

Data Mining I: Basic Methods and Techniques
Laboratory Assignment #4:
Due 11/06/2017

John M. Warlop

1 Describe the Classification Rule method.

This type of classifier uses If...then... based rules to classify. This type of classifier uses the covering approach. This method works chooses the attribute-value pair that maximizes the probability of a desired classification. Thus, the goal is maximum accuracy. If t is the total number of instances covered by rule and p is positive examples covered by rule. The classifier chooses a rule that maximizes p/t . This occurs when $p/t == 1$ or set of instances cannot be split anymore.

Data Mining I: Basic Methods and Techniques
Laboratory Assignment #4:
Due 11/06/2017

- 2 Use the Classification rule production method PRISM (weka.classifiers.Prism) on the Weather.nominal data set. How many rules did it produce? Compare this to the Decision tree produced on the same data. What is the difference between the two models?

PRISM & Weather

The screenshot shows the Weka Classifier window with the Prism classifier selected. The 'Test options' section on the left has 'Cross-validation' selected with 10 folds. The 'Classifier output' pane on the right displays the following information:

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

Prism rules

```
-----
If outlook = overcast then yes
If humidity = normal
  and windy = FALSE then yes
If temperature = mild
  and humidity = normal then yes
If outlook = rainy
  and windy = FALSE then yes
If outlook = sunny
  and humidity = high then no
If outlook = rainy
  and windy = TRUE then no
-----
```

Time taken to build model: 0 seconds

==== Stratified cross-validation ====

==== Summary ====

Correctly Classified Instances	9	64.2857 %
Incorrectly Classified Instances	3	21.4286 %
Kappa statistic	0.4375	
Mean absolute error	0.25	
Root mean squared error	0.5	
Relative absolute error	59.2264 %	
Root relative squared error	105.9121 %	
Unclassified Instances	2	14.2857 %
Total Number of Instances	14	

==== Detailed Accuracy By Class ====

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.600	0.700	1.000	0.824	0.529	0.589	0.687	yes
	0.400	0.000	1.000	0.400	0.571	0.529	0.700	0.614	no
Weighted Avg.	0.750	0.350	0.825	0.750	0.718	0.529	0.635	0.657	

J48 & Weather

Data Mining I: Basic Methods and Techniques
Laboratory Assignment #4:
Due 11/06/2017

Classifier

Choose **J48 -C 0.25 -M 2**

Test options

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

(Nom) play

Start Stop

Result list (right-click for options)

16:28:27 - rules.Prism
 16:33:54 - trees.J48

Classifier output

==== Classifier model (from training set) ====

J48 pruned tree

```

outlook = sunny
|  humidity = high: no (3.0)
|  humidity = normal: yes (2.0)
outlook = overcast: yes (4.0)
outlook = rainy
|  windy = TRUE: no (2.0)
|  windy = FALSE: yes (3.0)
  
```

Number of Leaves : 5

Size of the tree : 8

Time taken to build model: 0.01 seconds

==== Stratified cross-validation ====

==== Summary ====

Correctly Classified Instances	7	50	%
Incorrectly Classified Instances	7	50	%
Kappa statistic	-0.0426		
Mean absolute error	0.4167		
Root mean squared error	0.5984		
Relative absolute error	87.5	%	
Root relative squared error	121.2987	%	
Total Number of Instances	14		

==== Detailed Accuracy By Class ====

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.556	0.600	0.625	0.556	0.588	-0.043	0.633	0.758	yes
	0.400	0.444	0.333	0.400	0.364	-0.043	0.633	0.457	no
Weighted Avg.	0.500	0.544	0.521	0.500	0.508	-0.043	0.633	0.650	

64.3% vs 50% performance difference. Neither one very good.

Data Mining I: Basic Methods and Techniques
Laboratory Assignment #4:
Due 11/06/2017

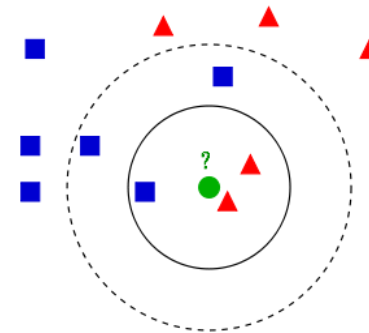
3 Describe the K-nearest neighbor method.



K nearest neighbors method is used in classification and regression. This method assumes the data under consideration does not have known probability distributions.

With this method, the parameters are determined by the data that is used to train the model.

Below is an example of how algorithm works for classification. The red triangles are class "A" and the blue squares are class "B". The green dot(GD) is the test case(which class should GD be in?). Using Euclidean distance, if k is 3 then the circle engulfs 2 "A"'s and 1 "B", but if K is 5, the circle engulfs 3 "B"'s and 2 "A"'s. So, given values of K give different classifications.



Data Mining I: Basic Methods and Techniques
Laboratory Assignment #4:
Due 11/06/2017

- 4 Produce a K-NN model (classifiers.lazy.IBk) for Weather data set. The standard K-nearest neighbor method can be found in the 'lazy' submenu of the list presented when you click 'Choose' in Explorer's Classify window. It is called 'IBk'. Select this and then click on IBk so you can modify the parameters. The default value of k is 1. Set it to 3 (or other value of your preference) and then click Start to run the programs. What is the output? How many instances did it classify correctly and how many incorrectly? As seen below, 9 out of 14 were classified correctly. Whereas 5 out of 14 were classified incorrectly.

Classifier

Choose IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last""

Test options

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

(Nom) play

Start Stop

Result list (right-click for options)

16:45:57 - lazy.IBk

Classifier output

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	9	64.2857 %
Incorrectly Classified Instances	5	35.7143 %
Kappa statistic	0.1026	
Mean absolute error	0.4414	
Root mean squared error	0.4747	
Relative absolute error	92.699 %	
Root relative squared error	96.2242 %	
Total Number of Instances	14	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.889	0.800	0.667	0.889	0.762	0.122	0.689	0.821	yes
	0.200	0.111	0.500	0.200	0.286	0.122	0.644	0.578	no
Weighted Avg.	0.643	0.554	0.607	0.643	0.592	0.122	0.673	0.734	

=== Confusion Matrix ===

a b <-- classified as

8	1	a = yes
4	1	b = no

=> Try changing the parameter K - the number of neighbors. Did that influence the model's performance?

No, I tried several values of K, but it did not change performance. This may be due to the low number of instances.

Data Mining I: Basic Methods and Techniques
Laboratory Assignment #4:
Due 11/06/2017

=> Try using different weighting schemes. Does this change influence the model's performance?
As seen below, when I use the distanceWeight, the performance improves to 71.4%, while leaving KNN @ 3. The weight parameter gives more weight to closer instances.

The screenshot displays the Weka GUI with the 'Classifier' tab selected. The 'Choose' button is set to 'IBK -K 3 -W 0 -I -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last"'. The 'Test options' section shows 'Cross-validation' selected with 'Folds' set to 10. The 'Classifier output' pane shows the following results:

Time taken to build model: 0 seconds
=== Stratified cross-validation ===
=== Summary ===

Metric	Value	Percentage
Correctly Classified Instances	10	71.4286 %
Incorrectly Classified Instances	4	28.5714 %
Kappa statistic	0.3171	
Mean absolute error	0.4454	
Root mean squared error	0.48	
Relative absolute error	93.5264 %	
Root relative squared error	97.2879 %	
Total Number of Instances	14	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MC
0.889	0.600	0.727	0.889	0.800	0.3	
0.400	0.111	0.667	0.400	0.500	0.3	
Weighted Avg.	0.714	0.425	0.706	0.714	0.693	0.3

=== Confusion Matrix ===

```
a b <-- classified as
8 1 | a = yes
3 2 | b = no
```

The 'Result list (right-click for options)' shows a list of instances classified by lazy.IBk. The 'Status' bar at the bottom indicates the classifier is running.

The 'About' dialog box for 'K-nearest neighbours classifier.' is also visible, showing settings for KNN (3), batchSize (100), crossValidate (False), debug (False), distanceWeighting (Weight by 1/distance), doNotCheckCapabilities (False), meanSquared (False), nearestNeighbourSearchAlgorithm (LinearNNSearch -A "weka.core.Euc), numDecimalPlaces (2), and windowSize (0).

Data Mining I: Basic Methods and Techniques
Laboratory Assignment #4:
Due 11/06/2017

- 5 Upload the soybean.arff data set. Before running Weka, it is worth having a brief look at the data file under the Preprocess tab click Edit button. Alternatively, you can take a look at the data file using a text editor (Notepad or WordPad would work). Lines beginning with % are comments. Typically the beginning of the file provides background information on the data set. This includes details of the data itself and references to previous work using the data. The Soybean file contains 683 examples, each of which has 35 attributes plus the class attribute. The task is to assign examples to one of 19 disease classes. Apply the k-nearest neighbor classifier to the soybean data set.

J48 Unpruned & Soybean.arff

Classifier
tour i
into a

Choose J48 -U -M 2

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Nom) class ▼

Start Stop

Result list (right-click for options)

17:13:20 - trees.J48

Classifier output

Correctly Classified Instances 624 91.3616 %

Incorrectly Classified Instances 59 8.6384 %

Kappa statistic 0.9052

Mean absolute error 0.012

Root mean squared error 0.0864

Relative absolute error 12.5318 %

Root relative squared error 39.4263 %


Total Number of Instances 683

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.950	0.002	0.950	0.950	0.950	0.948	0.948	0.974	0.873	diaporthe-stem-canker
0.900	0.000	1.000	0.900	0.900	0.947	0.947	0.949	0.903	charcoal-rot
0.950	0.002	0.950	0.950	0.950	0.948	0.948	0.974	0.951	rhizoctonia-root-rot
0.989	0.005	0.967	0.989	0.978	0.974	0.993	0.961	0.961	phytophthora-rot
0.977	0.000	1.000	0.977	0.989	0.988	0.988	0.988	0.979	brown-stem-rot
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	powdery-mildew
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	downy-mildew
0.935	0.014	0.915	0.935	0.925	0.913	0.983	0.922	0.922	brown-spot
1.000	0.002	0.952	1.000	0.976	0.975	0.999	0.924	0.924	bacterial-blight
0.850	0.000	1.000	0.850	0.919	0.920	0.920	0.854	0.854	bacterial-pustule
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	purple-seed-stain
0.932	0.005	0.932	0.932	0.932	0.927	0.987	0.946	0.946	anthracnose
0.750	0.005	0.833	0.750	0.789	0.785	0.946	0.820	0.820	phyllosticta-leaf-spot
0.868	0.025	0.840	0.868	0.854	0.831	0.958	0.856	0.856	alternaria-leaf-spot
0.813	0.024	0.841	0.813	0.827	0.801	0.954	0.837	0.837	frog-eye-leaf-spot
1.000	0.003	0.882	1.000	0.938	0.938	0.997	0.746	0.746	diaporthe-pod-&-stem-blight
1.000	0.004	0.824	1.000	0.903	0.905	0.996	0.667	0.667	cyst-nematode
0.813	0.004	0.813	0.813	0.813	0.808	0.995	0.745	0.745	2-4-d-injury
0.500	0.003	0.667	0.500	0.571	0.573	0.928	0.699	0.699	herbicide-injury
Weighted Avg.	0.914	0.010	0.914	0.914	0.913	0.904	0.975	0.897	

Data Mining I: Basic Methods and Techniques
Laboratory Assignment #4:
Due 11/06/2017

IBk & Soybeans.arff



Waikato Environment for Intelligence
version 3.8.1
© 1999 - 2016
The University of Waikato
Hamilton, New Zealand

Classifier
Choose **IBk** -K 3 -W 0 -I -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\"""

Test options
☐ Use training set
☐ Supplied test set Set...
☒ Cross-validation Folds **10**
☐ Percentage split % 66
More options...
(Nom) class

Classifier output

```

===== Summary =====
Correctly Classified Instances      625          91.5081 %
Incorrectly Classified Instances    58           8.4919 %
Kappa statistic                    0.9068
Mean absolute error                 0.0126
Root mean squared error             0.0806
Relative absolute error             13.1169 %
Root relative squared error        36.7888 %
Total Number of Instances         683

===== Detailed Accuracy By Class =====

```

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1.000	0.003	0.909	1.000	0.952	0.952	1.000	1.000	diaporthe-stem-canker
1.000	0.002	0.952	1.000	0.976	0.975	1.000	1.000	charcoal-rot
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	rhizoctonia-root-rot
1.000	0.003	0.978	1.000	0.989	0.987	1.000	1.000	phytophthora-rot
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	brown-stem-rot
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	powdery-mildew
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	downy-mildew
0.891	0.020	0.872	0.891	0.882	0.863	0.979	0.950	brown-spot
1.000	0.003	0.909	1.000	0.952	0.952	0.999	0.972	bacterial-blight
0.850	0.000	1.000	0.850	0.919	0.920	0.990	0.957	bacterial-pustule
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	purple-seed-stain
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	anthracnose
0.700	0.008	0.737	0.700	0.718	0.710	0.995	0.846	phyllosticta-leaf-spot
0.879	0.046	0.748	0.879	0.808	0.779	0.982	0.848	alternarialeaf-spot
0.747	0.012	0.907	0.747	0.819	0.799	0.960	0.910	frog-eye-leaf-spot
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	diaporthe-pod-&-stem-blight
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	cyst-nematode
0.688	0.000	1.000	0.688	0.815	0.826	1.000	1.000	2-4-d-injury
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	herbicide-injury
0.915	0.011	0.919	0.915	0.914	0.905	0.989	0.954	

What % of examples are correctly classified? Compare the result to the same result of the unpruned decision tree procedure. Try investigating the effect of repeating the run with different values for k. Compare and contrast the 2 methods and their outputs.

On first inspection, these two classifiers have similar performance. The unpruned tree @ 91.4% and IBk @ 91.6%.

Appendix

Soybeans.arff Header & Meta Info

```
%
% Notes: The large soybean database (soybean-large-data.arff) and it's
%         corresponding test database (soybean-large-test.arff) combined
%         into a single file (soybean-large.arff).
%
% 1. Title: Large Soybean Database
%
% 2. Sources:
%     (a) R.S. Michalski and R.L. Chilausky "Learning by Being Told and
%         Learning from Examples: An Experimental Comparison of the Two
%         Methods of Knowledge Acquisition in the Context of Developing
%         an Expert System for Soybean Disease Diagnosis", International
%         Journal of Policy Analysis and Information Systems, Vol. 4,
%         No. 2, 1980.
%     (b) Donor: Ming Tan & Jeff Schlimmer (Jeff.Schlimmer%cs.cmu.edu)
%     (c) Date: 11 July 1988
%
% 3. Past Usage:
%     1. See above.
%     2. Tan, M., & Eshelman, L. (1988). Using weighted networks to represent
%         classification knowledge in noisy domains. Proceedings of the Fifth
%         International Conference on Machine Learning (pp. 121-134). Ann Arbor,
%         Michigan: Morgan Kaufmann.
%         -- IWN recorded a 97.1% classification accuracy
%         -- 290 training and 340 test instances
%     3. Fisher, D.H. & Schlimmer, J.C. (1988). Concept Simplification and
%         Predictive Accuracy. Proceedings of the Fifth
%         International Conference on Machine Learning (pp. 22-28). Ann Arbor,
%         Michigan: Morgan Kaufmann.
%         -- Notes why this database is highly predictable
%
% 4. Relevant Information Paragraph:
%     There are 19 classes, only the first 15 of which have been used in prior
```

Data Mining I: Basic Methods and Techniques
Laboratory Assignment #4:
Due 11/06/2017

```
% work. The folklore seems to be that the last four classes are
% unjustified by the data since they have so few examples.
% There are 35 categorical attributes, some nominal and some ordered. The
% value ``dna'' means does not apply. The values for attributes are
% encoded numerically, with the first value encoded as ``0,'' the second as
% ``1,'' and so forth. An unknown values is encoded as ``?'.
%
% 5. Number of Instances: 683
%
% 6. Number of Attributes: 35 (all have been nominalized)
%
% 7. Attribute Information:
% -- 19 Classes
% diaporthes-stem-canker, charcoal-rot, rhizoctonia-root-rot,
% phytophthora-rot, brown-stem-rot, powdery-mildew,
% downy-mildew, brown-spot, bacterial-blight,
% bacterial-pustule, purple-seed-stain, anthracnose,
% phyllosticta-leaf-spot, alternarialeaf-spot,
% frog-eye-leaf-spot, diaporthes-pod-&-stem-blight,
% cyst-nematode, 2-4-d-injury, herbicide-injury.
%
% 1. date:      april,may,june,july,august,september,october,?.
% 2. plant-stand: normal,lt-normal,?.
% 3. precip:      lt-norm,norm,gt-norm,?.
% 4. temp:      lt-norm,norm,gt-norm,?.
% 5. hail:      yes,no,?.
% 6. crop-hist: diff-1st-year,same-1st-yr,same-1st-two-yrs,
%               same-1st-sev-yrs,?.
% 7. area-damaged: scattered,low-areas,upper-areas,whole-field,?.
% 8. severity:  minor,pot-severe,severe,?.
% 9. seed-tmt:  none,fungicide,other,?.
% 10. germination: '90-100%','80-89%','lt-80%',?.
% 11. plant-growth: norm,abnorm,?.
% 12. leaves:    norm,abnorm.
% 13. leafspots-halo: absent,yellow-halos,no-yellow-halos,?.
% 14. leafspots-marg: w-s-marg,no-w-s-marg,dna,?.
% 15. leafspot-size: lt-1/8,gt-1/8,dna,?.
% 16. leaf-shread: absent,present,?.
% 17. leaf-malf: absent,present,?.
```

Data Mining I: Basic Methods and Techniques
Laboratory Assignment #4:
Due 11/06/2017

```
% 18. leaf-mild:  absent,upper-surf,lower-surf,?.
% 19. stem:      norm,abnorm,?.
% 20. lodging:   yes,no,?.
% 21. stem-cankers:  absent,below-soil,above-soil,above-sec-nde,?.
% 22. canker-lesion: dna,brown,dk-brown-blk,tan,?.
% 23. fruiting-bodies: absent,present,?.
% 24. external decay: absent,firm-and-dry,watery,?.
% 25. mycelium:  absent,present,?.
% 26. int-discolor: none,brown,black,?.
% 27. sclerotia: absent,present,?.
% 28. fruit-pods: norm,diseased,few-present,dna,?.
% 29. fruit spots:  absent,colored,brown-w/blk-specks,distort,dna,?.
% 30. seed:        norm,abnorm,?.
% 31. mold-growth: absent,present,?.
% 32. seed-discolor: absent,present,?.
% 33. seed-size:  norm,lt-norm,?.
% 34. shriveling: absent,present,?.
% 35. roots:      norm,rotted,galls-cysts,?.
```

@RELATION soybean

```
@ATTRIBUTE date           {april,may,june,july,august,september,october}
@ATTRIBUTE plant-stand    {normal,lt-normal}
@ATTRIBUTE precip         {lt-norm,norm,gt-norm}
@ATTRIBUTE temp           {lt-norm,norm,gt-norm}
@ATTRIBUTE hail           {yes,no}
@ATTRIBUTE crop-hist      {diff-1st-year,same-1st-yr,same-1st-two-yrs, same-1st-sev-yrs}
@ATTRIBUTE area-damaged   {scattered,low-areas,upper-areas,whole-field}
@ATTRIBUTE severity       {minor,pot-severe,severe}
@ATTRIBUTE seed-tmt       {none,fungicide,other}
@ATTRIBUTE germination    {90-100,80-89,lt-80}
@ATTRIBUTE plant-growth   {norm,abnorm}
@ATTRIBUTE leaves         {norm,abnorm}
@ATTRIBUTE leafspots-halo {absent,yellow-halos,no-yellow-halos}
@ATTRIBUTE leafspots-marg {w-s-marg,no-w-s-marg,dna}
@ATTRIBUTE leafspot-size  {lt-1/8,gt-1/8,dna}
@ATTRIBUTE leaf-shread    {absent,present}
@ATTRIBUTE leaf-malf      {absent,present}
@ATTRIBUTE leaf-mild      {absent,upper-surf,lower-surf}
```

Data Mining I: Basic Methods and Techniques
Laboratory Assignment #4:
Due 11/06/2017

```
@ATTRIBUTE stem          {norm,abnorm}
@ATTRIBUTE lodging       {yes,no}
@ATTRIBUTE stem-cankers   {absent,below-soil,above-soil,above-sec-nde}
@ATTRIBUTE canker-lesion  {dna,brown,dk-brown-blk,tan}
@ATTRIBUTE fruiting-bodies {absent,present}
@ATTRIBUTE external-decay {absent,firm-and-dry,watery}
@ATTRIBUTE mycelium       {absent,present}
@ATTRIBUTE int-discolor   {none,brown,black}
@ATTRIBUTE sclerotia      {absent,present}
@ATTRIBUTE fruit-pods      {norm,diseased,few-present,dna}
@ATTRIBUTE fruit-spots     {absent,colored,brown-w/blk-specks,distort,dna}
@ATTRIBUTE seed           {norm,abnorm}
@ATTRIBUTE mold-growth    {absent,present}
@ATTRIBUTE seed-discolor  {absent,present}
@ATTRIBUTE seed-size      {norm,lt-norm}
@ATTRIBUTE shriveling     {absent,present}
@ATTRIBUTE roots          {norm,rotted,galls-cysts}
@ATTRIBUTE class          {diaporthe-stem-canker, charcoal-rot, rhizoctonia-root-rot, phytophthora-rot,
brown-stem-rot, powdery-mildew, downy-mildew, brown-spot, bacterial-blight, bacterial-pustule, purple-seed-stain,
anthracnose, phyllosticta-leaf-spot, alternarialeaf-spot, frog-eye-leaf-spot, diaporthe-pod-&-stem-blight,
cyst-nematode, 2-4-d-injury, herbicide-injury}
```

J48 Unpruned & Soybean.arff

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

```
J48 unpruned tree
```

```
-----
```

```
leafspot-size = lt-1/8
|   canker-lesion = dna
|   |   leafspots-marg = w-s-marg
|   |   |   seed-size = norm: bacterial-blight (21.0/1.0)
```

Data Mining I: Basic Methods and Techniques
Laboratory Assignment #4:
Due 11/06/2017

```

| | | seed-size = lt-norm: bacterial-pustule (3.23/1.23)
| | leafspots-marg = no-w-s-marg
| | | seed-size = norm: bacterial-pustule (12.0)
| | | seed-size = lt-norm
| | | | roots = norm: bacterial-pustule (2.2/0.2)
| | | | roots = rotted: bacterial-pustule (3.3/0.3)
| | | | roots = galls-cysts: cyst-nematode (0.4/0.04)
| | leafspots-marg = dna: bacterial-blight (0.0)
| canker-lesion = brown: bacterial-blight (0.0)
| canker-lesion = dk-brown-blk: phytophthora-rot (4.78/0.1)
| canker-lesion = tan
| | fruit-pods = norm: purple-seed-stain (2.0)
| | fruit-pods = diseased: purple-seed-stain (9.0)
| | fruit-pods = few-present: cyst-nematode (0.23)
| | fruit-pods = dna: purple-seed-stain (0.0)
leafspot-size = gt-1/8
| roots = norm
| | mold-growth = absent
| | | fruit-spots = absent
| | | | leaf-malf = absent
| | | | fruiting-bodies = absent
| | | | | date = april: brown-spot (5.0)
| | | | | date = may: brown-spot (24.0/1.0)
| | | | | date = june
| | | | | precip = lt-norm: phyllosticta-leaf-spot (4.0)
| | | | | precip = norm
| | | | | | area-damaged = scattered: brown-spot (2.0/1.0)
| | | | | | area-damaged = low-areas: brown-spot (0.0)
| | | | | | area-damaged = upper-areas: brown-spot (2.0)
| | | | | | area-damaged = whole-field: phyllosticta-leaf-spot (1.0)
| | | | | precip = gt-norm: brown-spot (21.0)
| | | | | date = july
| | | | | precip = lt-norm: phyllosticta-leaf-spot (1.0)
| | | | | precip = norm: phyllosticta-leaf-spot (2.0)
| | | | | precip = gt-norm
| | | | | | area-damaged = scattered: frog-eye-leaf-spot (6.0/2.0)
| | | | | | area-damaged = low-areas: brown-spot (2.0/1.0)
| | | | | | area-damaged = upper-areas: frog-eye-leaf-spot (2.0)
| | | | | | area-damaged = whole-field: brown-spot (1.0)

```

```
| | | | date = august  
| | | | plant-growth = norm  
| | | | seed = norm  
| | | | hail = yes  
| | | | leaf-shread = absent  
| | | | seed-tmt = none  
| | | | area-damaged = scattered: frog-eye-leaf-spot (3.0/1.0)  
| | | | area-damaged = low-areas: alternarialeaf-spot (1.0)  
| | | | area-damaged = upper-areas: alternarialeaf-spot (5.0/1.0)  
| | | | area-damaged = whole-field: alternarialeaf-spot (3.0)  
| | | | seed-tmt = fungicide  
| | | | plant-stand = normal: frog-eye-leaf-spot (5.0)  
| | | | plant-stand = lt-normal: alternarialeaf-spot (5.0/1.0)  
| | | | seed-tmt = other: alternarialeaf-spot (0.0)  
| | | | leaf-shread = present: alternarialeaf-spot (2.0)  
| | | | hail = no: frog-eye-leaf-spot (3.0)  
| | | | seed = abnorm: alternarialeaf-spot (3.0)  
| | | | plant-growth = abnorm: frog-eye-leaf-spot (2.0)  
| | | | date = september  
| | | | stem = norm  
| | | | temp = lt-norm: alternarialeaf-spot (0.0)  
| | | | temp = norm  
| | | | leaf-shread = absent  
| | | | crop-hist = diff-1st-year: alternarialeaf-spot (3.0)  
| | | | crop-hist = same-1st-yr: frog-eye-leaf-spot (3.0/1.0)  
| | | | crop-hist = same-1st-two-yrs: alternarialeaf-spot (3.0)  
| | | | crop-hist = same-1st-sev-yrs: frog-eye-leaf-spot (2.0)  
| | | | leaf-shread = present: alternarialeaf-spot (5.0)  
| | | | temp = gt-norm: alternarialeaf-spot (28.0)  
| | | | stem = abnorm: frog-eye-leaf-spot (2.0)  
| | | | date = october: alternarialeaf-spot (31.0/1.0)  
| | | | fruiting-bodies = present: brown-spot (34.0)  
| | | | leaf-malf = present: phyllosticta-leaf-spot (10.0)  
| | | | fruit-spots = colored  
| | | | fruit-pods = norm: brown-spot (2.0)  
| | | | fruit-pods = diseased: frog-eye-leaf-spot (62.0)  
| | | | fruit-pods = few-present: frog-eye-leaf-spot (0.0)  
| | | | fruit-pods = dna: frog-eye-leaf-spot (0.0)  
| | | | fruit-spots = brown-w/blk-specks
```

Data Mining I: Basic Methods and Techniques
Laboratory Assignment #4:
Due 11/06/2017

```
| | | | crop-hist = diff-1st-year: brown-spot (0.0)
| | | | crop-hist = same-1st-yr: brown-spot (2.0)
| | | | crop-hist = same-1st-two-yrs: brown-spot (0.0)
| | | | crop-hist = same-1st-sev-yrs: frog-eye-leaf-spot (2.0)
| | | | fruit-spots = distort: brown-spot (0.0)
| | | | fruit-spots = dna: brown-stem-rot (9.0)
| | | mold-growth = present
| | | | leaves = norm: diaporthe-pod-&-stem-blight (7.25)
| | | | leaves = abnorm: downy-mildew (20.0)
| | roots = rotted
| | | area-damaged = scattered: herbicide-injury (1.1/0.1)
| | | area-damaged = low-areas: phytophthora-rot (30.03)
| | | area-damaged = upper-areas: phytophthora-rot (0.0)
| | | area-damaged = whole-field: herbicide-injury (3.66/0.66)
| | roots = galls-cysts
| | | area-damaged = scattered: diaporthe-pod-&-stem-blight (0.02)
| | | area-damaged = low-areas: cyst-nematode (4.37)
| | | area-damaged = upper-areas: cyst-nematode (3.28)
| | | area-damaged = whole-field: diaporthe-pod-&-stem-blight (0.15)
| leafspot-size = dna
| | int-discolor = none
| | | leaves = norm
| | | | plant-growth = norm
| | | | | stem-cankers = absent
| | | | | | canker-lesion = dna: diaporthe-pod-&-stem-blight (5.53)
| | | | | | canker-lesion = brown: purple-seed-stain (0.0)
| | | | | | canker-lesion = dk-brown-blk: purple-seed-stain (0.0)
| | | | | | canker-lesion = tan: purple-seed-stain (9.0)
| | | | | stem-cankers = below-soil: anthracnose (0.0)
| | | | | stem-cankers = above-soil: anthracnose (0.0)
| | | | | stem-cankers = above-sec-nde: anthracnose (24.0)
| | | | plant-growth = abnorm: rhizoctonia-root-rot (19.0)
| | | leaves = abnorm
| | | | stem = norm
| | | | | plant-growth = norm
| | | | | | date = april: 2-4-d-injury (0.38)
| | | | | | date = may: powdery-mildew (3.27/0.27)
| | | | | | date = june: powdery-mildew (3.27/0.27)
| | | | | | date = july: powdery-mildew (2.26/0.26)
```


Data Mining I: Basic Methods and Techniques
Laboratory Assignment #4:
Due 11/06/2017

```
| | | | | date = august: powdery-mildew (4.27/0.27)
| | | | | date = september: powdery-mildew (4.27/0.27)
| | | | | date = october: powdery-mildew (4.27/0.27)
| | | | | plant-growth = abnorm: cyst-nematode (4.3/0.39)
| | | | | stem = abnorm
| | | | | plant-stand = normal
| | | | | leaf-malf = absent
| | | | | | seed = norm: diaporthe-stem-canker (21.0/1.0)
| | | | | | seed = abnorm: anthracnose (9.0)
| | | | | leaf-malf = present: 2-4-d-injury (3.0)
| | | | | plant-stand = lt-normal
| | | | | area-damaged = scattered
| | | | | | leaf-malf = absent: anthracnose (2.0)
| | | | | | leaf-malf = present: herbicide-injury (3.99/1.72)
| | | | | area-damaged = low-areas
| | | | | | fruiting-bodies = absent
| | | | | | | date = april: phytophthora-rot (6.35/0.04)
| | | | | | | date = may: phytophthora-rot (11.94/0.43)
| | | | | | | date = june: phytophthora-rot (14.54/0.45)
| | | | | | | date = july: phytophthora-rot (14.0/0.08)
| | | | | | | date = august: phytophthora-rot (1.92/0.01)
| | | | | | | date = september: 2-4-d-injury (0.37)
| | | | | | | date = october: 2-4-d-injury (0.37)
| | | | | | fruiting-bodies = present
| | | | | | | roots = norm: anthracnose (3.11/0.11)
| | | | | | | roots = rotted: phytophthora-rot (4.7/0.16)
| | | | | | | roots = galls-cysts: phytophthora-rot (0.0)
| | | | | area-damaged = upper-areas: anthracnose (4.71/1.71)
| | | | | area-damaged = whole-field
| | | | | | leaf-malf = absent: anthracnose (4.0/1.0)
| | | | | | leaf-malf = present: 2-4-d-injury (2.05/0.76)
| | | | | int-discolor = brown
| | | | | | leaf-malf = absent: brown-stem-rot (35.73/0.73)
| | | | | | leaf-malf = present: 2-4-d-injury (3.15/0.68)
| | | | | int-discolor = black
| | | | | | area-damaged = scattered: 2-4-d-injury (0.62/0.26)
| | | | | | area-damaged = low-areas: 2-4-d-injury (0.62/0.26)
| | | | | | area-damaged = upper-areas
| | | | | | | date = april: 2-4-d-injury (0.09)
```

Data Mining I: Basic Methods and Techniques
Laboratory Assignment #4:
Due 11/06/2017

```
| | | date = may: charcoal-rot (0.0)
| | | date = june: 2-4-d-injury (0.12/0.03)
| | | date = july: charcoal-rot (3.2/0.2)
| | | date = august: charcoal-rot (1.07/0.07)
| | | date = september: charcoal-rot (5.02/0.02)
| | | date = october: charcoal-rot (1.09/0.09)
| | area-damaged = whole-field
| | | date = april: 2-4-d-injury (0.18/0.09)
| | | date = may: charcoal-rot (0.0)
| | | date = june: charcoal-rot (0.0)
| | | date = july: 2-4-d-injury (0.09)
| | | date = august: charcoal-rot (4.1/0.1)
| | | date = september: charcoal-rot (1.0/0.0)
| | | date = october: charcoal-rot (5.02/0.02)
```

Number of Leaves : 121

Size of the tree : 175

Time taken to build model: 0.02 seconds