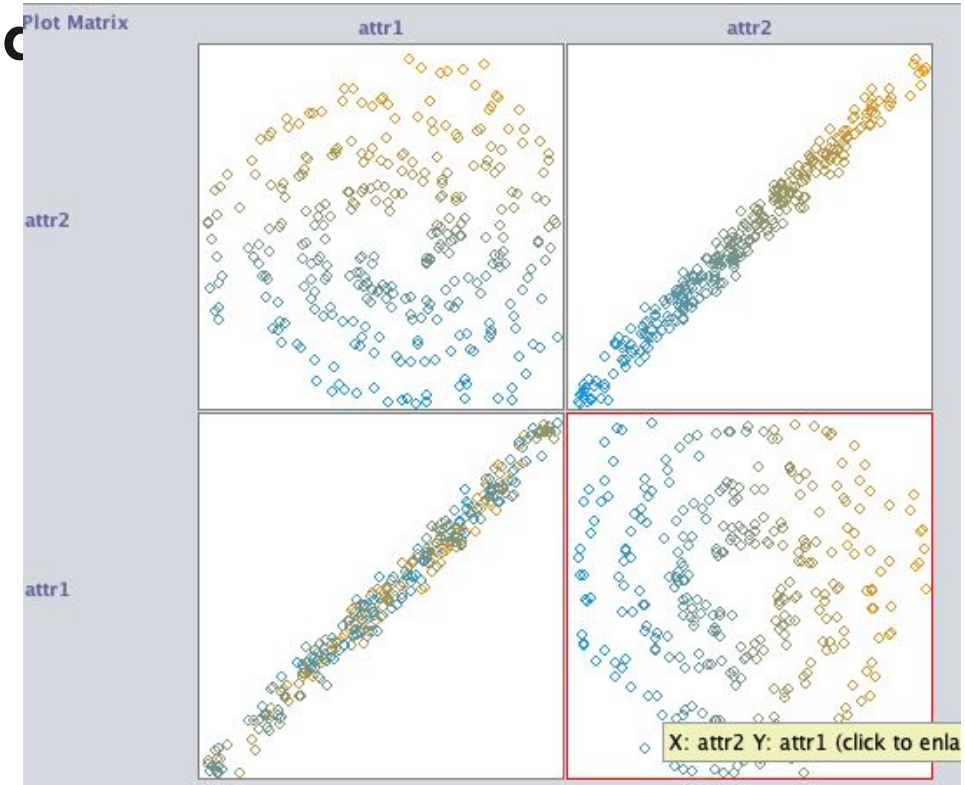
DWDM Assignment 03 Presentation

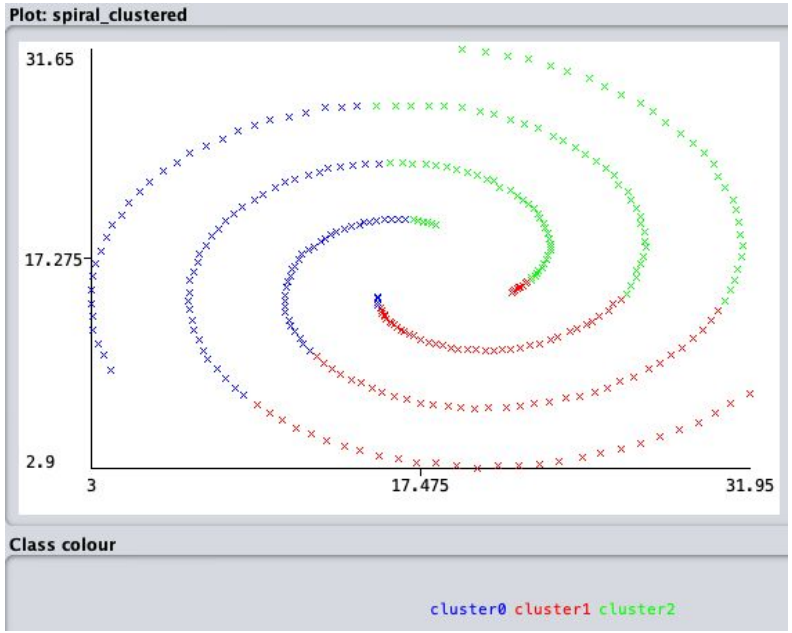
Qinyang Wu, st174540@stud.Uni-Stuttgart.de, 3519174
Huicheng Qian, st169665@stud.uni-stuttgart.de, 3443114
Kuang-Yu Li, st169971@stud.uni-stuttgart.de, 3440829

Task1.3: Explore Clustering Algorithms

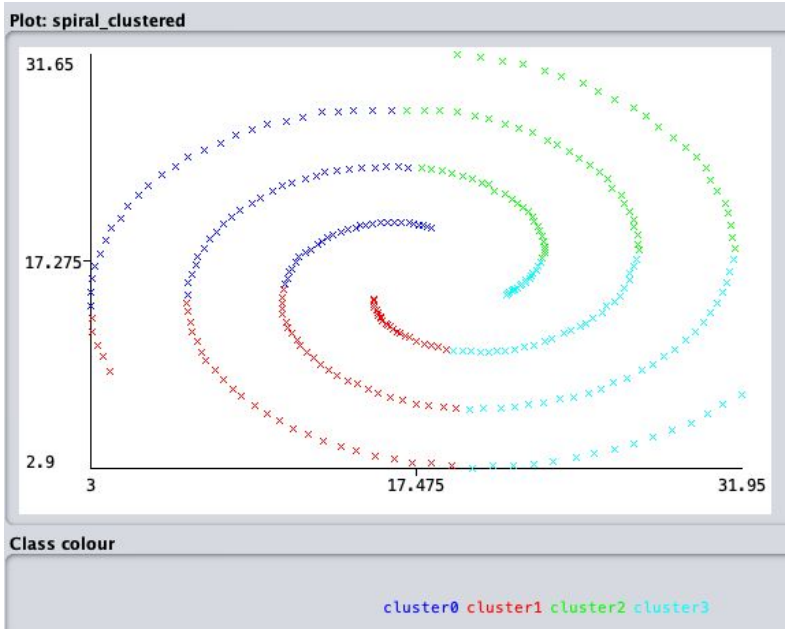
Spiral.csv d


Plot Matrix
attr1
attr2
attr2
attr1
X: attr2 Y: attr1 (click to enla

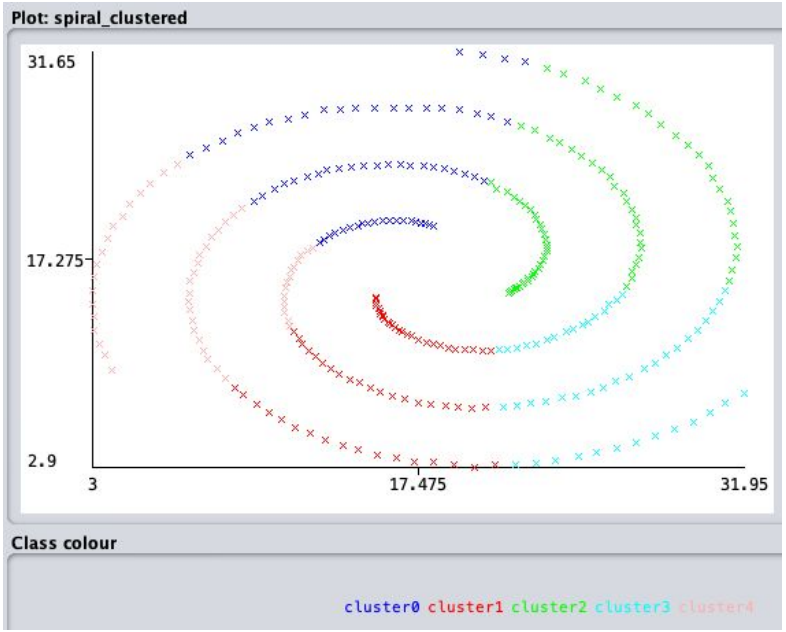
SimpleKMeans: values using k-means++ K = 3



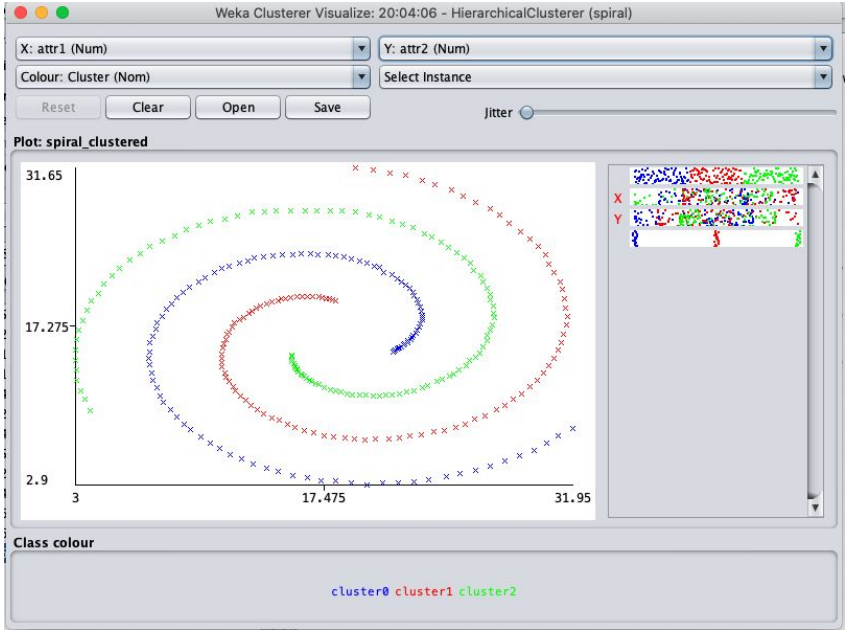
SimpleKMeans: values using k-means++ K = 4



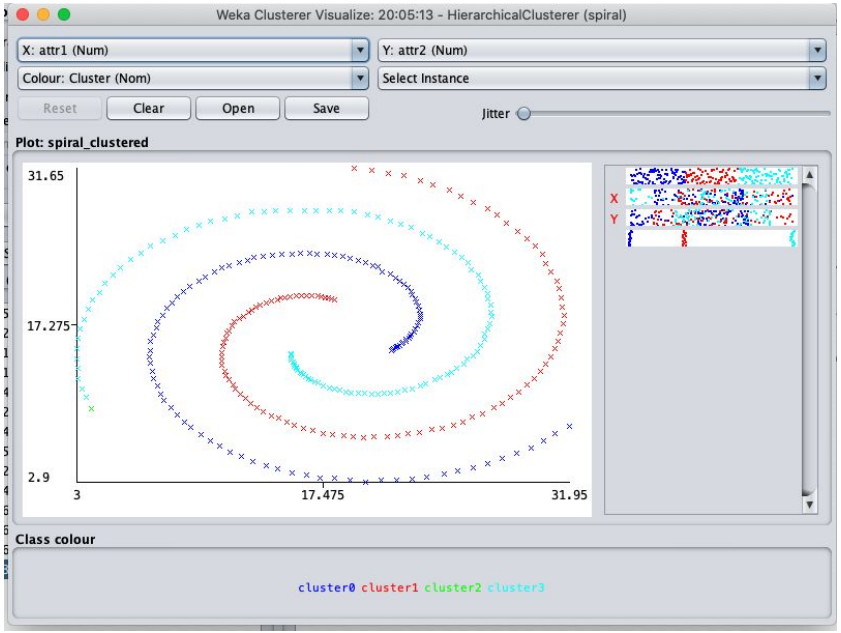
SimpleKMeans: values using k-means++ K = 5



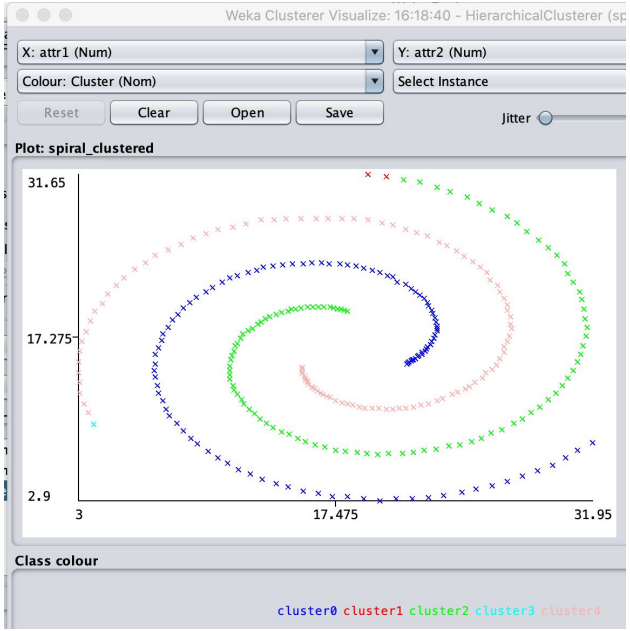
HierachyCluster:



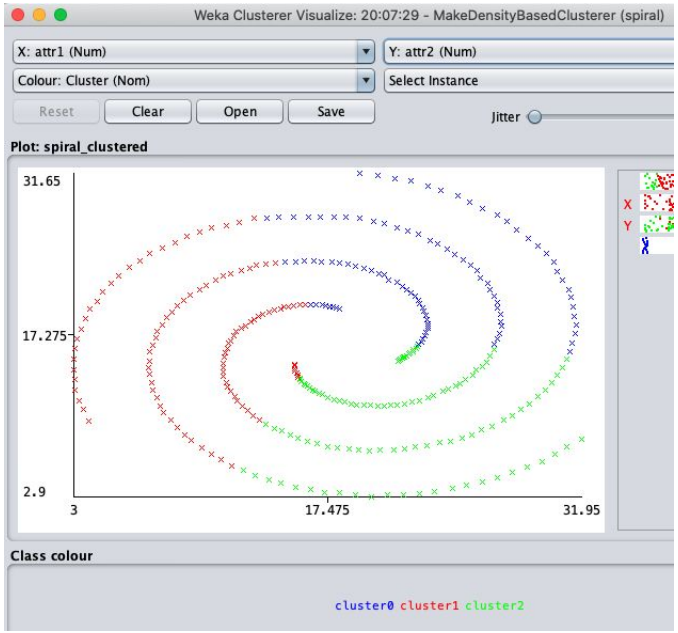
HierachyCluster:



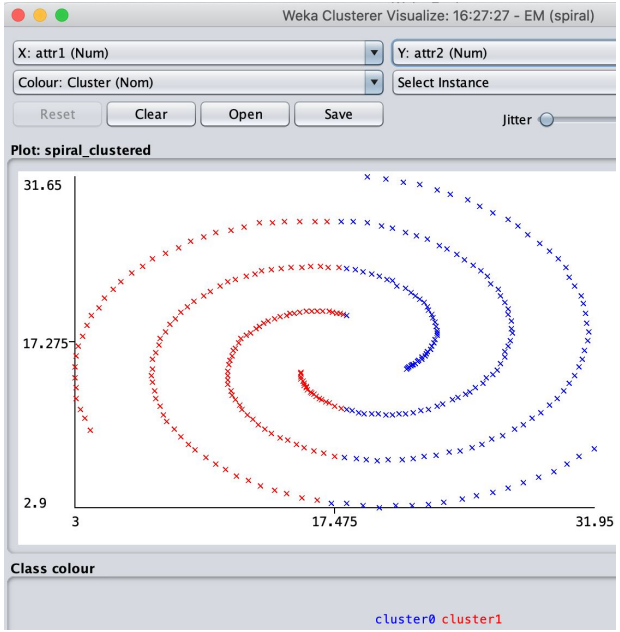
DensityCluster K = 5



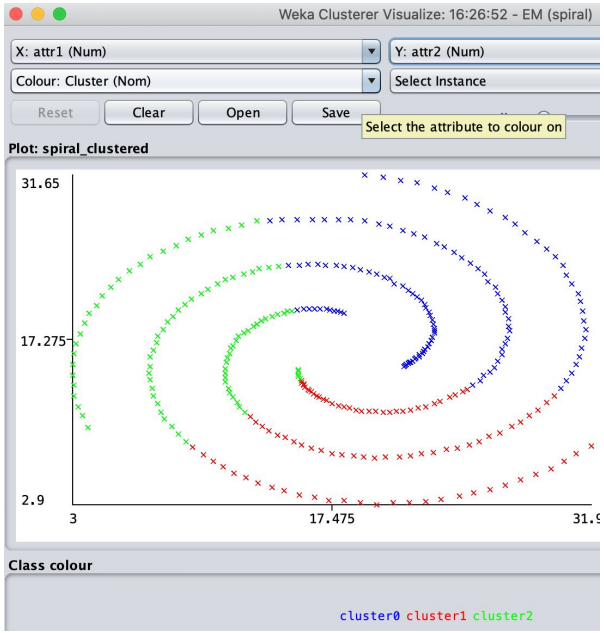
DensityCluster K = 3



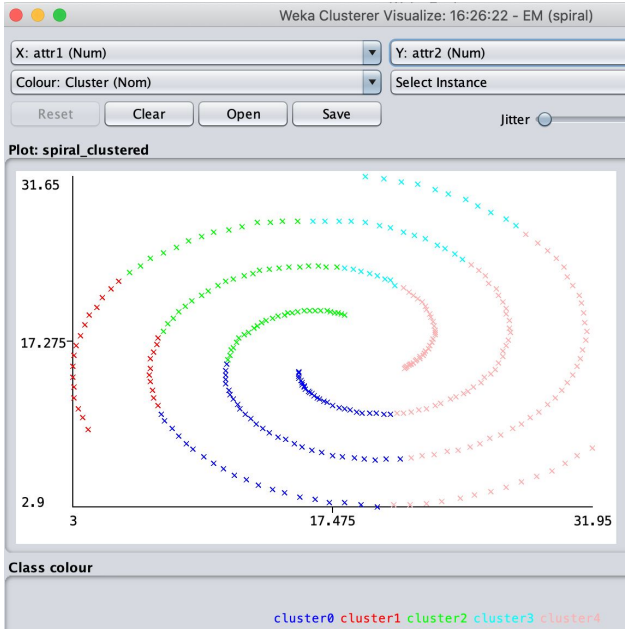
EM


Weka Clusterer Visualize: 16:27:27 - EM (spiral)
X: attr1 (Num)
Y: attr2 (Num)
Colour: Cluster (Nom)
Select Instance
Reset
Clear
Open
Save
Jitter
Plot: spiral_clustered
31.65
17.275
2.9
3
17.475
31.95
Class colour
cluster0 cluster1

EM


Weka Clusterer Visualize: 16:26:52 - EM (spiral)
X: attr1 (Num)
Y: attr2 (Num)
Colour: Cluster (Nom)
Select Instance
Reset
Clear
Open
Save
Select the attribute to colour on
Plot: spiral_clustered
31.65
17.275
2.9
3
17.475
Class colour
cluster0 cluster1 cluster2

EM


Weka Clusterer Visualize: 16:26:22 - EM (spiral)
X: attr1 (Num)
Y: attr2 (Num)
Colour: Cluster (Nom)
Select Instance
Reset
Clear
Open
Save
Jitter
Plot: spiral_clustered
31.65
17.275
2.9
3
17.475
31.95
Class colour
cluster0 cluster1 cluster2 cluster3 cluster4

Running HierachyCluster with Small Guasian takes longer and memory explode

What is your conclusion: Is it a good idea to always run KMeans as clustering algorithm and with the same parameter setting for all datasets? Explain your answer.

Not Suitable to run KMeans on all datasets

Task2.3 Explore different algorithms and parameters

Observation: Apriori to upperBoundMinSupport to **0.1**

Runtime: much longer

Frequency of best rules: every time

Cycles: 18Minimum metric <confidence>: 0.9

Number of cycles performed: 18 vs 11 (prev.)

```
Best rules found:

 1. water-project-cost-sharing=n physician-fee-freeze=y export-administration-act-south-africa=y 44 ==> crime=y 44    <conf:(1)> lift:(1.75) lev:(0.04
 2. water-project-cost-sharing=n physician-fee-freeze=y export-administration-act-south-africa=y 44 ==> Class=republican 44    <conf:(1)> lift:(2.59)
 3. water-project-cost-sharing=y religious-groups-in-schools=n superfund-right-to-sue=n 44 ==> aid-to-nicaraguan-contras=y 44    <conf:(1)> lift:(1.8)
 4. water-project-cost-sharing=y religious-groups-in-schools=n crime=n 44 ==> aid-to-nicaraguan-contras=y 44    <conf:(1)> lift:(1.8) lev:(0.04) [19]
 5. water-project-cost-sharing=y religious-groups-in-schools=n crime=n 44 ==> Class=democrat 44    <conf:(1)> lift:(1.63) lev:(0.04) [16] conv:(16.99)
 6. education-spending=y export-administration-act-south-africa=n Class=republican 44 ==> adoption-of-the-budget-resolution=n 44    <conf:(1)> lift:(2
 7. religious-groups-in-schools=n immigration=y crime=n 44 ==> physician-fee-freeze=n 44    <conf:(1)> lift:(1.76) lev:(0.04) [19] conv:(19.02)
 8. physician-fee-freeze=n religious-groups-in-schools=y superfund-right-to-sue=y 44 ==> Class=democrat 44    <conf:(1)> lift:(1.63) lev:(0.04) [16] c
 9. education-spending=y duty-free-exports=n export-administration-act-south-africa=n 44 ==> physician-fee-freeze=y 44    <conf:(1)> lift:(2.46) lev:(
10. education-spending=y export-administration-act-south-africa=n Class=republican 44 ==> physician-fee-freeze=y 44    <conf:(1)> lift:(2.46) lev:(0.0
```

Observation: FPGrowth upperBoundMinSupport and lowerBoundMinSupport to **0.1.**

Runtime: Much faster

40 rules found

```
FPGrowth found 40 rules (displaying top 10)

 1. [crime=y, immigration=y, el-salvador-aid=y, water-project-cost-sharing=y]: 44 ==> [religious-groups-in-schools=y]: 44   <conf:(1)> lift:(1.6) lev:(0.04) conv:(16.49)
 2. [adoption-of-the-budget-resolution=y, mx-missile=y, water-project-cost-sharing=y, handicapped-infants=y]: 44 ==> [aid-to-nicaraguan-contras=y]: 44   <conf:(1)> lift:(1.8) lev:(0.04) conv:(19.52)
 3. [export-administration-act-south-africa=y, crime=y, immigration=y, education-spending=y, Class=republican]: 44 ==> [physician-fee-freeze=y]: 44   <conf:(1)> lift:(2.46) lev:(0.06) conv:(26.1)
 4. [export-administration-act-south-africa=y, immigration=y, el-salvador-aid=y, education-spending=y, Class=republican]: 44 ==> [physician-fee-freeze=y]: 44   <conf:(1)> lift:(2.46) lev:(0.06) conv:(26.1)
 5. [immigration=y, el-salvador-aid=y, superfund-right-to-sue=y, water-project-cost-sharing=y]: 45 ==> [religious-groups-in-schools=y]: 44   <conf:(0.98)> lift:(1.56) lev:(0.04) conv:(8.43)
 6. [export-administration-act-south-africa=y, immigration=y, education-spending=y, Class=republican]: 45 ==> [crime=y]: 44   <conf:(0.98)> lift:(1.72) lev:(0.04) conv:(9.67)
 7. [export-administration-act-south-africa=y, immigration=y, education-spending=y, Class=republican]: 45 ==> [el-salvador-aid=y]: 44   <conf:(0.98)> lift:(2.01) lev:(0.05) conv:(11.53)
 8. [religious-groups-in-schools=y, export-administration-act-south-africa=y, crime=y, immigration=y, education-spending=y]: 45 ==> [el-salvador-aid=y]: 44   <conf:(0.98)> lift:(2.01) lev:(0.05) conv:(11.53)
 9. [religious-groups-in-schools=y, export-administration-act-south-africa=y, immigration=y, el-salvador-aid=y, education-spending=y]: 45 ==> [crime=y]: 44   <conf:(0.98)> lift:(1.72) lev:(0.04) conv:(9.67)
10. [export-administration-act-south-africa=y, crime=y, immigration=y, physician-fee-freeze=y, education-spending=y]: 45 ==> [el-salvador-aid=y]: 44   <conf:(0.98)> lift:(2.01) lev:(0.05) conv:(11.53)
```

Observation: FPGrowth setting upperBoundMinSupport and lowerBoundMinSupport to **1.0**

No rules found!

Three different settings of metricType and minMetric combinations.

lift / confidnece↑ (higher)

rules↓ (fewer)

run-time ↓ (faster)

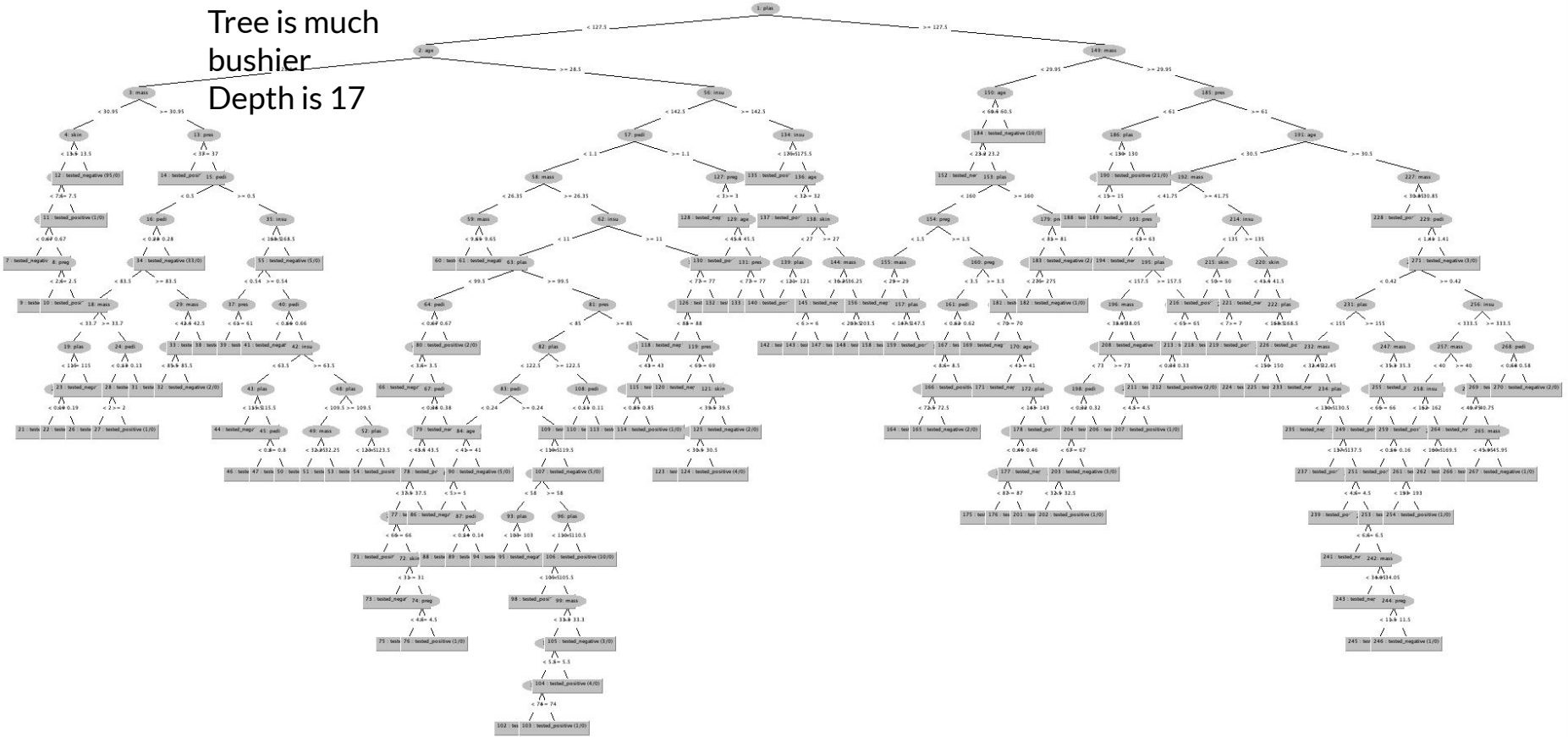
Aprior often crash with confidence < 0.9 and supportMin = Max = 0.1

(Min, Max) = (0.1, 1.0) is too fast to tell difference

Task3.3: Explore different algorithms and parameters

	Precision	Recall	F-Measure
J48 Tree	0.735	0.738	0.736
Random Tree	0.684	0.681	0.682

Tree is much bushier
Depth is 17



	Precision	Recall	F-Measure
J48 Tree	0.735	0.738	0.736
Random Tree D=17	0.684	0.681	0.682
Random Tree D=10	0.702	0.704	0.703
Random Tree D=7	0.723	0.727	0.724
Random Forest	0.754	0.758	0.755
NaiveBayes	0.759	0.763	0.760
Logisctic	0.767	0.772	0.765
AdaBoostM1	0.737	0.743	0.738
IBk (Instance Based Learner)	0.696	0.702	0.698