Bin-Weighted Ensemble Classifiers

Karsten Maurer
Miami University

July 31, 2017

Overview

Classifiers

$$y_i \in C \equiv \{\text{Set of true classes}\}$$

$$x_i \in F \equiv \{\text{Feature space}\}$$

$C_m(.)$ is defined as a classifier if

$$\hat{Y}_{i,m} = C_m(x_i) : F \to C$$

See *The Elements of Statistical Learning* (Hastie et al., 2013) for a general resource on classifier methods

Ensemble Classifiers

Suppose we have set of classifiers $\{C_1(.), C_2(.), ..., C_M(.)\}$

An ensemble classifier is a function such that

$$W(x_i) = G(C_1(x_i), C_2(x_i), ..., C_M(x_i)) : F \rightarrow C$$

A simple majority vote ensemble classifier, can be defined as:

$$\hat{Y}_i = W(x_i) = k \text{ where } \max_k\{\sum_{m=1}^{M} v_{imk}\}$$

$$\text{where } v_{imk} = \mathbf{I}(C_m(x_i) = k)$$

v_{imk} represents the vote for observation i for $C_m(.)$ toward class k

See *Combining Pattern Classifiers* (Kuncheva, 2004) for a comprehensive overview of ensemble methods

Weighted Ensemble Classifiers

We can additionally add weights to the votes from each classifier

$$W(x_i) = k \text{ where } \max_k \{\sum_{m=1}^{M} \omega_m v_{imk}\}$$

v_{iml} = vote for observation i for $C_m(.)$ toward class k

ω_m = weight for $C_m(.)$ [typically based on classification accuracy]

(Kuncheva, 2004)

Weighted Ensemble Classifiers

$$W(x_i) = k \text{ where } \max_k \{\sum_{m=1}^{M} \omega_m v_{imk}\}$$

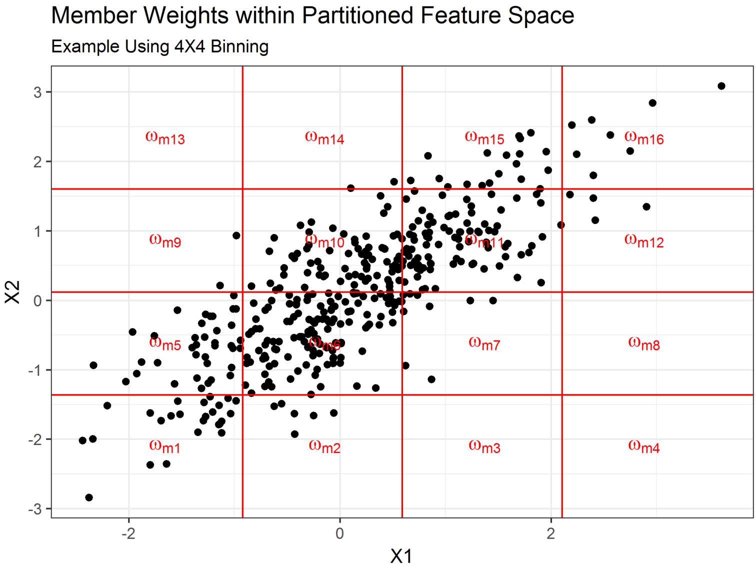
How could we make ω_m depend more specifically on x_i?

Base weights on member accuracy from **similar** training data observations

Related Work:

- Weights in neighborhoods with kNN adaptation (Aha, 1991)
- Weights in cluster based regions (Liu and Yuan, 2001)

Binned Partitions of Feature Space



Bin-Weighted Ensemble Classifiers

Proposed Bin-Weighted Ensemble Classifier:

$$x_i \in \text{Feature Bin } b$$

$$W(x_i) = k \text{ where } \max_k \{\sum_{m=1}^{M} \omega_{mb} v_{imk}\}$$

$$\omega_{mb} = \text{ weight for } C_m \text{ in feature bin } b$$

Application to Example Data

Abalone data collected originally by Nash et al., 1994

- $n = 4177$ Abalone Measurements
- Categorical Response: Sex
- Numeric Features: Physical Attributes
- Accessed via *UCI Machine Learning Repository*

Variable	Description
Sex	Infantile, Adult Female, Adult Male
Length	length on shell (mm)
Diameter	size perpendicular to length (mm)
Height	height laying flat (mm)
Whole Weight	weight of abalone (g)
Shuck Weight	weight of meat (g)
Viscera Weight	weight of gut (g)
Shell Weight	weight of shell (g)
Rings	rings visible on shell (count)

image source: wikipedia.org

Example of Fitting Bin-Weighted Ensemble

Need fitted values for bin weights (ω_{mb}) and votes (v_{imk})

$$W(x_i) = k \text{ where } \max_k\{\sum_{m=1}^{M} \omega_{mb} v_{imk}\}$$

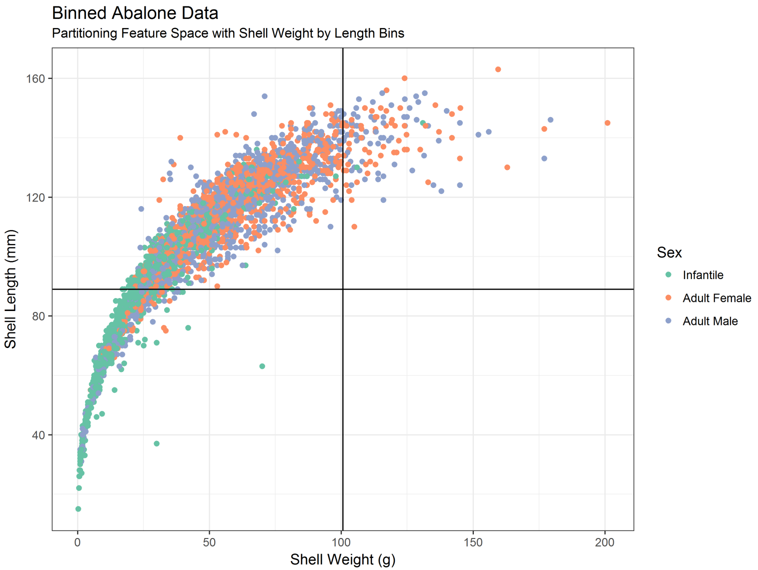
v_{imk} generated based on predictions from fitted member classifiers $\hat{C}_m(.)$

ω_{mb} calculated based on:
- Partitioned variables and numbers of bins
- Accuracy of cross-validated predictions from $\hat{C}_m(.)$ within bin partitions

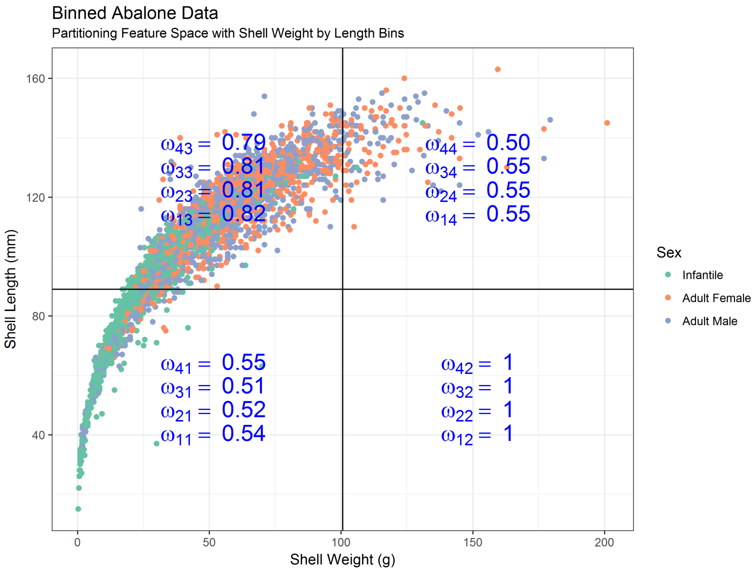
Example of Abalone Bin-Weighted Ensemble Specification:

- $\hat{C}_1(.)$=Naive Bayes, $\hat{C}_2(.)$=Random Forest, $\hat{C}_3(.)$=Bagging, $\hat{C}_4(.)$=Support Vector Machine
- Partition based on two shell weight and two shell length bins

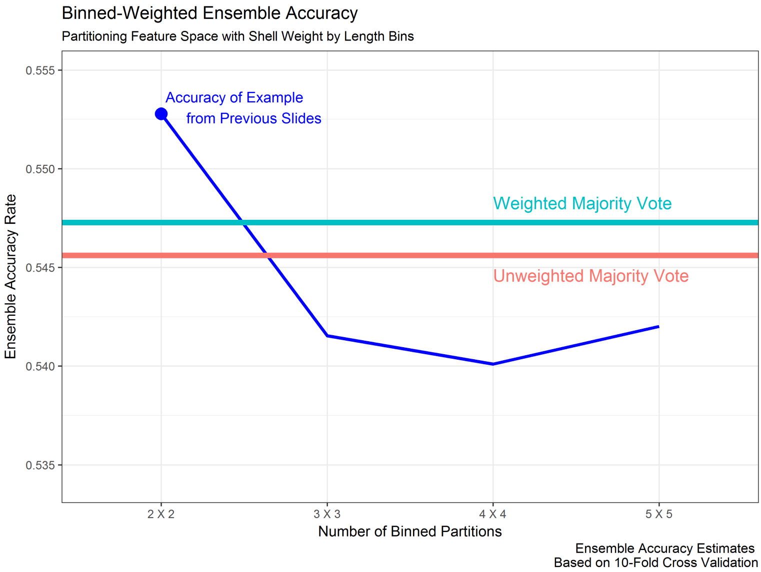
Example of Fitting Bin-Weighted Ensemble



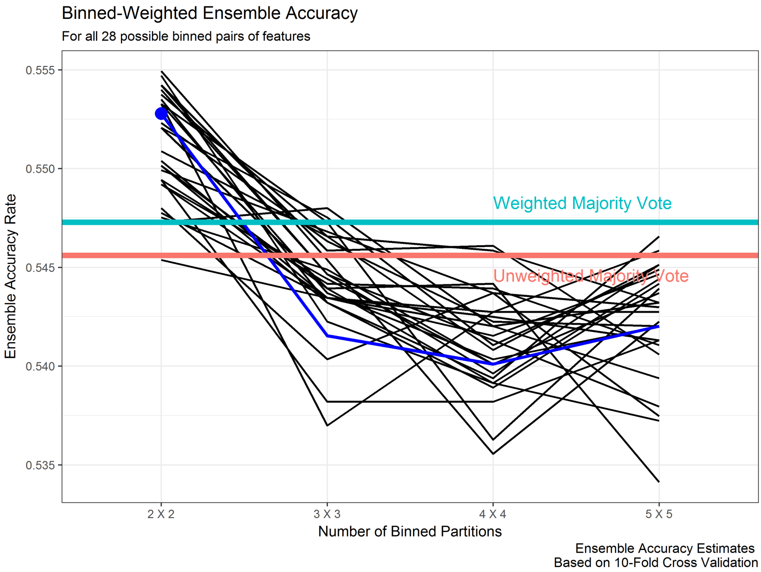
Example of Fitting Bin-Weighted Ensemble



Results


Binned-Weighted Ensemble Accuracy
Partitioning Feature Space with Shell Weight by Length Bins
Ensemble Accuracy Rate
0.555
0.550
0.545
0.540
0.535
Accuracy of Example
from Previous Slides
Weighted Majority Vote
Unweighted Majority Vote
2 X 2
3 X 3
4 X 4
5 X 5
Number of Binned Partitions
Ensemble Accuracy Estimates
Based on 10-Fold Cross Validation

Results


Binned-Weighted Ensemble Accuracy
For all 28 possible binned pairs of features
Weighted Majority Vote
Unweighted Majority Vote
Ensemble Accuracy Rate
0.555
0.550
0.545
0.540
0.535
2 X 2
3 X 3
4 X 4
5 X 5
Number of Binned Partitions
Ensemble Accuracy Estimates
Based on 10-Fold Cross Validation

Conclusions

1. Bin-weighted ensembles showed improvement over simple and weighted majority vote ensembles in Abalone Data

2. Number of bins is a tuning parameter for bin-weighted ensembles

- Can "overfit" with too many bins
- Simple binning scheme can leave many empty/sparse partitions

3. Features selected for partitioning have impact on accuracy

- Creates variable selection problem in specification
- Most Abalone pairs provided improvements with 2 X 2 binning

Current and Future Work

Current work (omitted for time/simplicity):

- Non-uniform binning sizes to balanced sample sizes
- Partial voting with posterior probabilities
- Winner take all bin-voting

Future work:

- Asymptotic voting properties
- Comparing computation time and accuracy with other “local competence” weighted ensembles (Kuncheva, 2004)

Acknowledgements

Collaborators:

- Dr. Walter Bennette : Industrial Engineer, Air Force Research Lab
- Yuexi Wang: Research Assistant, Miami University
- Amy Underwood: Research Assistant, Miami University

Sounding Board:

- Dr. John Bailer : Department Chair of Statistics, Miami University
- Dr. Tom Fisher : Associate Professor, Miami University

Financial Support of Research:

- Summer Research Appointment through Miami University Committee for Faculty Research

Bibliography

Aggarwal, C., & Zhai, C. (2012). *A Survey of Text Classification Algorithms*. Mining Text Data, 163-222. `http://doi.org/10.1007/978-1-4614-3223-4_6`

Aha, D. W., Kibler, D., & Albert, M. K. (1991). *Instance-Based Learning Algorithms*. Machine Learning, 6(1), 37-66. Retrieved from `http://link.springer.com.proxy.lib.iastate.edu/article/10.1007/BF00153759`

Alcalá-Fdez, J., Fernandez, A., Luengo, J., Derrac, J., García, S., Sánchez, L., & Herrera, F. (2011) *KEEL Data-Mining Software Tool: Data Set Repository, Integration of Algorithms and Experimental Analysis Framework*. Journal of Multiple-Valued Logic and Soft Computing 17:2-3, 255-287.

Breiman, L. (University of C. (1999). *Random forest*. Machine Learning, 45(5), 1-35.

Cover, T., & Hart, P. (1967). *Nearest neighbor pattern classification*. IEEE Transactions on Information Theory, 13(1), 21-27. `http://doi.org/10.1109/TIT.1967.1053964`

Frank, E., Hall, M., & Witten, I. (2016). *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"*, Morgan Kaufmann, Fourth Edition, 2016.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Elements, 1, 337-387.

Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques (Second Edi)*. Diane Cerra.

Hornik K, Buchta C and Zeileis A (2009). *Open-Source Machine Learning: R Meets Weka.* Computational Statistics, 24(2), pp. 225-232. `http://doi.org/10.1007/s00180-008-0119-7`).

Bibliography (continued)

Kuncheva, L. (2004) *Combining Pattern Classifiers: Methods and Algorithms*. John Wiley & Sons, Inc.

R Core Team (2016). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. `https://www.R-project.org/`.

R. Liu & B. Yuan. (2001) *Multiple classifier combination by clustering and selection*. Information Fusion, 2:163-168. `https://link.springer.com/article/10.1007%2FBF00153759`

Rish, I. (2001). *An empirical study of the naive Bayes classifier. IJCAI 2001 Workshop on Empirical Methods in Artificial* ., 0000(0000), 41-46. `http://doi.org/10.1039/b104835j`

Scholkopf, B, & Smola, A. (2001) *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.

Sebban, M., Nock, R., Chauchat, J. H., & Rakotomalala, R. (2000). *Impact of learning set quality and size on decision tree performances. International Journal of Computers, Systems and Signals*, 1(1), 85-105. Retrieved from
`http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.62.6365`

Quinlan, J. R. (1992). C4.5: *programs for machine learning*.

Witten, I. H., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Zhang, H. (2004). *The optimality of naive Bayes*. AA, 1(2), 3. Retrieved from
`http://www.aaai.org/Papers/FLAIRS/2004/Flairs04-097.pdf`

Thank You

Thank you!

Questions?

`maurerkt@miamioh.edu`