FUZZY LOGIC

By Alyssa Dalton

OUTLINE

- Problem Description
- Classification
- Overtraining
- Hard Classification vs. Fuzzy Logic
- Logical Benefit to Fuzzy Logic
- Iris Data Set
- Cluster Data Set
- C4.5 algorithm
- Results

PROBLEM DESCRIPTION

- This analysis evaluates the use of fuzzy logic compared to hard classification methods. This is done by determining the results of the fuzzy logic algorithm when applied to the manipulated cluster data set compared to the results determined by the C4.5 decision tree induction algorithm.

CLASSIFICATION

- Used to predict group membership for data instances
- Example: Predict whether a day will be sunny, rainy, or cloudy

OVERTRAINING

- Overtraining happens when a classification method uses its training data so that each instance is correctly classified.
- Rule set is too specific to training set and will cause error when new data is run through
- Fuzzy logic and pruning are ways to avoid overtraining

HARD CLASSIFICATION VS. FUZZY LOGIC

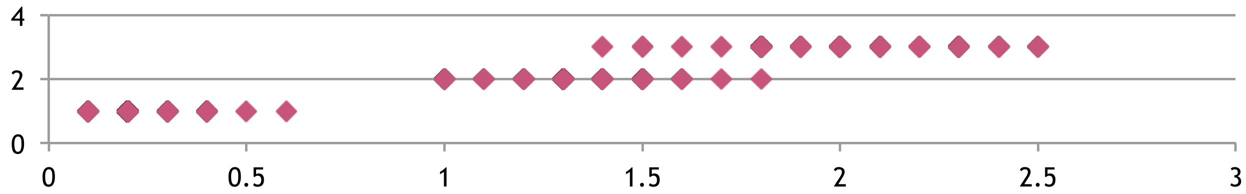
- Hard Classification: Instance belongs to one class only.
- Fuzzy Logic: Instance can belong to more than one class.
- Example: a day can be partly sunny

LOGICAL BENEFIT TO FUZZY LOGIC

- Bank wants to come up with a rule set to determine whether or not someone is a good candidate (low risk) for a loan.
- Someone who is under the age of 25 poses much more risk than someone who is currently 25 or above.
- Hard classification: Until the day that the person turns 25, they are a bad (high risk) candidate.
- Fuzzy Logic: As they approach their birthday, they become a better (lower risk) candidate.

IRIS DATA

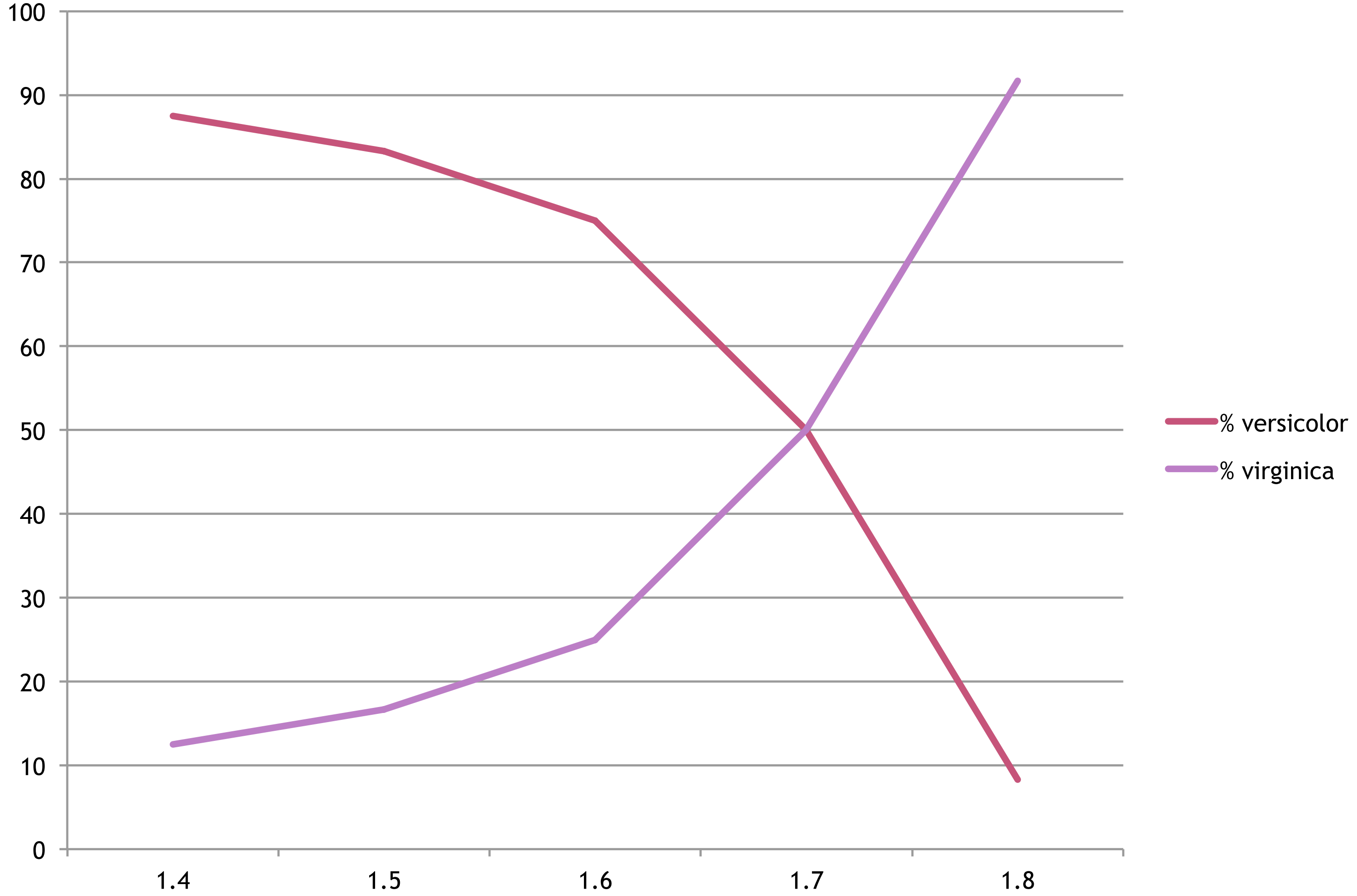
- 150 Instances
- 3 classes: (1) setosa, (2) versicolor, and (3) virginica
- 50 instances per class
- 4 attributes: petal width, petal length, sepal width, and sepal length
- Highest correlation to class: petal width



IRIS PARTIAL MEMBERSHIP

value	No. of instances	Versicolor instances	Virginica instances	% versicolor	% virginica
1.4	8	7	1	87.5	12.5
1.5	12	10	2	83.3	16.7
1.6	4	3	1	75	25
1.7	2	1	1	50	50
1.8	12	1	11	8.3	91.7

IRIS PARTIAL MEMBERSHIP



CLUSTER DATA SET

- Artificial Data Set (Aleshunas, 2011)
  - Advantage: can be manipulated to clearly test performance of fuzzy logic algorithm
- 4 classes
- 500 instances
- 4 attributes, very similar standard deviations
- Class 1 removed
- Only attribute C (highest correlation to class) used
- Divided into Training set and Test set

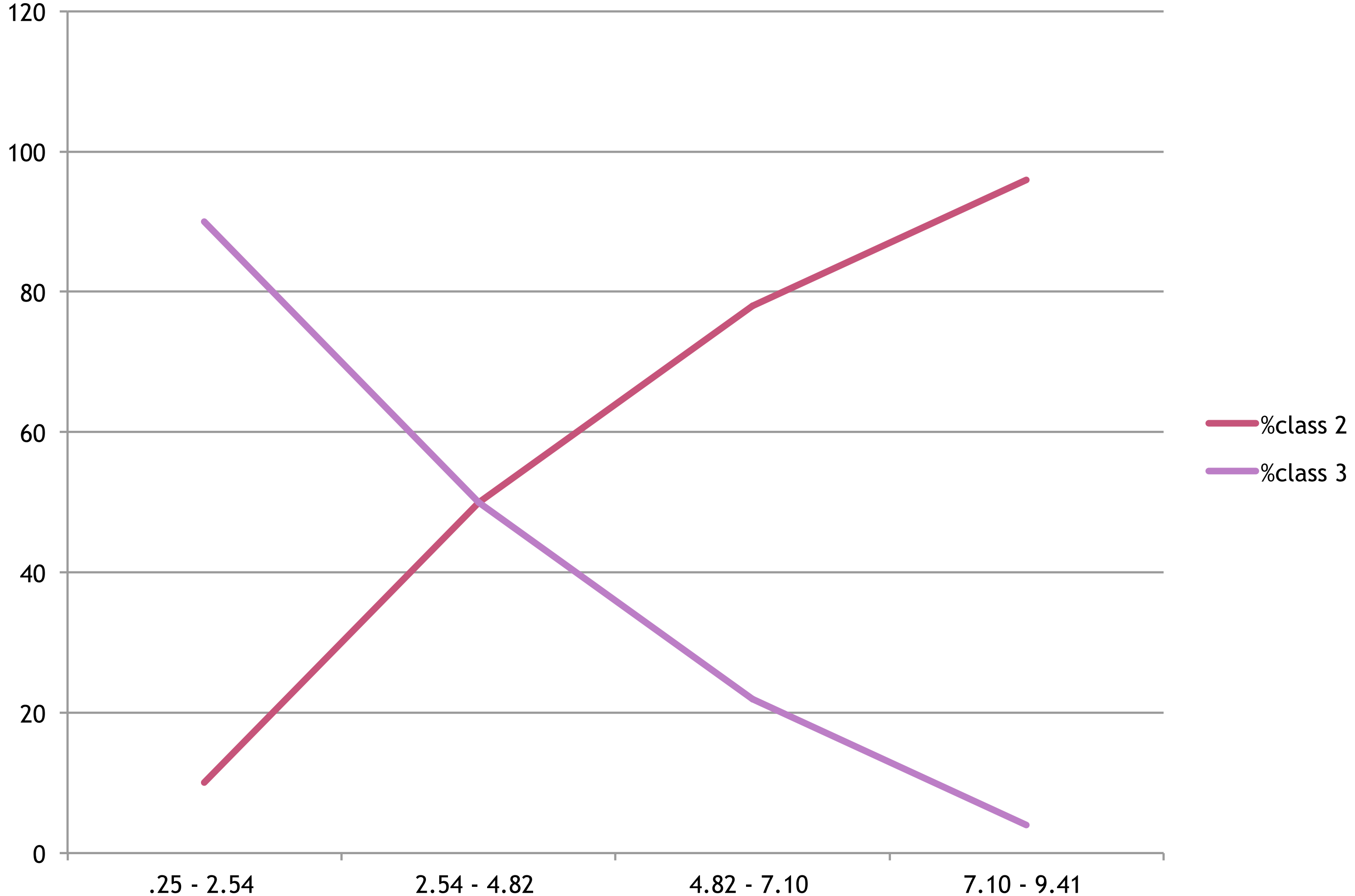
RULES DEVELOPED FOR NON-OVERLAP REGION

1. If $C < -.44$ the instance belongs to Class 4
2. If $-.44 \leq C \leq .25$ the instance belongs to Class 3
3. If $C > 9.41$ the instance belongs to Class 2

CLUSTER SET PARTIAL MEMBERSHIP

[2.29]	No. of instances	class 2	class 3	%class 2	%class 3
.25 - 2.54	29	3	26	10	90
2.54 - 4.82	8	4	4	50	50
4.82 - 7.10	27	21	8	78	22
7.10 - 9.41	55	53	2	96	4

CLUSTER SET PARTIAL MEMBERSHIP



PARTIAL MEMBERSHIP RULES

4. If $.26 \le C \le 2.54$ instance belongs 10% to Class 2 and 90% to Class 3
5. If $2.54 < C < 4.82$ instance belongs 50% to Class 2 and 50% to Class 3
6. If $4.82 \le C \le 7.10$ instance belongs 78% to Class 2 and 22% to Class 3
7. If $7.10 < C < 9.41$ instance belongs 96% to Class 2 and 4% to Class 3

C4.5 RULE SET

Rule 1:
Attribute C <= -17.8466
-> class4 [98.4%]

Rule 7:
Attribute C > 6.28904
-> class2 [95.7%]

Rule 6:
Attribute C > -17.8466
Attribute C <= 6.28904
-> class3 [88.3%]

- Default class: class3

RESULTS

- Both rule sets gave a 5.4% error rate.
- Fuzzy errors cannot be considered strictly errors.
- Instances that were “errors” had partial membership to their actual class.
- Both methods give accurate classification rules.
- Fuzzy logic rules provide more informational detail about the instances that fall into the overlap region.

REFERENCES

- Aleshunas, J. (2011). *Cluster Set*. Retrieved November 30, 2011, from Mercury: http://mercury.webster.edu/aleshunas/Data%20Sets/Supplemental%20Excel%20Data%20Sets.htm
- Chapple, M. (2011). *Classification*. Retrieved November 20, 2011, from About.com: http://databases.about.com/od/datamining/g/classification.htm
- Frank, A. & Asuncion, A. (2010). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science