

A Comparison of Multiclass SVM Methods

Ben Aisen

December 15, 2006

1 Introduction

The support vector machine is a powerful tool for binary classification, capable of generating very fast classifier functions following a training period. There are several approaches to adopting SVMs to classification problems with three or more classes:

- **Multiclass ranking SVMs**, in which one SVM decision function attempts to classify all classes.
- **One-against-all classification**, in which there is one binary SVM for each class to separate members of that class from members of other classes.
- **Pairwise classification**, in which there is one binary SVM for each pair of classes to separate members of one class from members of the other.

These three methods have been tested on a data set from Jeff Orkin's project, *Speech Act Classification of Video Game Data*. This data has a large, sparsely populated feature space and classes that are defined more qualitatively than mathematically. The single multiclass SVM fared poorly on this data set, while the one-against-all and pairwise methods both performed reasonably well.

2 Notes on the data set and methods

The data was extracted from restaurant game transcripts using Jeff Orkin's FeatureCreep program, set to use the presence or absence of the top 500 words, bigrams, and trigrams from each of six classes:

1. Assertion
2. (*)
3. Directive
4. Expressive
5. Greeting
6. Promise
7. Question

There is no class 2 because that slot was taken by the *Continuation* class, which has been omitted from ranking; for details, see Orkin's report.

The data was classified with the SVMlight engine by Thorsten Joachims; I have written a set of Python scripts to automate the operation of SVMlight and summarize its results. In all cases below, 10-fold cross-validation was performed and the results of each fold added to produce final results on the data set.

3 Multiclass ranking SVM

Multiclass ranking SVMs are generally considered unreliable for arbitrary data, since in many cases no single mathematical function exists to separate all classes of data from one another. This data set proved to be no exception, as SVMlight ran for over two hours in training mode before aborting due to a lack of progress and producing garbage results. As the engine failed to optimize a function for the problem, no results are presented.

4 One-Against-All Methods

One-against-all (OAA) SVMs were first introduced by Vladimir Vapnik in 1995. [4] The initial formulation of the one-against-all method required unanimity among all SVMs: a data point would be classified under a certain class if and only if that class's SVM accepted it and all other classes' SVMs rejected it. While accurate for tightly clustered classes, this method leaves regions of the feature space undecided where more than one class accepts or all classes reject. In the case of this data set, about 25% were unaccounted for.

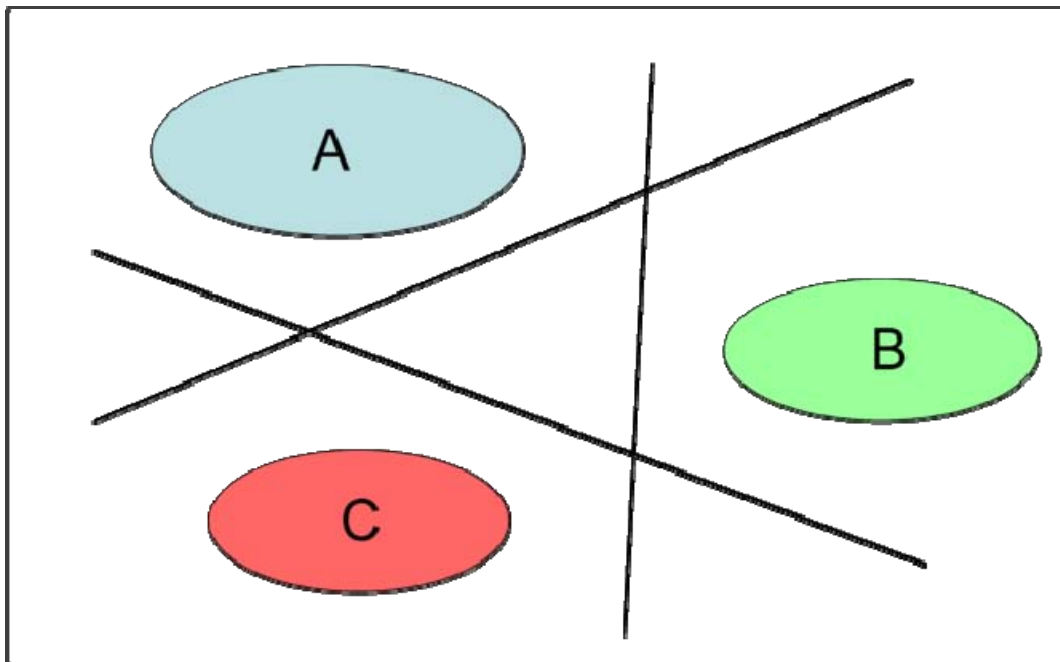


Figure 1: Diagram of binary OAA region boundaries on a basic problem

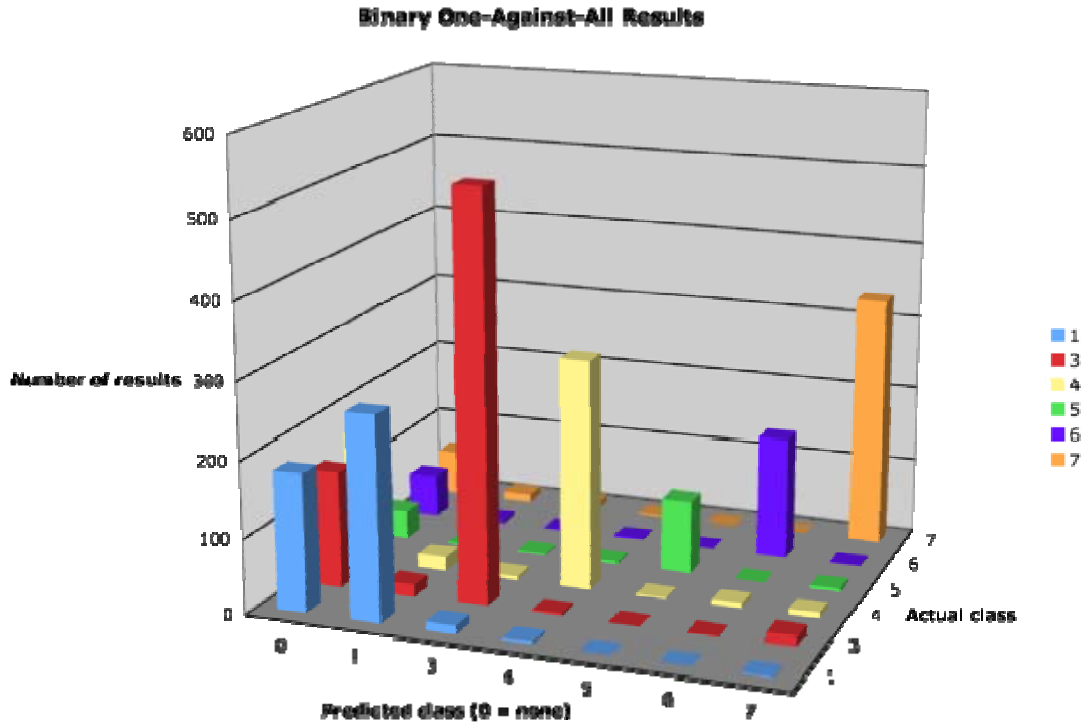


Figure 2: Graph of binary OAA results on data set

Table 1: Confusion matrix for binary OAA

	Estimated class							
Actual	0	1	3	4	5	6	7	
1	182	266	11	4	0	0	1	
3	153	16	532	3	0	1	15	
4	172	19	5	300	2	7	7	
5	38	2	2	3	97	0	4	
6	56	5	2	2	0	161	1	
7	60	10	9	2	0	0	332	

One method for improving the performance of OAA SVMs was suggested by Vapnik in 1998. [5] It involves using the continuous values of SVM decision functions rather than simply their signs. The class of a data point is whichever class has a decision function with highest value, regardless of sign. This appears to be the most common method for multiclass SVM classification in use today.

On this data set, the continuous OAA method had a performance of approximately 81% correctly classified. About half of the datapoints unclassified by binary OAA were correctly classified by continuous OAA.

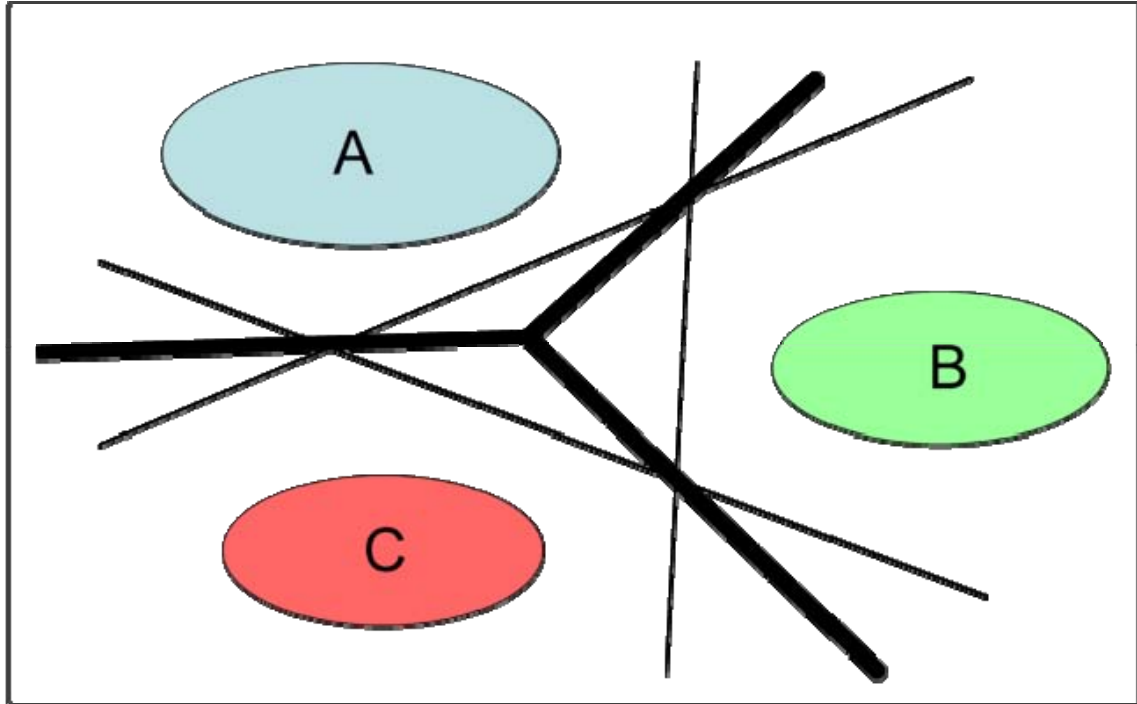


Figure 3: Diagram of continuous OAA region boundaries on a basic problem

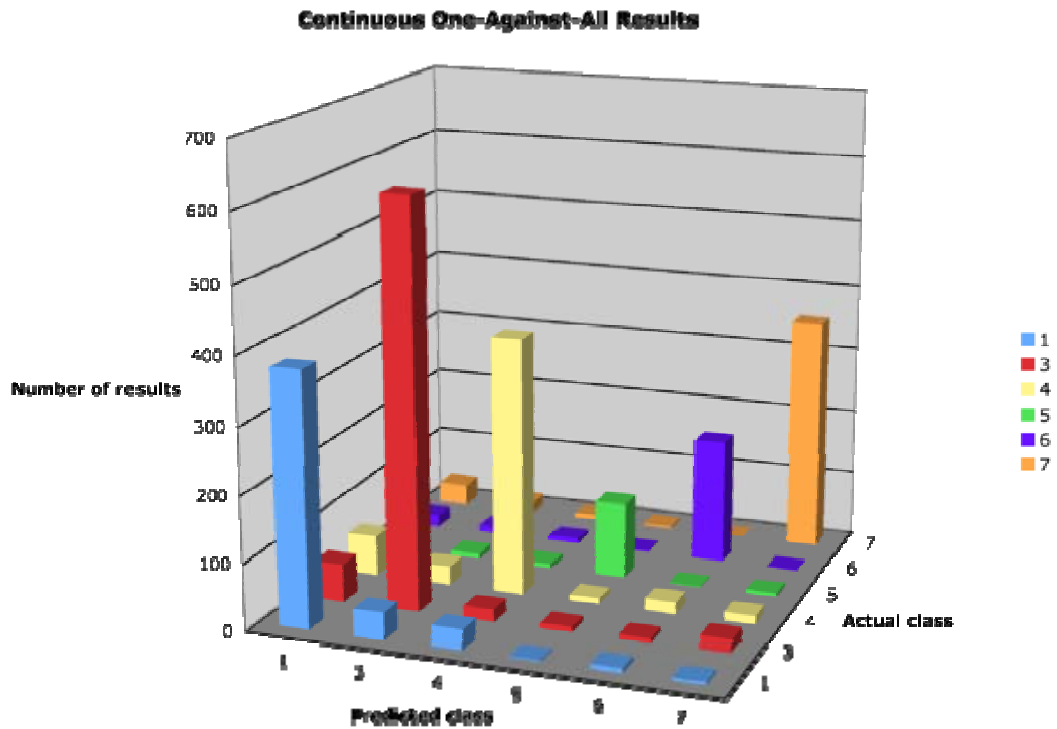


Figure 4: Graph of continuous OAA results on data set

Table 2: Confusion matrix for continuous OAA

Actual \ Estimated class	1	3	4	5	6	7
1	378	41	32	2	7	4
3	57	607	19	8	7	22
4	60	28	384	9	18	13
5	11	8	7	114	1	5
6	18	11	7	2	188	1
7	33	16	7	3	0	354

5 Pairwise Methods

In pairwise SVMs, there is one SVM for each pair of classes trained to separate the data from each. The simplest form of classification with pairwise SVMs selects the class chosen by the maximal number of pairwise SVMs; more advanced methods include using decision graphs to determine the class selected in a similar manner to knockout tournaments. [3] I have used the latter method on the data set, and retrieved results comparable to the continuous OAA SVM.

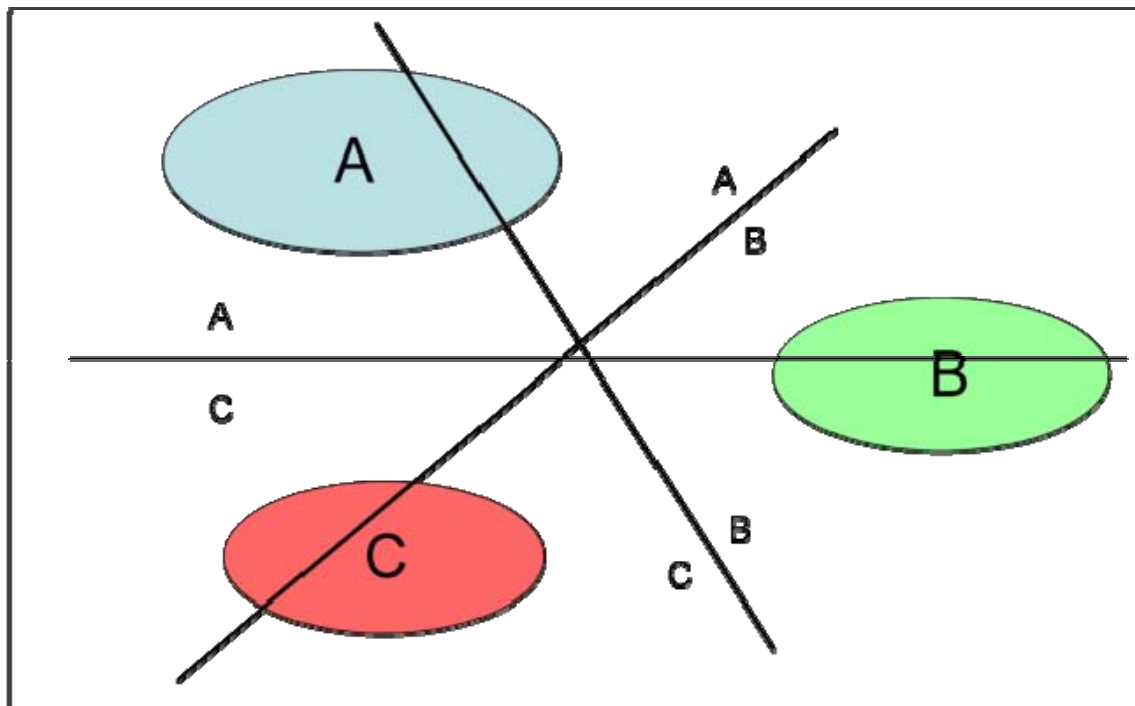


Figure 5: Diagram of pairwise SVM decision boundaries on a basic problem

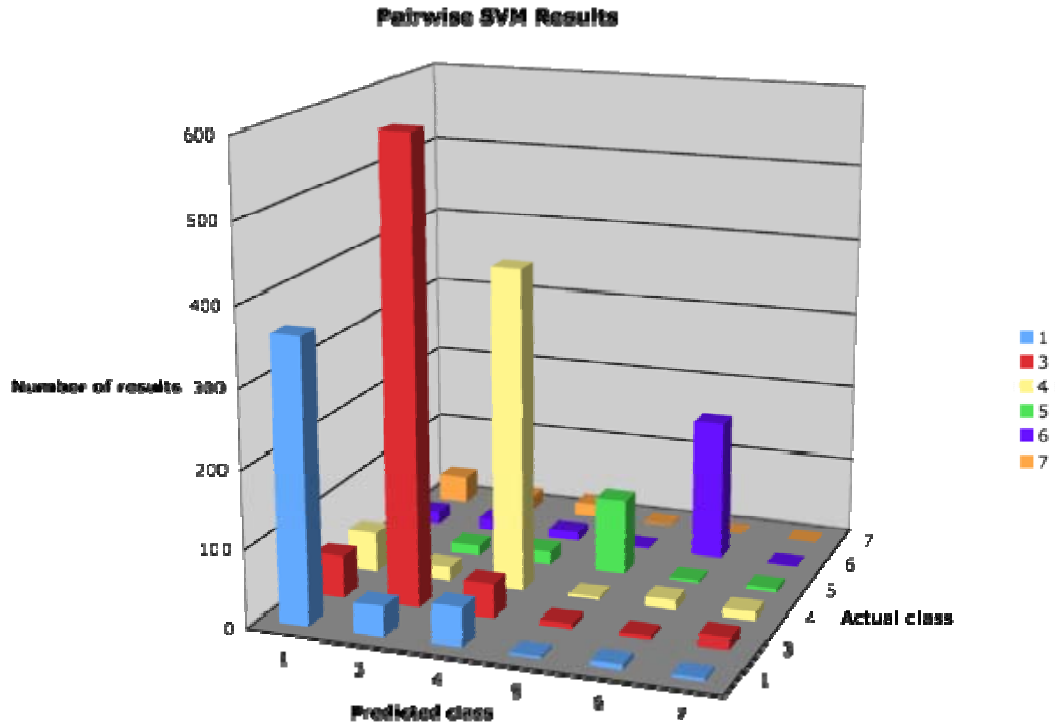


Figure 6: Graph of pairwise SVM results on data set

Table 3: Confusion matrix for pairwise SVM

	Estimated class						
Actual	1	3	4	5	6	7	
1	362	38	50	3	7	4	
3	55	592	44	7	5	17	
4	52	18	414	4	13	11	
5	7	14	18	99	3	5	
6	16	16	11	2	181	1	
7	35	18	18	1	0	341	

6 Conclusion

The expected results were that pairwise SVMs would be somewhat slower than OAA but produce more accurate results on the data set; surprisingly, both results and running time were comparable. It is possible that this is merely a feature of this particular data set, and a decisive advantage to one method or the other may emerge if different data sets are used. The performance of SVMs on this data is largely dependent upon the methods of feature extraction, and it may be worth following up on this work as the feature extraction on the data set improves.

References

- [1] Abe, S. (2003) Analysis of Multiclass Support Vector Machines. *International Conference on Computational Intelligence for Modelling Control and Automation (CIMCA 2003)*, 385-396.
- [2] Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):955-974.
- [3] Kijssirikul, B. & Ussivakul, N. (2002) Multiclass support vector machines using adaptive directed acyclic graph. *Proceedings of International Joint Conference on Neural Networks (IJCNN 2002)*, 980-985.
- [4] Vapnik, V. (1995) *The Nature of Statistical Learning Theory*. Springer-Verlag, London.
- [5] Vapnik, V. (1998). *Statistical Learning Theory*. Wiley-Interscience, NY.