Supervised Classification Using Graph-based Space Partitioning for Multiclass Problems

Adam Krzyżak and Karima Ben Suliman
Concordia University, Montreal, Canada
Nicola Yanev and Ventzeslav Valev
Bulgarian Academy of Sciences, Sophia, Bulgaria

Outline

- Supervised Classification as G-cut Problem
- Minimum Clique Cover Problem
- Definition of Box Classifier
- Application of Box Classifier to classification of normal data and nominal data from UCI Repository

Box Classifier

- The supervised multiclass classification algorithm called Box Classifier (BC) uses partition of the training data by multidimensional parallelepipeds called boxes.

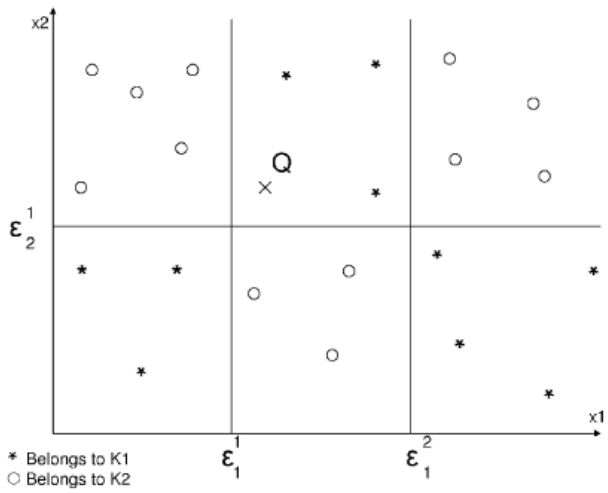
- We demonstrate how multiclass classification problems can be solved by combining the heuristic minimum clique cover approach and the *k*-nearest neighbor rule.

- Our Box algorithm is motivated by an algorithm for partitioning graph into a minimal number of cliques. .

- The main advantage of the Box classifier is that it optimally utilizes the geometrical structure of the training set by decomposing the *l*-class problem ($l > 2$) into *l* binary classification problems.

Supervised Classification as *G*-cut Problem

- The supervised classification problem can be formulated as the feature space partitioning problem, so-called ***G-cut problem***



- ***G-cut problem:***
  - Partition *n*-dimensional hyperparallelepiped into minimal number of hyperparallelepipeds (boxes) so that each of them contains either patterns belonging to only one of the classes or is an empty box.

*V. Valev, Supervised pattern recognition by parallel feature partitioning. *Pattern Recognition*, vol. 37, no.3, 2004, pp. 463–467

Supervised Classification as *G*-cut Problem

- *G-cut* was first formulated and solved in Valev (2004) using parallel feature partitioning.

- The solution was obtained by partitioning the feature space into a minimal number of nonintersecting regions by solving an integer-valued optimization problem.

- The learning phase consists of geometrical construction of the decision regions for classes in *n*-dimensional feature space.

- Let X_b and X_r be two training sets of patterns from two different classes and consider them as points colored in blue and red, respectively in the hypercube $F \in R^n$.

- During the learning phase the problem is to find $f(x)$ for $x \in R^n$ such that $f(x)<0$ for the blue points and $f(x)>0$ for the red points.

Supervised Classification as *G*-cut Problem

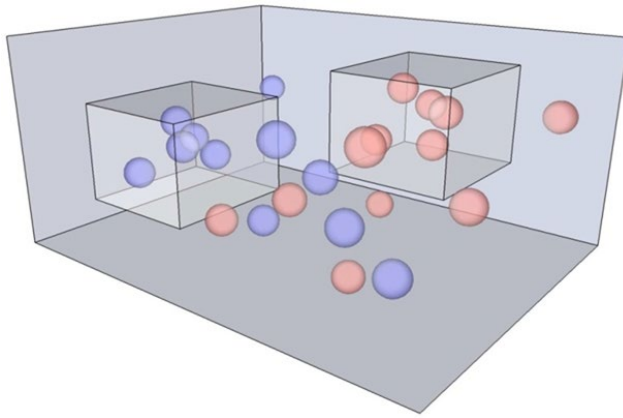
- In case when convex hulls $conv(X_r)$ and $conv(X_b)$ are linearly separable the classifier is linear, i.e., $f(x) = w \cdot x + b = 0$ is a linear discriminant function and SVM finds a hyperplane that maximizes the minimum distance to the training patterns (margin).

- In nonlinear case we look for a nonlinear function f that separates red and blue points.

- In the nonlinear case the notion of margin becomes complicated because the blue and red regions could be disconnected.

- The BC algorithm solves the supervised classification problem by reducing it to heuristically solving good clique cover problem satisfying the nearest neighbor rule.

Covering Classes by Colored Boxes

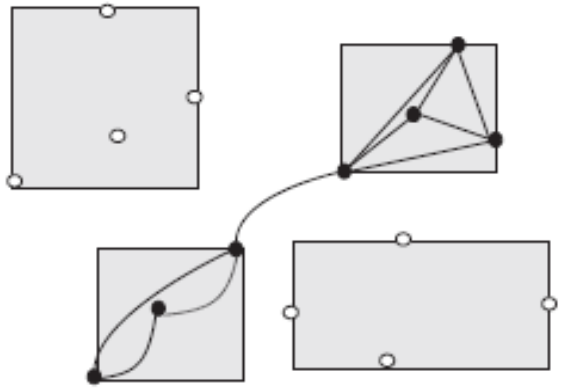


❑ ***Master Problem (MP):*** Cover all points in the training set with a minimal number of painted boxes, where painted boxes acquire the color of unicolor points they contain.

➢ MP reduces to the well-known *NP*-complete problem of Minimum Clique Cover Problem (*MCCP*).

➢ MCCP is to partition the vertex set of a graph into a minimum number of cliques.

➢ The following heuristic algorithm based on a greedy approach seems to be quite appropriate for MP (Östergard, 2002) which finds the maximum cardinality cliques in a given graph G.



Black and white points are patterns from two classes. The graph and the minimal clique cover are shown for one of the classes.

- Set $G = G_c$
- While $V_c \neq \varnothing$ do
- $V_{cur} = MCCP(G)$
- Set G to be the induced subgraph of G by the vertex set $V_c - V_{cur}$
- End do

A Minimal Clique Cover Box Algorithm

- **Problem : Cover the graph G_{Xb} with the minimum number of blue colored cliques (repeat for G_{Xr}).**

Input: graph graph *G(V,E)*

- ***Step 1.*** (reduce the graph *G*) Create the graph *WG =(WV,WE)* by use of non-dominated edges of *G*.
- ***Step 2***. (Clique enlargement) while $WV \neq 0$ **do**:

  - Call **try-to-extend**.

  - Create *WE*, save all isolated vertices (boxes) in FB (Final Boxes) and remove them from *WG*

  – end do

Try-to-extend (*WG*) :

input :graph *WG*

output : vertex set *WV* as a result from *d*-clique cover of the input graph *WG*

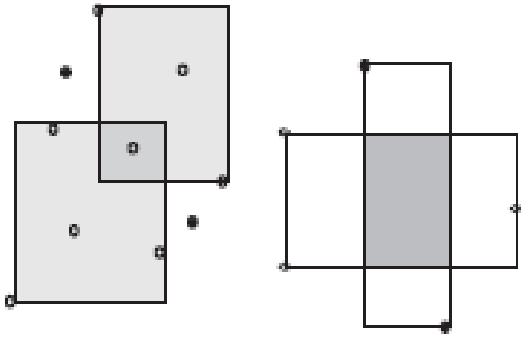
- (2-clique cover of *WG*) For all $\{u,v\} \in WE$ save $u \oplus v$ in *BS*
- return *WV* = *BS*

Box Algorithm Classifier

❑ If pattern **x** from the test dataset falls in a single-colored box or in the union of boxes with the same color the element **x** is assigned to the class that corresponds to this color.

❑ If pattern **x** from the test dataset falls in an empty (uncolored) box then the pattern **x** is not classified.

❑ Alternatively, if pattern **x** falls in the white region then it can be assigned to a class with color that corresponds to majority of the adjacent colored boxes.



Overlapping example. Left: overlapping for the same-colored boxes. Right: overlapping for two-colored boxes - the shaded area is empty.

Relation of Box Algorithm Classifier to Nearest Neighbor Rule

➢ If box $B = l_i \le x_i \le u_i, i = 1,...,n$ contains training patterns and ρ is the Manhattan distance, then for the pattern y the distance is equal to

$$\rho(y, B) = \sum_{i=1}^{n} \left[\max(0, l_i - y_i) + \max(0, y_i - u_i) \right]$$

➢ We first approximate the above-mentioned painted areas (not known in advance) by painted boxes (perfect candidates for Manhattan distance) and then classify test patterns according to point-to-box distance rule.

➢BC is similar to NN classifier if the Voronoi cells are taken to be boxes.

➢Note that instead of boxes we can consider convex hulls of patterns. Applying nearest neighbor rule we get better classification, but this approach is computationally intractable either for constructing convex hulls or for computing the point-to-set distances.

➢Now the MP(c) problem can be formulated as **heuristic good clique cover problem** satisfying the nearest neighbor rule.

Relation of Box Classifier to Tree Classifier

- The leaves of the decision trees (DT) are painted rectangles (boxes) obtained by starting with the sets of intervals and priority (for univariate trees) for creating successors of a given node. Internal nodes are rectangles containing a mixture of colored points.

- BC creates painted boxes (under Manhattan distance) that are convex hull of given sets of unicolored sets of points .

- Generally BC creates DT with a root node and a list of successors, i.e., colored boxes, containing unicolored points, whose coordinates satisfy the test on the arc for box membership.

- BA aims to create trees with minimum number of leaves and thus it attempts to reduce the generalization errors.

Experimental Results for Normal Attributes

- The samples for binary classification problem are generated with normal attributes.

- These samples are generated from three 3-dimensional normal distributions with mean vectors and covariance matrices given in the Table below, where $e=(1,1,1)^T$

- For each distribution, 100 samples are generated half of which are used for training and the rest for testing.

Case	Covariance matrices		Mean vectors	
1	*I*	*I*	0	*0.5e*
2	*I*	*2I*	0	*0.6e*
3	*I*	*4I*	0	*0.8e*

Parameter settings for normal distributions

Experimental Results

Normal Attributes

	Normal distribution					
	Accuracy			**Sensitivity**		
	1	2	3	1	2	3
k-NN	0.61	0.66	0.78	0.61	0.60	0.70
DT	0.58	0.62	0.75	0.58	0.63	0.74
SVM	0.64	0.69	0.81	0.55	0.57	0.73
BC	**0.98**	**0.98**	**0.98**	**0.98**	**0.99**	**0.99**
	Specificity			**Precision**		
	1	2	3	1	2	3
k-NN	0.60	0.73	0.86	0.61	0.69	0.83
DT	0.59	0.62	0.75	0.58	0.62	0.75
SVM	0.72	0.81	0.89	0.66	0.75	0.89
BC	**0.98**	**0.97**	**0.96**	**0.98**	**0.97**	**0.96**

Accuracy, sensitivity, specificity, and precision of k-NN, DT, SVM classifiers and Box classifier for normally distributed data for average of 50 runs.

Experimental Results for Nominal Attributes for Real Data from UCI Repository

Name	Data type	# of attributes	# of instances	# of classes
Bank authentication	Real	5	1372	2
Breast Cancer Coimbra	Integer	10	116	2
Cardiotocography	Real	23	2126	3
Teaching Assistant Evaluation (TA)	Categorical, Integer	5	151	3
Breast tissue	Real	10	106	6
Lenses	Categorical	4	24	3
Glass	Real	10	214	6
Hepatitis C Virus (HSV)	Integer, Real	29	106	4

Experimental Results

Binary Classification

	Bank authentication	Breast Cancer Coimbra
	Accuracy	
k-NN	0.99	0.67
DT	0.98	0.63
SVM	0.99	0.67
RF	0.99	0.67
BC	0.99	0.67
Sensitivity		
k-NN	0.98	0.83
DT	0.98	0.83
SVM	0.99	0.33
RF	0.99	0.17
BC	0.99	0.17

	Bank authentication	Breast Cancer Coimbra
	Specificity	
k-NN	1.00	0.61
DT	0.98	0.56
SVM	0.98	0.78
RF	0.98	0.83
BC	1.00	0.83
False positive rate		
k-NN	0	0.40
DT	0.02	0.45
SVM	0.02	0.22
RF	0.02	0.17
BC	**0**	0.17

Accuracy, sensitivity, specificity, and precision of k-NN, DT, SVM, RF classifiers and Box classifier for binary data using 80% and 20% splitting

Experimental Results

Multiclass Classification

	Cardiotocography	TA	Breast tissue	Lenses	Glass	HSV
	Accuracy					
k-NN	0.80	0.45	0.64	0.60	0.53	0.32
DT	0.81	0.45	0.64	0.40	0.49	0.32
SVM	0.80	0.52	0.68	0.60	0.53	0.27
RF	0.89	0.48	0.64	0.60	0.53	0.32
BC	**0.90**	**0.74**	**0.73**	0.60	**0.56**	**0.36**
	Sensitivity					
k-NN	0.33	0.46	0.64	0.67	0.35	0.30
DT	**0.86**	0.40	0.64	0.33	0.41	0.33
SVM	0.33	0.53	0.69	0.67	0.38	0.29
RF	0.75	0.46	0.64	0.67	**0.47**	0.31
BC	0.84	**0.79**	**0.74**	0.67	0.41	**0.37**

Accuracy and sensitivity of k-NN, DT, SVM, RF classifiers and Box classifier for different data from UCI using 80% and 20% splitting

Experimental Results

Multiclass Classification

	Cardiotocography	TA	Breast tissue	Lenses	Glass	HSV
	Specificity					
K-NN	0.67	0.73	0.93	0.78	0.89	0.77
DT	0.93	0.72	0.93	0.67	0.89	0.77
SVM	0.67	0.76	0.94	0.78	0.90	0.76
RF	0.82	0.74	0.93	**0.83**	0.90	0.76
BC	**0.95**	**0.89**	**0.95**	0.78	0.90	**0.79**
	False positive rate					
K-NN	0.33	0.27	0.07	0.22	0.11	0.23
DT	0.07	0.28	0.07	0.33	0.11	0.23
SVM	0.33	0.24	0.06	0.22	0.10	0.24
RF	0.18	0.26	0.07	**0.17**	0.10	0.24
BC	**0.05**	**0.11**	**0.05**	0.22	**0.09**	**0.21**

Specificity and FP rate of k-NN, DT, SVM, RF classifiers and Box classifier for different data from UCI using 80% and 20% splitting

Experimental Results

Multiclass Classification

Data	BC	k-NN	DT	SVM	RF
	Accuracy				
Bank authentication	**1.00**	0.99	0.98	0.99	0.99
Breast Cancer Coimbra	**0.83**	0.76	0.75	0.79	**0.83**
Teaching Assistant Evaluation	**0.93**	0.58	0.58	0.57	0.53
Breast tissue	**0.72**	0.70	0.68	0.71	0.57
Lenses	**0.88**	0.79	0.87	0.87	0.83
Glass	**0.68**	0.67	0.66	**0.68**	0.54
Hepatitis C Virus (HSV)	1.00	1.00	1.00	1.00	1.00

The accuracy of k-NN, DT, SVM, RF classifiers and Box algorithm for different data from UCI using 10-fold cross validation

Experimental Results

Computational Complexity

Data	Execution Times (in seconds)				
	BC	**k-NN**	**DT**	**SVM**	**RF**
Bank authentication	0.93	1.00	**0.85**	1.40	1.54
Breast Cancer Coimbra	**0.40**	0.81	0.65	1.10	1.38
Teaching Assistant Evaluation	**0.09**	0.79	0.78	1.33	1.45
Breast tissue	**0.04**	0.73	0.78	3.13	1.69
Lenses	**0.03**	0.73	0.73	1.14	1.18
Glass	**0.31**	0.89	0.69	1.49	1.26
Hepatitis C Virus (HSV)	**0.08**	0.91	0.68	7.22	1.38
Time on average	**0.27**	0.83	0.73	2.40	1.41

Conclusions

- An efficient geometrical approach for solving multiclass supervised classification problem based on graph optimization approach is introduced.

- This efficient classifier is motivated by the graph optimization approach based on partitioning the graph into a minimum number of maximal size cliques which are subsequently merged using the nearest neighbor rule.

- The multiclass supervised classification problem is solved by means of a heuristic minimum clique cover problem satisfying the nearest neighbor rule called the Box classifier.

- Comprehensive experiments on real binary/mutliclass and simulated data are presented showing superior performance of Box classifier versus k-NN, DT, SVM and RF.

- It is demonstrated in experiments that the proposed Box classifier has low computational complexity when compared to k-NN, DT, SVM and RF.