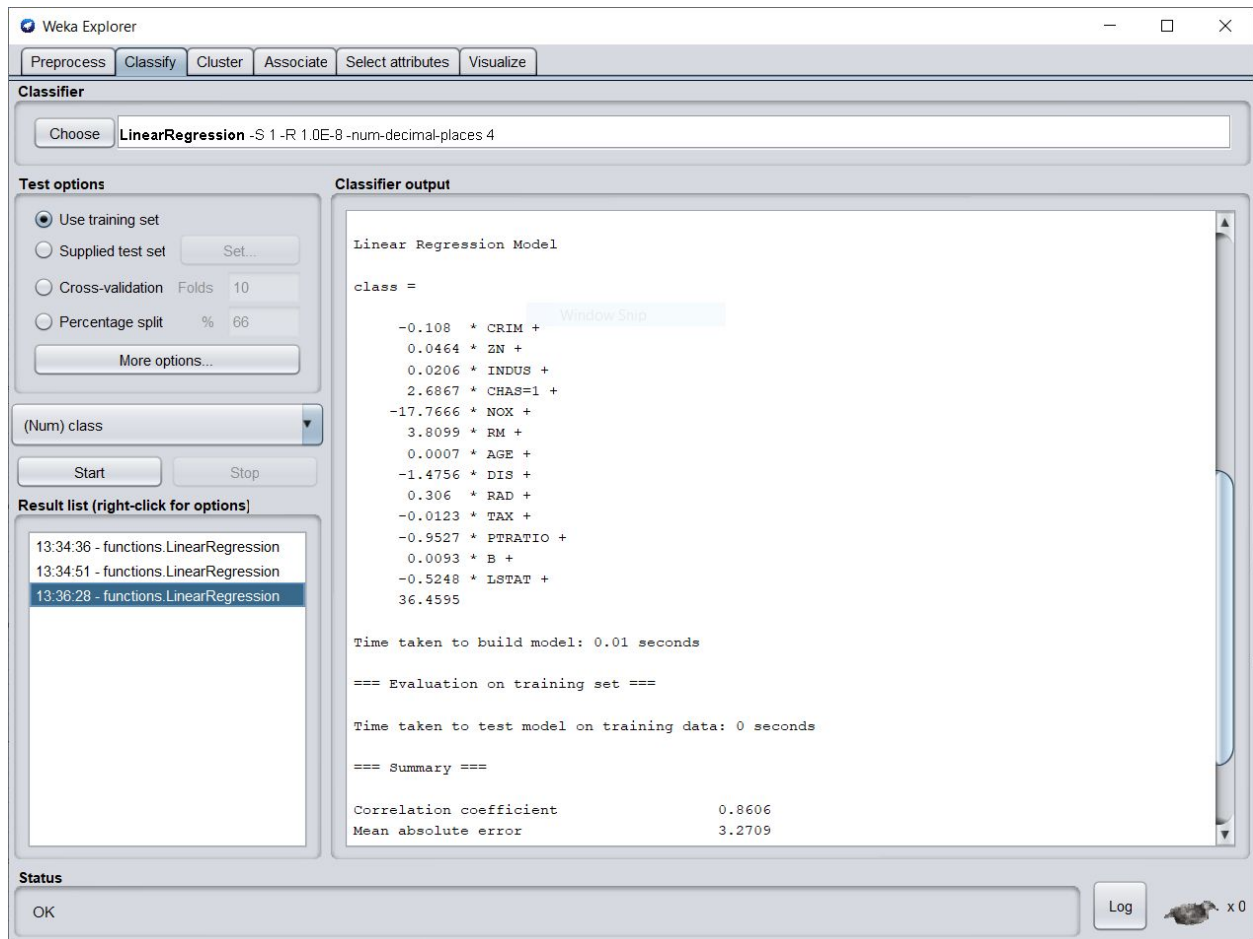
If they got a different screenshot and number of variables for the M5 method, and they have mentioned the reasons for why they got those numbers, it is acceptable.

Otherwise,

If by any chance, the screenshots do not match (number of terms other than 12, different coefficients, different summary) that means actually the student has done it incorrectly; we can deduct 0.25 overall.

Method2:

If attributeSelectionMethod = No attribute selected



Explanation:

There are 14 terms in this equation because each term affects the MEDV. A negative value would lower the median home price whereas a positive value raises it.

Screenshot - 0.25

Number of variables + explanation of 'why' - 0.25

(If the constant is not included as a term and **is mentioned separately**, that is okay)

Explanations that mean the same thing are accepted (e.g. terms like proportional and inversely proportional, explanation with examples etc.)

For partially correct explanations, -0.1

Note: If by any chance, the screenshots do not match (number of terms other than 12, different coefficients, different summary) that means actually the student has done it incorrectly; we can deduct 0.25 overall.

Q2: (1 point)

Screenshots of the entire NN model as well as of the NN.

Screenshot: **Either weights of each node OR the architecture is acceptable. (one of the 2 below)**

Weights Screenshot:

=== Classifier model (full training set) ===

```
Linear Node 0
Inputs      Weights
Threshold   1.2021459920004987
Node 1      0.9469793647859159
Node 2      -0.9618065508639184
Node 3      -0.5691565552512772
Node 4      -0.8654870866072637
Node 5      -0.9133751707145877
Node 6      1.4499192257820819
Node 7      -1.4797889220320624

Sigmoid Node 1
Inputs      Weights
Threshold   -1.1502116068380643
Attrib CRIM 1.9628047153063264
Attrib ZN    -0.27473716594392084
Attrib INDUS 1.2861395031515181
Attrib CHAS=1 -0.12394739576827847
Attrib NOX   0.06612266006626144
Attrib RM    5.20048123886782
Attrib AGE   -0.6842697182710415
Attrib DIS   0.9006229861981174
Attrib RAD   -2.0108602126705923
Attrib TAX   -1.7559327738392663
Attrib PTRATIO -1.8138366192299706
Attrib B     -2.7565808983496782
Attrib LSTAT -1.9810942057668155

Sigmoid Node 2
Inputs      Weights
Threshold   -1.20963057481019
Attrib CRIM 2.909331846277324
Attrib ZN    -1.093056355840263
Attrib INDUS 0.49361454486357
Attrib CHAS=1 0.4865246481064803
Attrib NOX   0.38365596328863105
Attrib RM    3.4563773414121606
Attrib AGE   -1.7189129601752886
Attrib DIS   1.9686040514966563
Attrib RAD   1.8953435691774174
Attrib TAX   1.866239540662793
Attrib PTRATIO 1.011084350129147
Attrib B     -1.0278766874033696
Attrib LSTAT 0.5785209695318989

Sigmoid Node 3
Inputs      Weights
Threshold   -2.4501571390341397
Attrib CRIM 2.27761691047333
Attrib ZN    0.05220369576443878
Attrib INDUS 0.345049812840946
Attrib CHAS=1 -0.7842583661439886
Attrib NOX   0.8357176581311002
Attrib RM    -2.2130405147033545
Attrib AGE   2.399146120677644
Attrib DIS   2.274694777178758
Attrib RAD   0.3486067195821872
Attrib TAX   -0.09570355788806648
Attrib PTRATIO 2.121525050157061
Attrib B     1.7200495463492116
Attrib LSTAT 2.7794342930688587

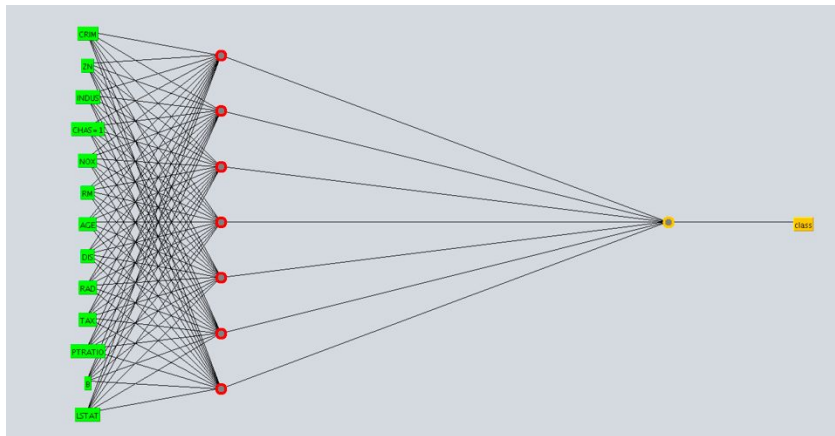
Sigmoid Node 4
Inputs      Weights
Threshold   -1.2406667852881832
Attrib CRIM 0.8329091820267875
Attrib ZN    0.5631380479824273
Attrib INDUS 0.3863175002959314
Attrib CHAS=1 1.0940379609259137
```

```

Attrib CHAS=1 1.0940379609259137
Attrib NOX 0.17850591214457867
Attrib RM -0.6191564054168993
Attrib AGE -1.4295821116663079
Attrib DIS 0.5534838004540318
Attrib RAD -0.021241281549475317
Attrib TAX -1.217810523538383
Attrib PTRATIO -2.403217402045735
Attrib B -1.3311068838785942
Attrib LSTAT -1.1172108871935837
Sigmoid Node 5
Inputs Weights
Threshold -1.4256118860751932
Attrib CRIM 0.0352101840865285
Attrib ZN -0.33517677557013664
Attrib INDUS 0.5276911902622392
Attrib CHAS=1 0.16654689988731636
Attrib NOX -0.05443358624254321
Attrib RM -0.8153217153425646
Attrib AGE -0.30818606034389345
Attrib DIS 0.9967544142513024
Attrib RAD -2.4675466177920895
Attrib TAX 0.49811983422966843
Attrib PTRATIO -0.38526814095461226
Attrib B -2.2813775634334004
Attrib LSTAT -0.3577466719956536
Sigmoid Node 6
Inputs Weights
Threshold 0.17481618618054354
Attrib CRIM 2.9196615154403855
Attrib ZN 0.9132690216672589
Attrib INDUS -1.3530265781052135
Attrib CHAS=1 0.6831920728841838
Attrib NOX -0.5890223727473257
Attrib RM 1.8028369728830815
Attrib AGE -1.4280319486148165
Attrib DIS -0.23654966446740522
Attrib RAD 0.14512694578503177
Attrib TAX 0.87721981946558
Attrib PTRATIO 0.24593390949081942
Attrib B -0.6442321426759797
Attrib LSTAT 1.0373796119000123
Sigmoid Node 7
Inputs Weights
Threshold 8.688211229130516
Attrib CRIM 0.07550444559565729
Attrib ZN -0.5282259090456871
Attrib INDUS 0.4503967404940967
Attrib CHAS=1 -0.515453184216586
Attrib NOX 1.9065022823443665
Attrib RM 0.4921823635326084
Attrib AGE 0.1543777887636418
Attrib DIS 5.214745586587453
Attrib RAD -1.7212823145503262
Attrib TAX -0.011173609919067469
Attrib PTRATIO 0.47855447929744127
Attrib B -0.843359180021431
Attrib LSTAT 4.609985528669575
Class
Input
Node 0
Time taken to build model: 0.51 seconds

```

Architecture Screenshot:



Answer:

The lowest RMSE - just under 2.6 (2.59). **Getting a lesser value is fine as long you mention the hyperparameters.**

Below is a screenshot of the RMSE.

Classifier

Choose **MultilayerPerceptron** -L 0.1 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Test options

☒ Use training set
☐ Supplied test set Set...
☐ Cross-validation Folds 10
☐ Percentage split % 66
More options...

(Num) class

Start Stop

Result list (right-click for options)

- 14:22:41 - functions.MultilayerPerceptron
- 14:30:25 - functions.MultilayerPerceptron
- 14:30:54 - functions.MultilayerPerceptron
- 14:31:29 - functions.MultilayerPerceptron
- 14:31:55 - functions.MultilayerPerceptron
- 14:32:31 - functions.MultilayerPerceptron
- 14:32:45 - functions.MultilayerPerceptron
- 14:33:03 - functions.MultilayerPerceptron
- 14:33:19 - functions.MultilayerPerceptron
- 14:33:36 - functions.MultilayerPerceptron
- 14:33:46 - functions.MultilayerPerceptron
- 14:34:16 - functions.MultilayerPerceptron
- 14:34:22 - functions.MultilayerPerceptron
- 14:34:35 - functions.MultilayerPerceptron
- 14:34:41 - functions.MultilayerPerceptron

Classifier output

```
rmse=2.59047126302134404342
Attrib CRIM -0.6736845973511719
Attrib ZN -0.3744976893568132
Attrib INDUS -0.9075369827338612
Attrib CHAS=1 0.22874215904788953
Attrib NOX -0.4978057237527395
Attrib RM -3.9401217682348175
Attrib AGE 0.6457412449450589
Attrib DIS 0.5277314947182936
Attrib RAD -1.2430486919857386
Attrib TAX -1.0191360058543237
Attrib PTRATIO 0.3788949638989887
Attrib B -1.8163576486977866
Attrib LSTAT 3.6085559439721493
Class
Input
Node 0

Time taken to build model: 0.46 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

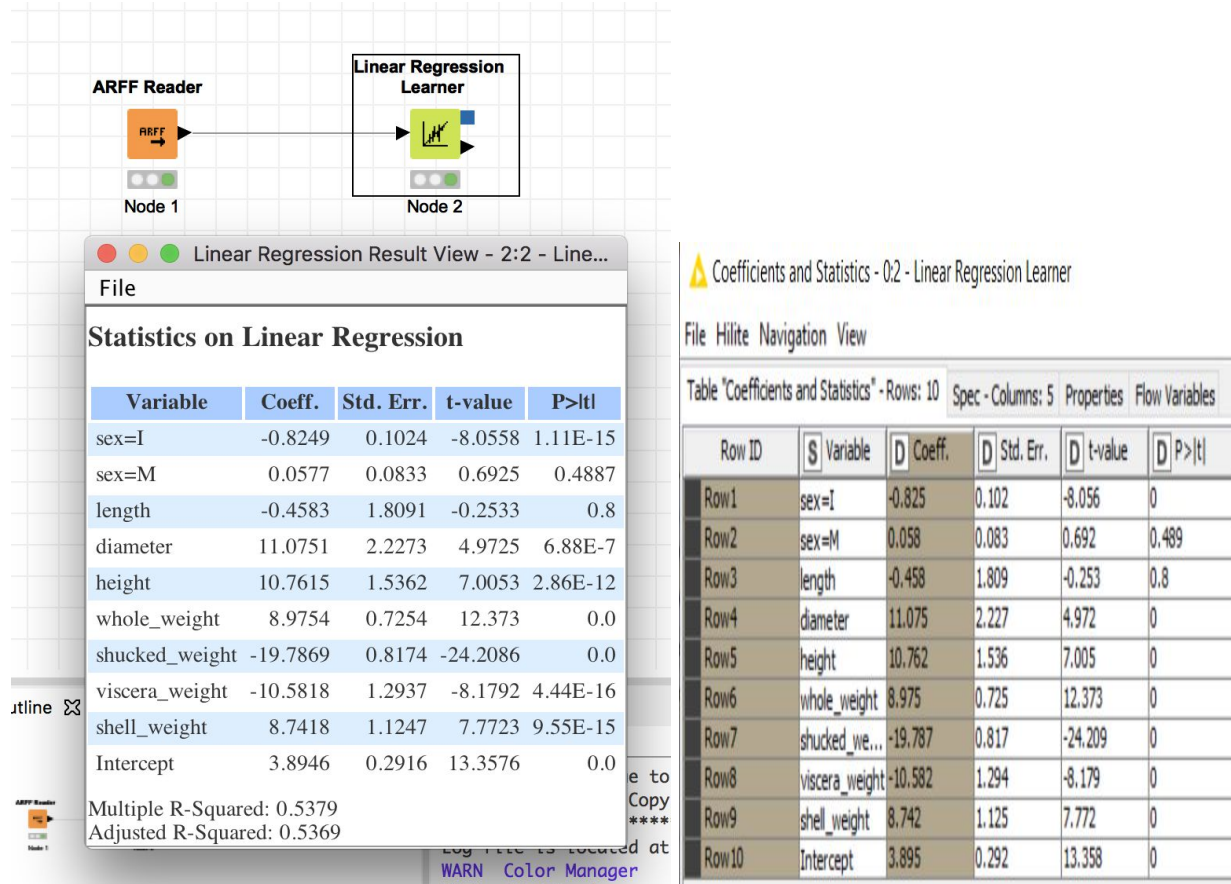
=== Summary ===
Correlation coefficient 0.9595
Mean absolute error 1.979
Root mean squared error 2.5966
Relative absolute error 29.7715 %
Root relative squared error 28.2609 %
Total Number of Instances 506
```

Screenshots = 0.5(NN) + 0.5(RMSE)

- RMSE from 2.4 to 2.6 is acceptable. No need of configs! If RMSE>2.6, deduct 0.25.
- ReadMe: If students get <2.4 and do not mention the exact hyperparameters for that particular value, then -0.25.
- Award points for RMSE values with 3.8654

Q3: (1 point)

A screenshot of the linear regression for the linear equation using the ARFF Reader and Linear Regression Learner nodes.



Linear equation(precision allowable decimals i.e. upto 3 or more decimal places):

$$-0.8249 * (\text{sex}=\text{I}) + 0.0577 * (\text{sex}=\text{M}) - 0.4583 * (\text{length}) + 11.0751 * (\text{diameter}) + 10.7615 * (\text{height}) + 8.9754 * (\text{whole_weight}) - 19.7869 * (\text{shucked_weight}) - 10.5818 * (\text{viscera_weight}) + 8.7418 * (\text{shell_weight}) + 3.8946$$

Screenshot - 0.5

Linear equation in a readme file according to the screenshot obtained- 0.5

Incorrectly written linear equation, -0.25

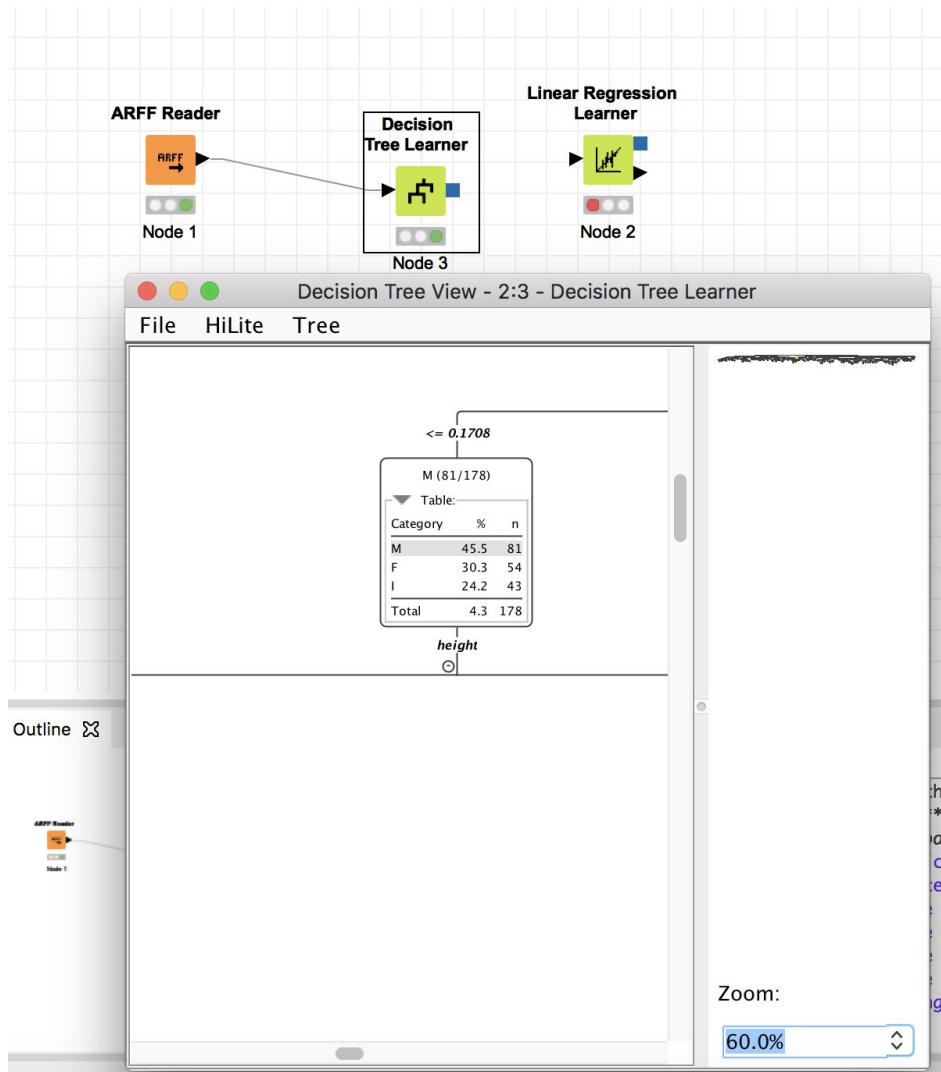
Any unreasonable mismatches which could arise by doing it incorrectly, -0.5

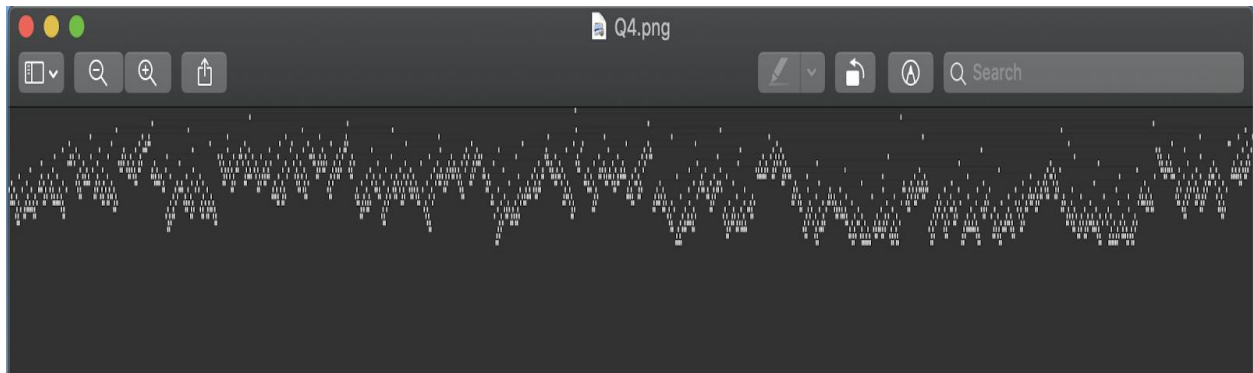
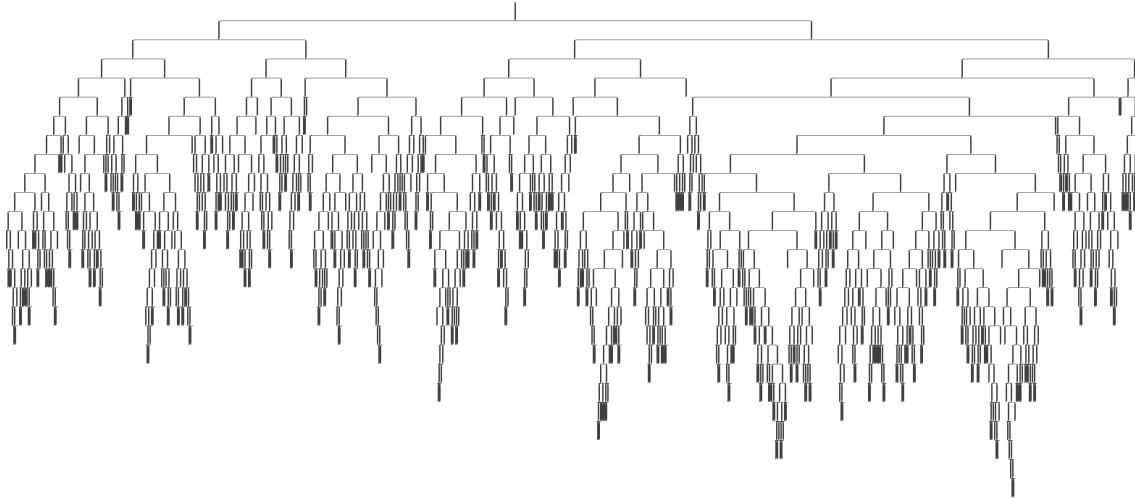
Q4: (1 point)

A screenshot of the decision tree learner.

One of the below three screenshots (or similar screenshots) is fine.

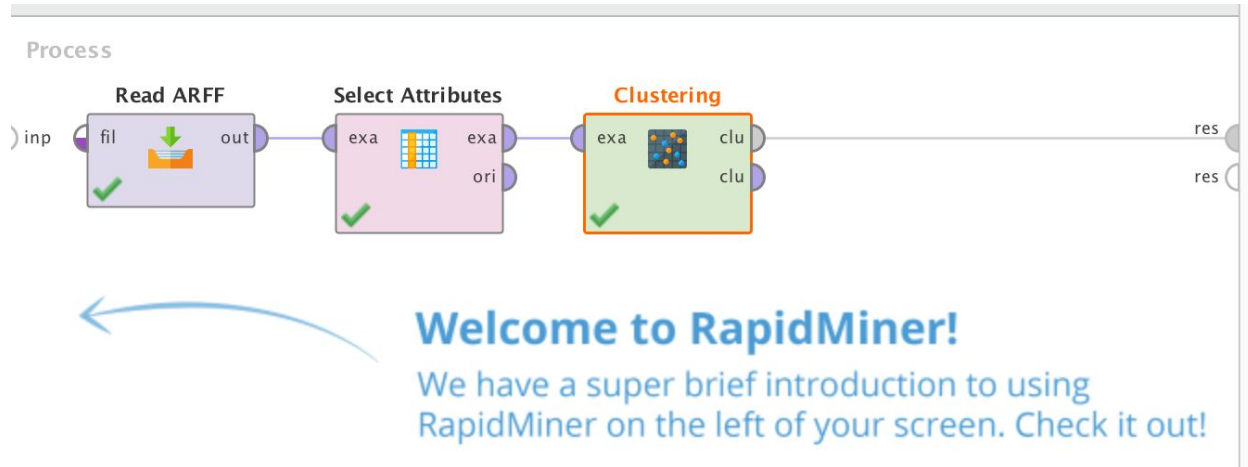
Variants like @1000 and @1051 are acceptable!!





Q5: (1 point)

Below are screenshots working with only length, diameter, height, and num_rings. The number of data points per cluster is provided in the second screenshot.



Screen shots: **Provide one of the below two screenshots.**

(6 clusters) Expected screenshot.

Cluster Model

```
Cluster 0: 1388 items
Cluster 1: 499 items
Cluster 2: 448 items
Cluster 3: 22 items
Cluster 4: 172 items
Cluster 5: 1648 items
Total number of items: 4177
```

Screenshot with clusters' centroids will look like this:

Attribute	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4	cluster_5
length	0.586	0.580	0.321	0.609	0.601	0.501
diameter	0.459	0.458	0.241	0.485	0.475	0.387
height	0.158	0.163	0.079	0.186	0.172	0.129
num_rings	10.736	14.068	5.368	23.682	18.395	8.181

0.5 - screenshot

- All students will have the same number of samples in each cluster if they have made exactly 6 clusters. (as above).

Readme: 0.5 - Number of samples/centroids for each cluster must be mentioned according to the screenshot that they got.


(1388, 499, 448, 22, 172, 1648)

- Make sure there are 6 clusters, else deduct 0.5.

Q6: (0.5 points)

Below are screenshots of the linear regression. The equation is determined by the first screenshot.

Screen shots: **One of the below two screenshots works**

 <new process*> - RapidMiner Studio Educational 9.6.000 @ LAPTOP-O35RR9JI

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Deployments

Result History LinearRegression (Linear Regression) X

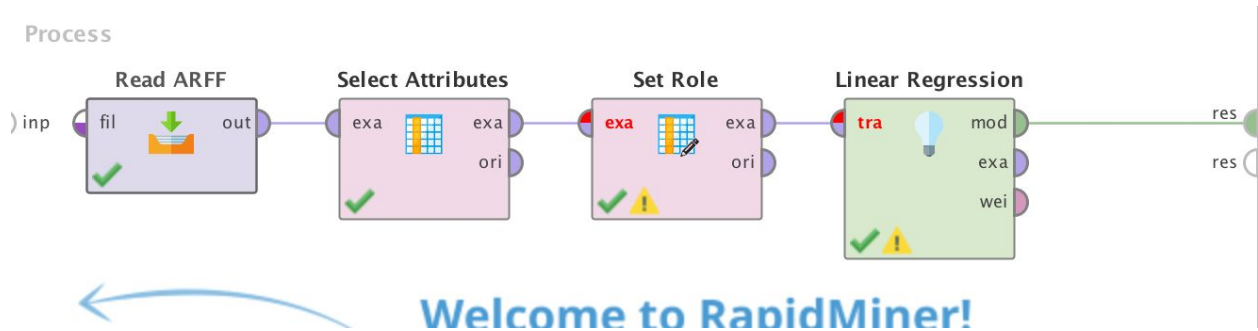
LinearRegression

- 11.933 * length
+ 25.766 * diameter
+ 20.358 * height
+ 2.836

Data
Description
Annotations

Variable	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
length	-11.933	2.064	-0.444	0.078	-5.781	0.000	****
diameter	25.766	2.539	0.793	0.094	10.147	0	****
height	20.358	1.737	0.264	0.319	11.719	0	****
(Intercept)	2.836	0.186	?	?	15.243	0	****

README: - 11.933 * length + 25.766 * diameter + 20.358 * height + 2.836



0.25 - screenshot

0.25 - linear equation in the readme file, written in the right format, as mentioned. Must match with their screenshot.

- Num_rings must be predicted from length, height and diameter. If parameters other than the ones mentioned are used, deduct 0.25. (in this case, the linear equation will be wrong)