# The Class Cover Problem and its Applications to Pattern Recognition

Jason G. DeVinney

devinney@mts.jhu.edu

Mathematical Sciences Department
Johns Hopkins University

History of the Class Cover Problem (CCP)

- History of the Class Cover Problem (CCP)
- Introduction to Class Cover Problem

- History of the Class Cover Problem (CCP)
- Introduction to Class Cover Problem
- Class Cover Catch Digraphs (CCCD)

- History of the Class Cover Problem (CCP)
- Introduction to Class Cover Problem
- Class Cover Catch Digraphs (CCCD)
  - Characterization of Euclidean CCCDs

- History of the Class Cover Problem (CCP)
- Introduction to Class Cover Problem
- Class Cover Catch Digraphs (CCCD)
  - Characterization of Euclidean CCCDs
- Random CCCD's

- History of the Class Cover Problem (CCP)
- Introduction to Class Cover Problem
- Class Cover Catch Digraphs (CCCD)
  - Characterization of Euclidean CCCDs
- Random CCCD's
  - Results in one dimension

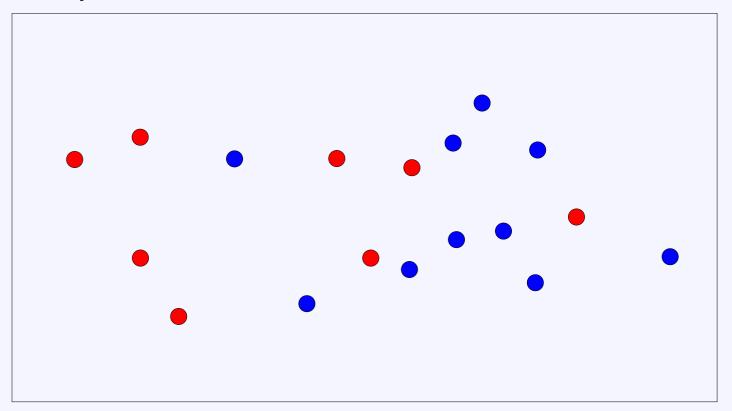
- History of the Class Cover Problem (CCP)
- Introduction to Class Cover Problem
- Class Cover Catch Digraphs (CCCD)
  - Characterization of Euclidean CCCDs
- Random CCCD's
  - Results in one dimension
- Application to Classification

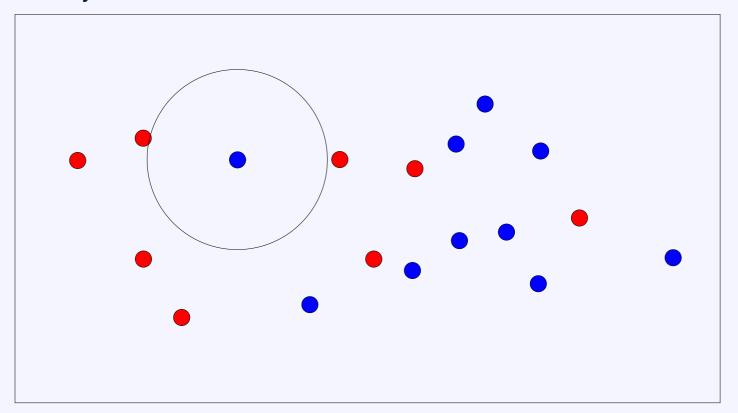
- History of the Class Cover Problem (CCP)
- Introduction to Class Cover Problem
- Class Cover Catch Digraphs (CCCD)
  - Characterization of Euclidean CCCDs
- Random CCCD's
  - Results in one dimension
- Application to Classification
  - CCP-based Classifiers

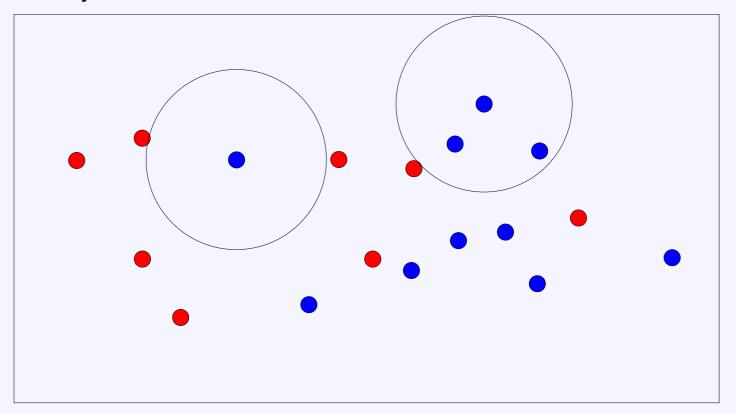
- History of the Class Cover Problem (CCP)
- Introduction to Class Cover Problem
- Class Cover Catch Digraphs (CCCD)
  - Characterization of Euclidean CCCDs
- Random CCCD's
  - Results in one dimension
- Application to Classification
  - CCP-based Classifiers
  - Simulation Results

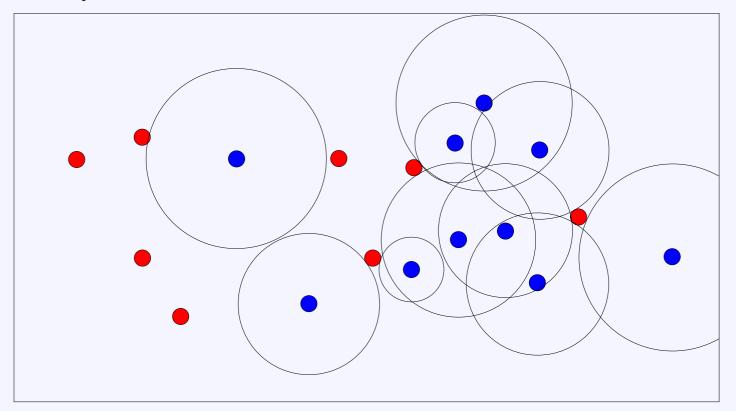
## **History of the Class Cover Problem**

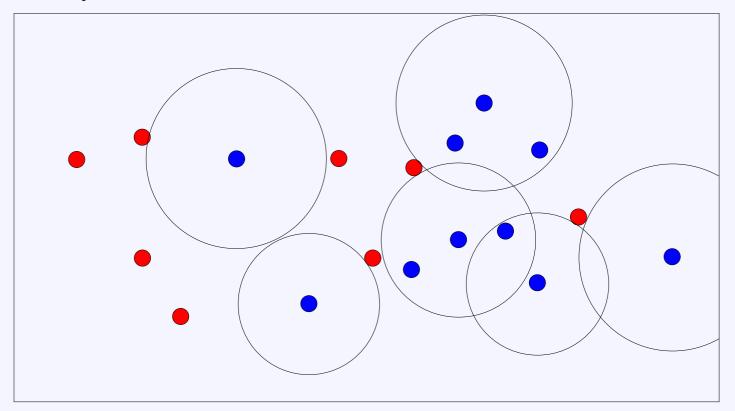
- Approximate Distance Clustering
   Cowen and Priebe
   Advances in Applied Mathematics, 19 (1997)
- Class Cover Problem
   Cannon and Cowen
   Annals of Math and A.I. to appear
- Closely Related Topics
  - Sphere Covers, Hochbaum and Maass (J. of the A.C.M. 1985)
  - Sphere Digraphs, Maehara (J.G.T. 1984)
  - Sphere of Attraction Graphs,
     McMorris and Wang (Con. Num. 2000)











# **Catch Digraphs**

Given a collection of sets  $\{S_1, S_2, \dots, S_n\}$  with associated "base" points  $\{t_1, t_2, \dots, t_n\}$ .

We form the catch digraph D with  $V = \{v_1, v_2, \dots, v_n\}$  and a directed edge from  $v_i$  to  $v_j$  iff  $t_j \in S_i$ .

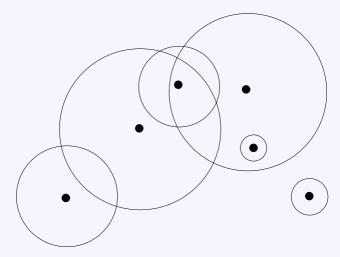
A *sphere digraph* is a catch digraph where the sets are spheres and the base point for each sphere is its center.

# **Catch Digraphs**

Given a collection of sets  $\{S_1, S_2, \dots, S_n\}$  with associated "base" points  $\{t_1, t_2, \dots, t_n\}$ .

We form the catch digraph D with  $V = \{v_1, v_2, \dots, v_n\}$  and a directed edge from  $v_i$  to  $v_j$  iff  $t_j \in S_i$ .

A *sphere digraph* is a catch digraph where the sets are spheres and the base point for each sphere is its center.

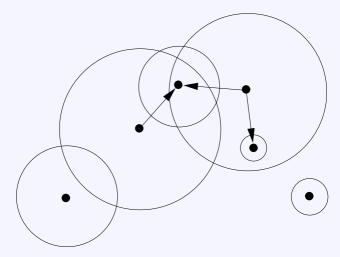


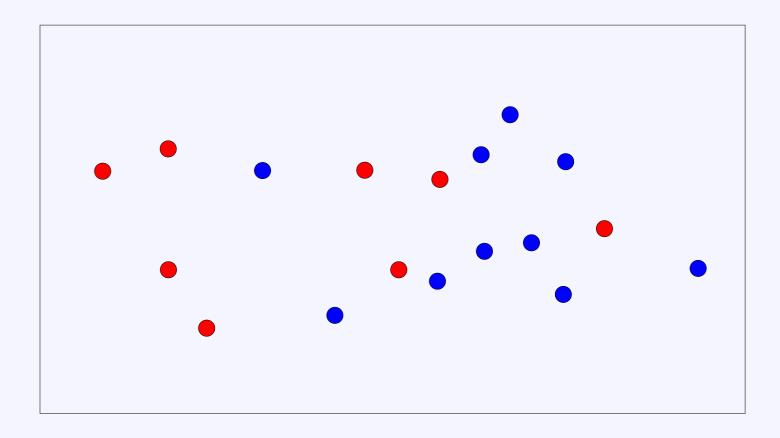
# **Catch Digraphs**

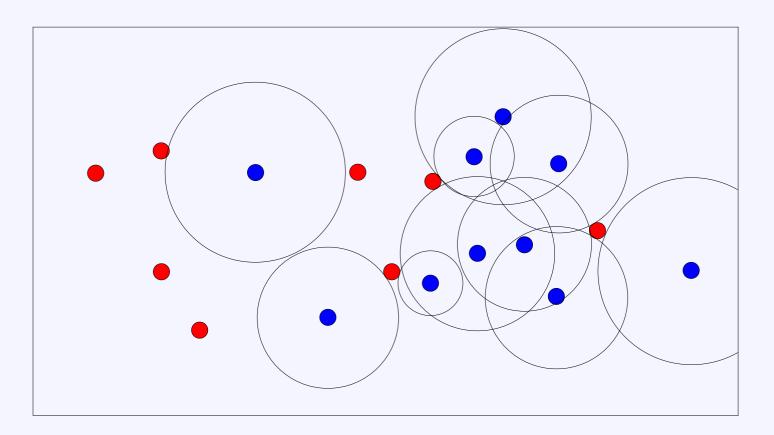
Given a collection of sets  $\{S_1, S_2, \ldots, S_n\}$  with associated "base" points  $\{t_1, t_2, \ldots, t_n\}$ .

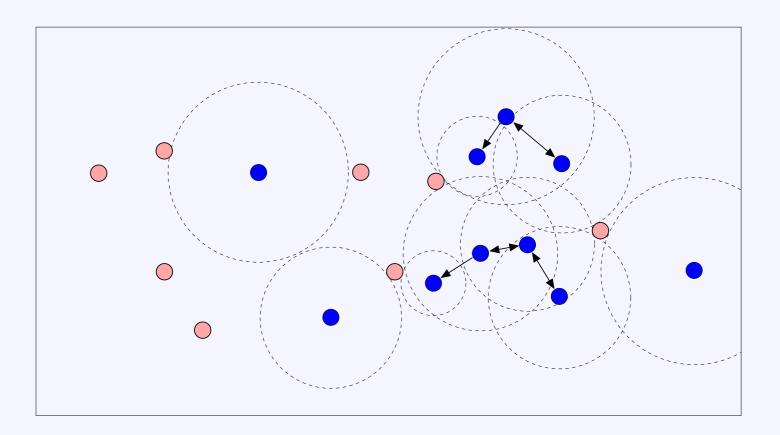
We form the catch digraph D with  $V = \{v_1, v_2, \dots, v_n\}$  and a directed edge from  $v_i$  to  $v_j$  iff  $t_j \in S_i$ .

A *sphere digraph* is a catch digraph where the sets are spheres and the base point for each sphere is its center.



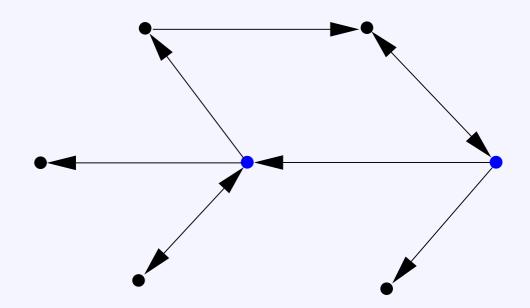






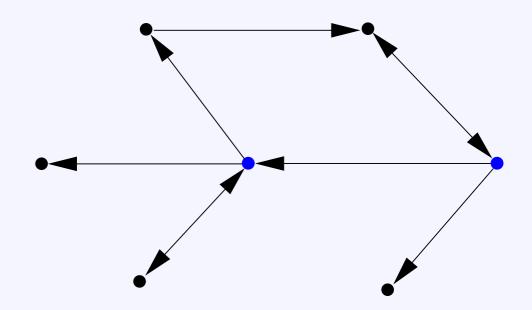
## **Dominating Sets**

Define a dominating set, S, of a digraph D=(V,A) as follows:  $S\subset V$  such that  $\forall v\in V,\ v\in S$  or  $\exists w\in S$  such that  $(w,v)\in A$ .



## **Dominating Sets**

Define a dominating set, S, of a digraph D=(V,A) as follows:  $S\subset V$  such that  $\forall v\in V,\ v\in S$  or  $\exists w\in S$  such that  $(w,v)\in A$ .

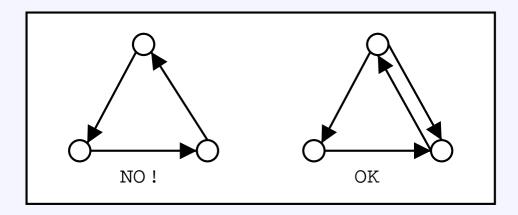


Solution to CCP ⇔ Minimum Size Dominating Set in CCCD

### **Euclidean CCCD's**

If  $X, Y \subset \Re^q$  and d is the standard Euclidean metric, then we call the resulting digraph a Euclidean CCCD.

What digraphs are Euclidean CCCD's?



Definition: A simple cycle is a directed cycle with no bidirected edges.

## **Euclidean CCCD's**

**Theorem 1** A digraph is a Euclidean CCCD if and only if it has no simple cycles.

In other words...

Given  $X, Y \in \mathbb{R}^q$ , they induce a digraph with no simple cycles. Given a digraph D with no simple cycles, we can find  $X, Y \in \mathbb{R}^q$  which induce a digraph isomorphic to D (in fact we will need only one Y point, but we may need many dimensions).

- $(\Leftarrow)$  Given D with a smallest simple cycle  $v_1, v_2, \ldots, v_l$ .
  - $(v_i,v_{i+1}) \in A \text{ and } (v_i,v_{i-1}) \not\in A \text{ implies } d(x_i,x_{i-1}) > d(x_i,x_{i+1})$   $\forall i \text{ modulo } l$

- $(\Leftarrow)$  Given D with a smallest simple cycle  $v_1, v_2, \ldots, v_l$ .
  - $(v_i, v_{i+1}) \in A \text{ and } (v_i, v_{i-1}) \not\in A \text{ implies } d(x_i, x_{i-1}) > d(x_i, x_{i+1})$   $\forall i \text{ modulo } l$
  - This leads to a set of cyclic inequalities

$$d(x_1, x_2) > d(x_2, x_3) > \ldots > d(x_1, x_2)$$

- $(\Leftarrow)$  Given D with a smallest simple cycle  $v_1, v_2, \ldots, v_l$ .
  - $(v_i, v_{i+1}) \in A \text{ and } (v_i, v_{i-1}) \not\in A \text{ implies } d(x_i, x_{i-1}) > d(x_i, x_{i+1})$   $\forall i \text{ modulo } l$
  - This leads to a set of cyclic inequalities  $d(x_1, x_2) > d(x_2, x_3) > \ldots > d(x_1, x_2)$
- $(\Rightarrow)$  Given D with no simple cycles.

- $(\Leftarrow)$  Given D with a smallest simple cycle  $v_1, v_2, \ldots, v_l$ .
  - $(v_i, v_{i+1}) \in A \text{ and } (v_i, v_{i-1}) \not\in A \text{ implies } d(x_i, x_{i-1}) > d(x_i, x_{i+1})$   $\forall i \text{ modulo } l$
  - This leads to a set of cyclic inequalities  $d(x_1, x_2) > d(x_2, x_3) > \ldots > d(x_1, x_2)$
- $(\Rightarrow)$  Given D with no simple cycles.
  - Create a poset on interpoint distances

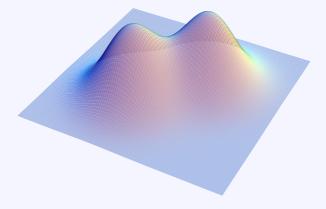
- $(\Leftarrow)$  Given D with a smallest simple cycle  $v_1, v_2, \ldots, v_l$ .
  - $(v_i, v_{i+1}) \in A \text{ and } (v_i, v_{i-1}) \not\in A \text{ implies } d(x_i, x_{i-1}) > d(x_i, x_{i+1})$   $\forall i \text{ modulo } l$
  - This leads to a set of cyclic inequalities  $d(x_1, x_2) > d(x_2, x_3) > \ldots > d(x_1, x_2)$
- $(\Rightarrow)$  Given D with no simple cycles.
  - Create a poset on interpoint distances
  - Extend poset to linear order

- $(\Leftarrow)$  Given D with a smallest simple cycle  $v_1, v_2, \ldots, v_l$ .
  - $(v_i, v_{i+1}) \in A \text{ and } (v_i, v_{i-1}) \not\in A \text{ implies } d(x_i, x_{i-1}) > d(x_i, x_{i+1})$   $\forall i \text{ modulo } l$
  - This leads to a set of cyclic inequalities  $d(x_1, x_2) > d(x_2, x_3) > \ldots > d(x_1, x_2)$
- $(\Rightarrow)$  Given D with no simple cycles.
  - Create a poset on interpoint distances
  - Extend poset to linear order
  - Assign interpoint distances

- $(\Leftarrow)$  Given D with a smallest simple cycle  $v_1, v_2, \ldots, v_l$ .
  - $(v_i, v_{i+1}) \in A \text{ and } (v_i, v_{i-1}) \not\in A \text{ implies } d(x_i, x_{i-1}) > d(x_i, x_{i+1})$   $\forall i \text{ modulo } l$
  - This leads to a set of cyclic inequalities  $d(x_1, x_2) > d(x_2, x_3) > \ldots > d(x_1, x_2)$
- $(\Rightarrow)$  Given D with no simple cycles.
  - Create a poset on interpoint distances
  - Extend poset to linear order
  - Assign interpoint distances
  - Show Euclidean interpoint distances exist which preserve the linear order using MDS.

#### **Random Model of CCP**

n target class points are chosen from a distribution  $F_X$  m non-target class points are chosen from a distribution  $F_Y$   $\Gamma_{n,m}$  := the domination number of the CCCD induced by the sample.



Question: What is the distribution of the random variable  $\Gamma_{n,m}$ ?

## Random Model in $\Re$

Let  $F_X$  and  $F_Y$  be the uniform distribution on [0,1]. In this case we use  $\Gamma_{n,m}$  to represent the case with n target class points and m non-target class points.

### Random Model in $\Re$

Let  $F_X$  and  $F_Y$  be the uniform distribution on [0,1]. In this case we use  $\Gamma_{n,m}$  to represent the case with n target class points and m non-target class points.

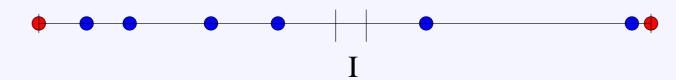


We need only consider the simplified case where  $y_1=0,y_2=1$  and  $F_X$  is the uniform distribution on [0,1]. We call such an induced CCCD  $C_n^*$ .



### Random Model in R

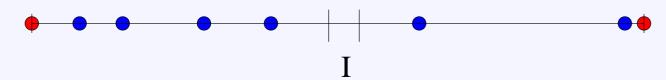
$$\Gamma(C_n^*) = \begin{cases} 0 & \text{if } n = 0 \\ 1 & \text{if } I \cap X \neq \emptyset \\ 2 & \text{otherwise} \end{cases}$$



where 
$$I = \left[\frac{x_n}{2}, \frac{1+x_1}{2}\right]$$
  
 $\kappa(n) = P[\Gamma(C_n^*) = 1] = \frac{5}{9} + \frac{4}{9} \cdot 4^{-(n-1)}$ 

### Random Model in R

$$\Gamma(C_n^*) = \begin{cases} 0 & \text{if } n = 0 \\ 1 & \text{if } I \cap X \neq \emptyset \\ 2 & \text{otherwise} \end{cases}$$



where 
$$I=\left[\frac{x_n}{2},\frac{1+x_1}{2}\right]$$
 
$$\kappa(n)=P[\Gamma(C_n^*)=1]=\frac{5}{9}+\frac{4}{9}\cdot 4^{-(n-1)}$$

This result allows the calculation of...

### Random Model in $\Re$

 $\dots$  probability mass function for  $\Gamma_{n,m}$ 

$$P[\Gamma_{n,m} = d] = \frac{n!m!}{(n+m)!} \sum_{\vec{n} \in \Delta_{n,m+1}^{Z_{n+1}}} \sum_{\vec{d} \in \Delta_{d,m+1}^{Z_3}} \alpha(d_1, n_1) \cdot \alpha(d_{m+1}, n_{m+1}) \prod_{j=2}^m \beta(d_j, n_j)$$

and...

### Random Model in $\Re$

 $\dots$  probability mass function for  $\Gamma_{n,m}$ 

$$P[\Gamma_{n,m} = d] = \frac{n!m!}{(n+m)!} \sum_{\vec{n} \in \Delta_{n,m+1}^{Z_{n+1}}} \sum_{\vec{d} \in \Delta_{d,m+1}^{Z_3}} \alpha(d_1, n_1) \cdot \alpha(d_{m+1}, n_{m+1}) \prod_{j=2}^{m} \beta(d_j, n_j)$$

and... expected value of  $\Gamma_{n,m}$ ,

$$E[\Gamma_{n,m}] = \frac{2n}{n+m} + \frac{n!m(m-1)}{(n+m)!} \sum_{i=1}^{n} \frac{(n+m-i-1)!}{(n-i)!} \cdot (2-\kappa(i))$$

and...

### Random Model in R

 $\dots$  probability mass function for  $\Gamma_{n,m}$ 

$$P[\Gamma_{n,m} = d] = \frac{n!m!}{(n+m)!} \sum_{\vec{n} \in \Delta_{n,m+1}^{Z_{n+1}}} \sum_{\vec{d} \in \Delta_{d,m+1}^{Z_3}} \alpha(d_1, n_1) \cdot \alpha(d_{m+1}, n_{m+1}) \prod_{j=2}^{m} \beta(d_j, n_j)$$

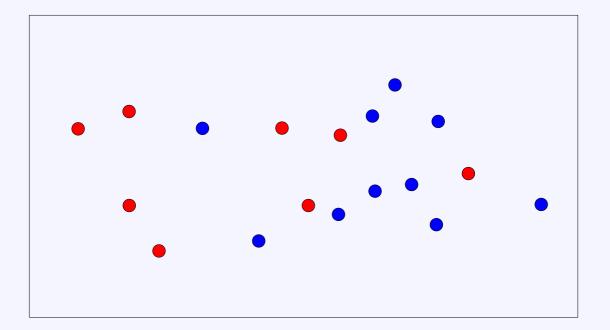
and... expected value of  $\Gamma_{n,m}$ ,

$$E[\Gamma_{n,m}] = \frac{2n}{n+m} + \frac{n!m(m-1)}{(n+m)!} \sum_{i=1}^{n} \frac{(n+m-i-1)!}{(n-i)!} \cdot (2-\kappa(i))$$

and... almost sure limit of  $\Gamma_{n,m}$ ,

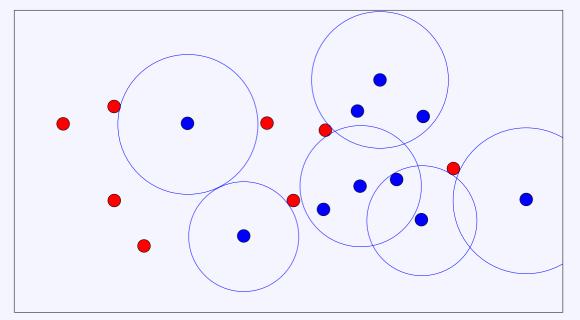
$$\lim_{n\to\infty}\frac{\Gamma_{\lfloor an\rfloor,n}}{n}=\frac{a(13a+12)}{3(a+1)(3a+4)}\text{ a.s. }a\in(0,\infty).$$

Our framework for using the CCP in classification is a reduced nearest neighbor with a dissimilarity dependent on the balls in the classifier.



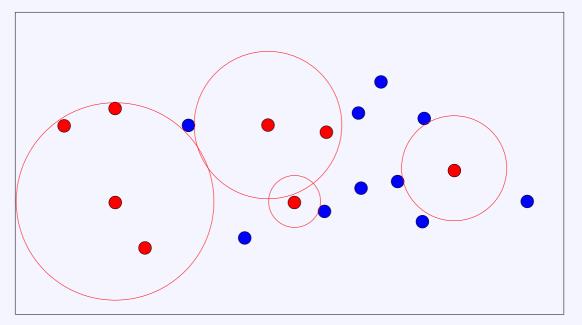
Our framework for using the CCP in classification is a reduced nearest neighbor with a dissimilarity dependent on the balls in the classifier.

Find a cover  $C_X$  for X.

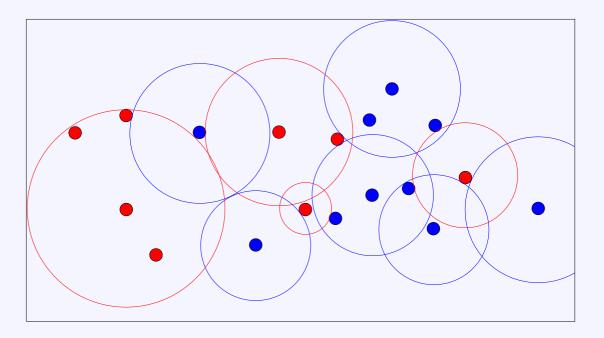


Our framework for using the CCP in classification is a reduced nearest neighbor with a dissimilarity dependent on the balls in the classifier.

Find a cover  $C_Y$  for Y.



Our framework for using the CCP in classification is a reduced nearest neighbor with a dissimilarity dependent on the balls in the classifier.



Create a dissimilarity function  $\rho(z,C)$  to measure the dissimilarity between a point  $z\in\Omega$  and a cover C.

#### **Naive or Pre-classifier**

Given training data  $X,Y \subset \Omega$  and covers  $C_X,C_Y$ . Define

$$\rho(z,C) := 1 - I\{z \in C\}$$

Using the reduced nearest neighbor framework,

$$g(z) := \left\{ egin{array}{ll} 1 &:& 
ho(z,C_1) < 
ho(z,C_2) \ 2 &:& 
ho(z,C_1) > 
ho(z,C_2) \ 0 &:& ext{otherwise} \end{array} 
ight.$$

#### **Naive or Pre-classifier**

Given training data  $X,Y \subset \Omega$  and covers  $C_X,C_Y$ . Define

$$\rho(z,C) := 1 - I\{z \in C\}$$

Using the reduced nearest neighbor framework,

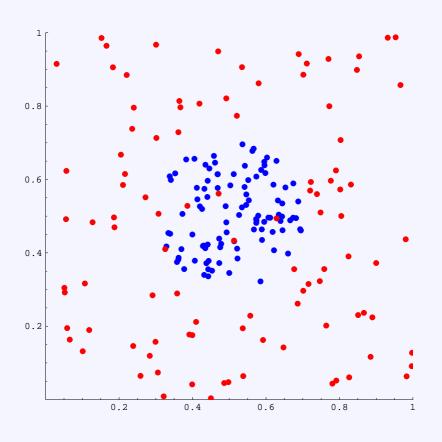
$$g(z) := \begin{cases} 1 & : & \rho(z, C_1) < \rho(z, C_2) \\ 2 & : & \rho(z, C_1) > \rho(z, C_2) \\ 0 & : & \text{otherwise} \end{cases}$$

or more simply,

$$g(z) := \begin{cases} 1 : z \in C_1 \cap C_2^c \\ 2 : z \in C_2 \cap C_1^c \\ 0 : \text{ otherwise} \end{cases}$$

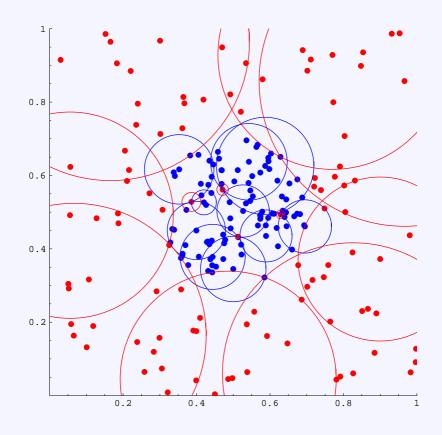
### **Simulation: Disk**

100 Red Points  $\sim U([0,1]\times[0,1]).$ 100 Blue Points  $\sim U(B((0.5,0.5),0.2))$ 



### **Simulation: Disk**

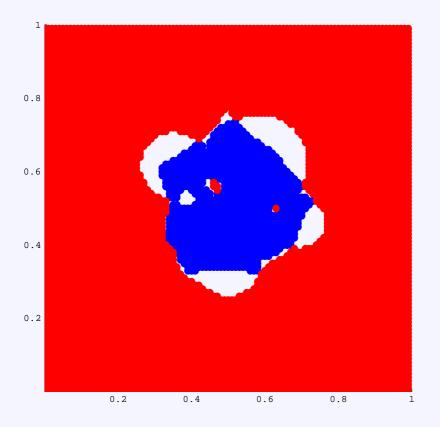
100 Red Points  $\sim U([0,1] \times [0,1]).$ 100 Blue Points  $\sim U(B((0.5,0.5),0.2))$ 



Drawbacks of Naive Classifier - overfitting

### **Simulation: Disk**

100 Red Points  $\sim U([0,1] \times [0,1])$ . 100 Blue Points  $\sim U(B((0.5,0.5),0.2))$ 



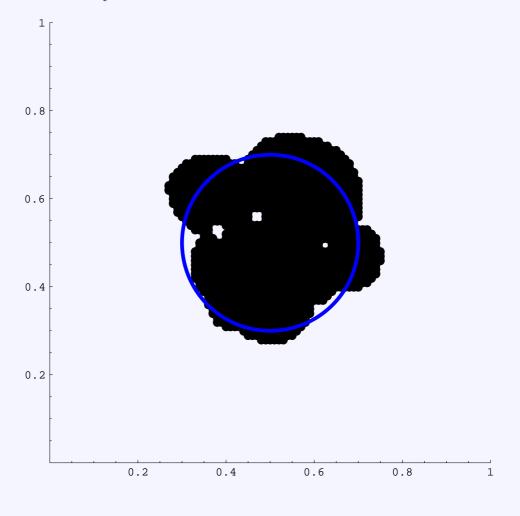
Drawbacks of Naive Classifier - overfitting and "no decision" regions.

# **Scaled Dissimilarity Function**

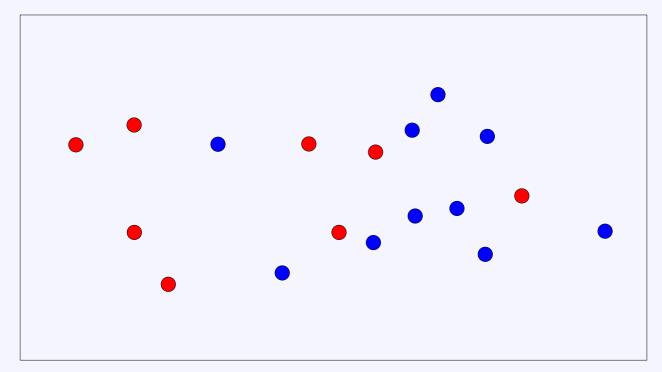
 $\rho(z,C):=\min_{B_i\in C} rac{d(c_i,z)}{r_i}$  where a ball  $B_i$  has center  $c_i$  and radius  $r_i$ .

## **Scaled Dissimilarity Function**

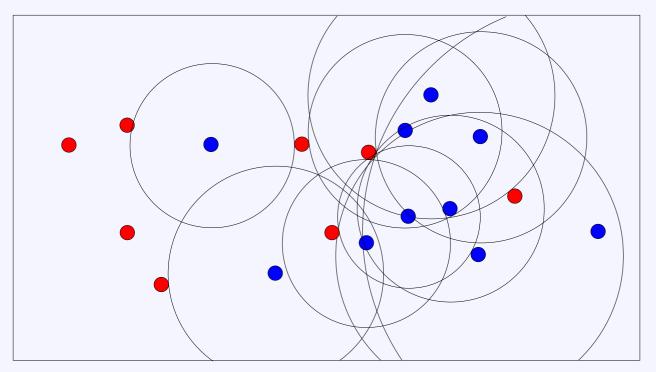
 $\rho(z,C) := \min_{B_i \in C} \frac{d(c_i,z)}{r_i}$  where a ball  $B_i$  has center  $c_i$  and radius  $r_i$ .



 $CCP(\alpha, \beta) :=$  the CCP where each ball may cover  $\beta$  of non-target class points and the cover may "miss" at most  $\alpha$  of the target class points.

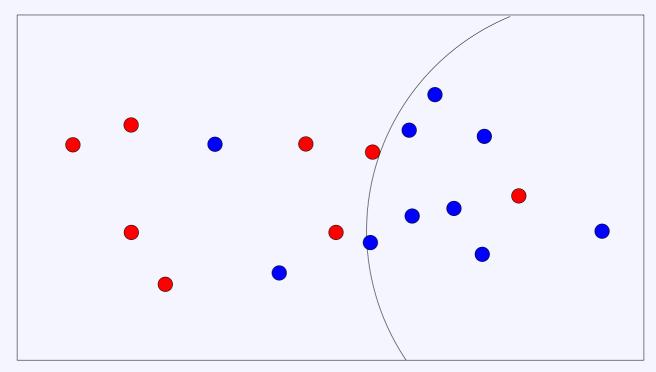


 $CCP(\alpha, \beta) :=$  the CCP where each ball may cover  $\beta$  of non-target class points and the cover may "miss" at most  $\alpha$  of the target class points.



$$\beta = 1$$

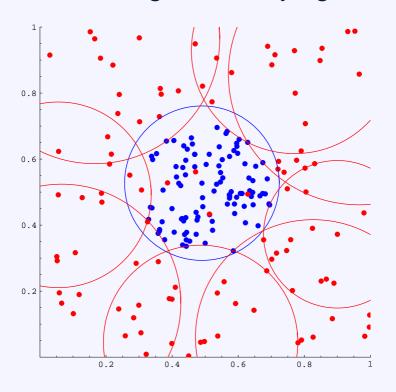
 $CCP(\alpha, \beta) :=$  the CCP where each ball may cover  $\beta$  of non-target class points and the cover may "miss" at most  $\alpha$  of the target class points.



$$\beta = 1$$
  $\alpha = 2$ 

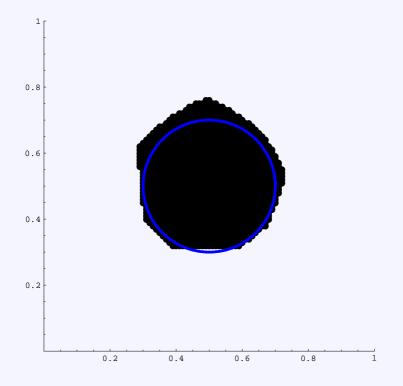
#### Benefits of $\alpha, \beta$ CCP.

- $lue{\alpha}$  parameter allows us to "ignore" outlying target class points. Moving toward modeling the discriminant region.
- $\blacksquare$   $\beta$  parameter allows us to "ignore" outlying non-target class points.



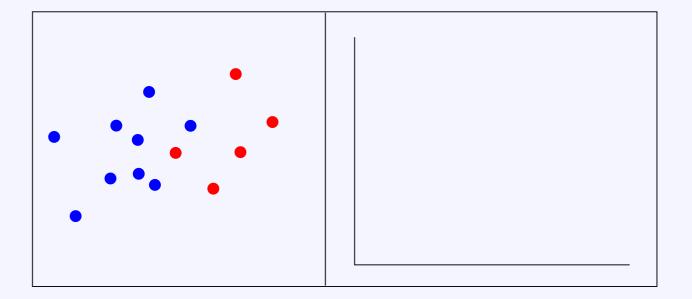
#### Benefits of $\alpha, \beta$ CCP.

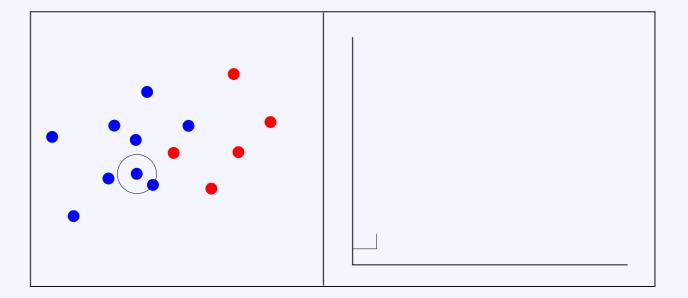
- $lue{\alpha}$  parameter allows us to "ignore" outlying target class points. Moving toward modeling the discriminant region.
- $\blacksquare$   $\beta$  parameter allows us to "ignore" outlying non-target class points.

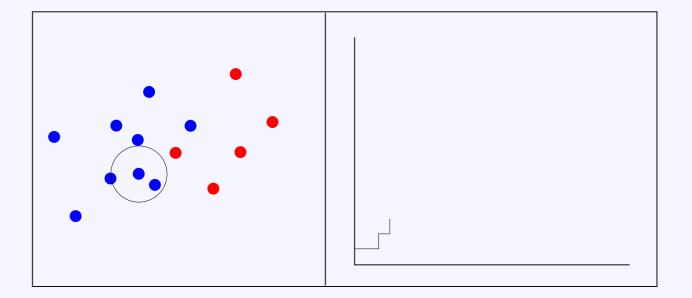


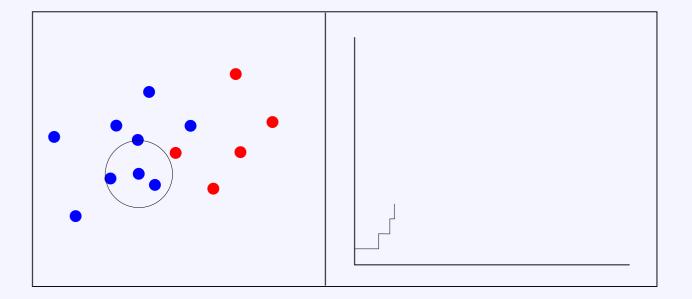
Disadvantages of  $\alpha$ , $\beta$  CCP.

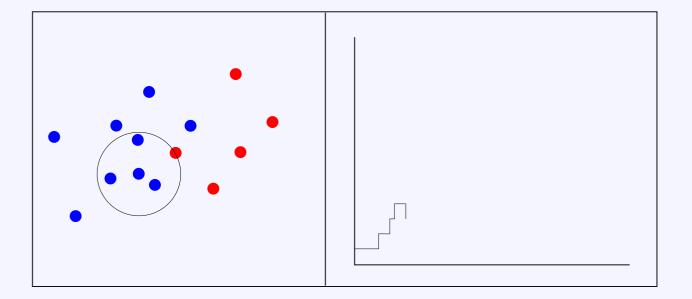
- User must choose parameters  $\alpha, \beta$  for each class.
- lacksquare parameters are global. We would only like balls to cover non-target class points in the discriminant region of the target class.

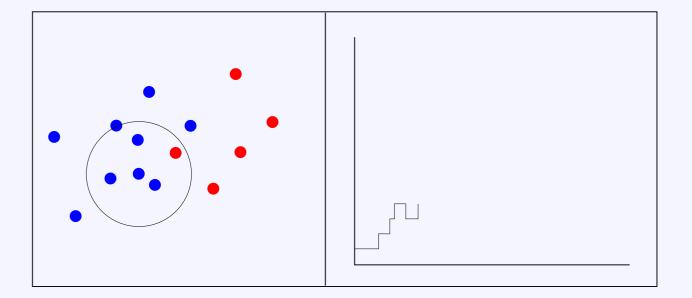


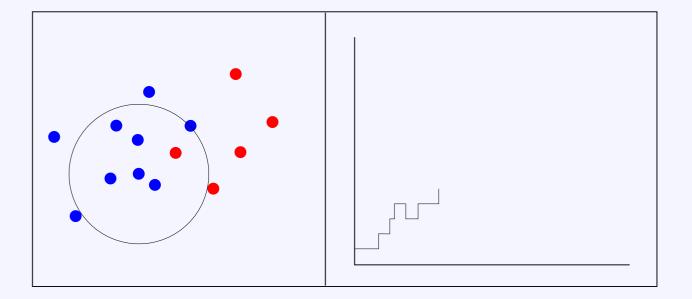




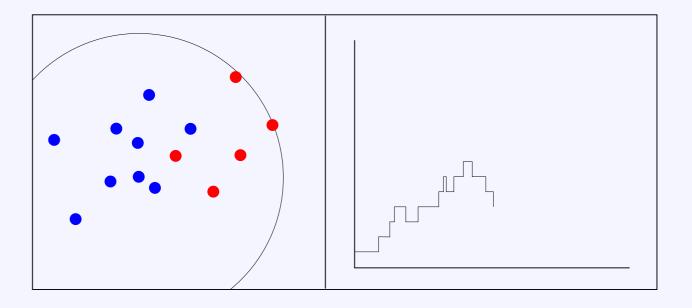






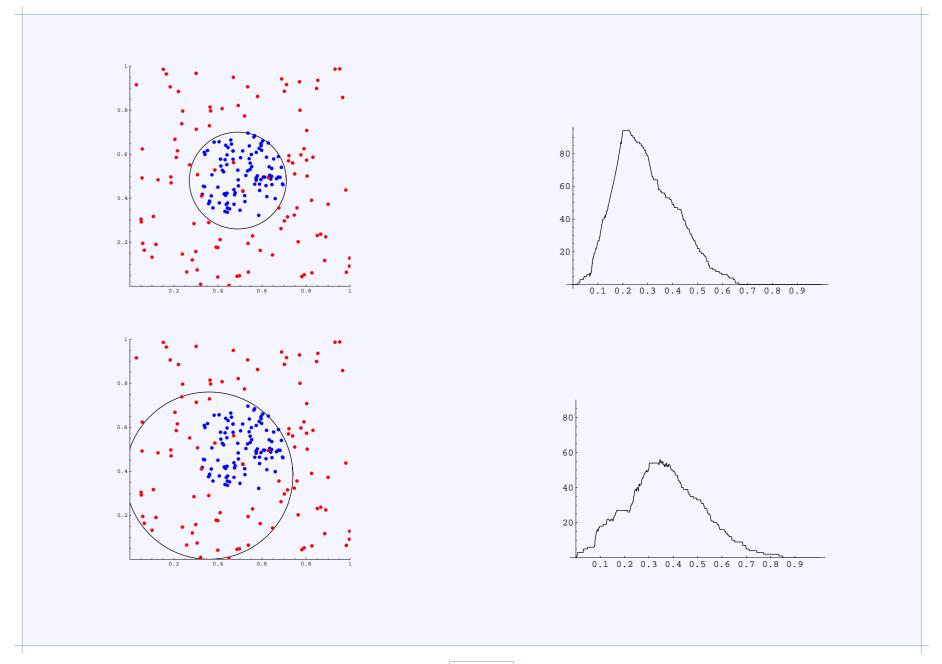


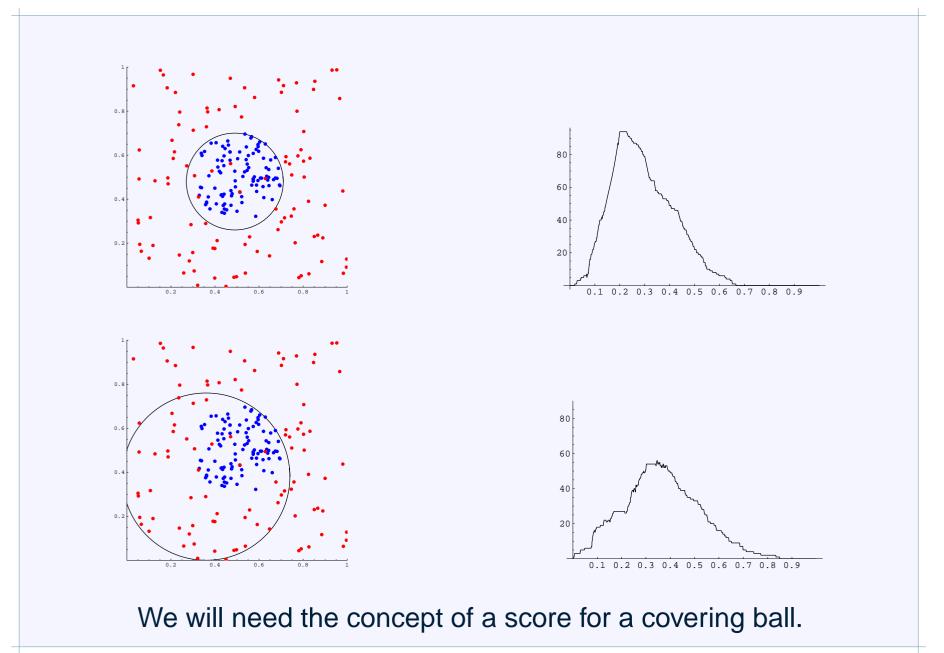
Idea: Let each point choose the radius of its covering ball.



We choose the radius according to:

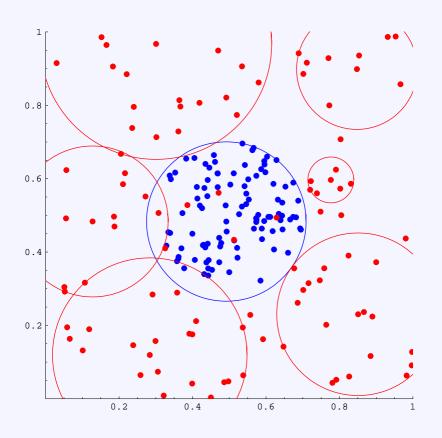
$$r^* = argmax_{r \ge 0} \{RW(r) - f(r)\}$$



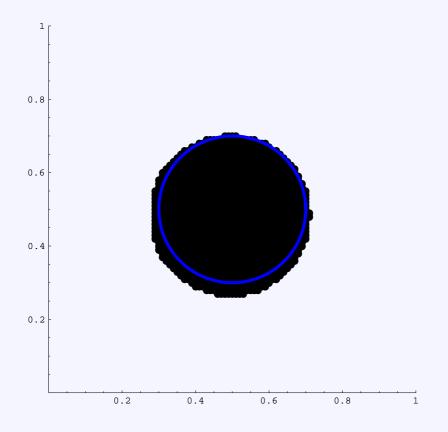


We want to chose the cover that is made up of the "best" or highest scoring covering balls instead of the fewest number of balls.

We want to chose the cover that is made up of the "best" or highest scoring covering balls instead of the fewest number of balls.



We want to chose the cover that is made up of the "best" or highest scoring covering balls instead of the fewest number of balls.



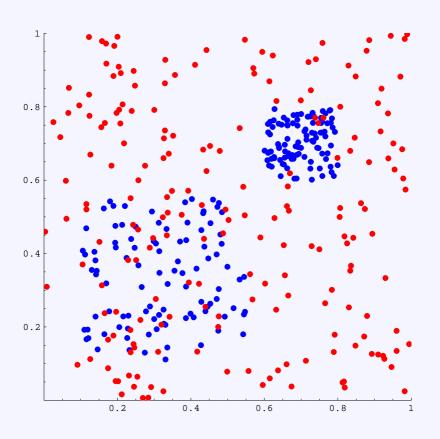
### **Random Walk CCP**

#### Benefits of Random Walk.

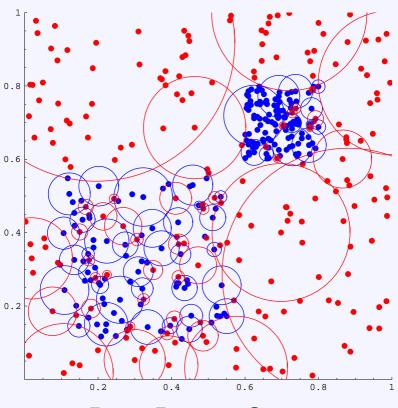
- Same motivation as  $\alpha$ ,  $\beta$  version, but adaptively chooses the  $\alpha$ ,  $\beta$  parameters for each ball.
- Each covering ball for a class lies approximately in the discriminant region of that class.
- Covers are less complex than pre-classifier and  $\alpha$ ,  $\beta$  classifiers. For Disk data,
  - Pre-classifier  $\Gamma_X = 10, \Gamma_Y = 12$
  - $\bullet$   $\alpha, \beta$  classifier  $\Gamma_X = 1, \Gamma_Y = 8$
  - Random Walk Classifier  $\Gamma_X = 1, \Gamma_Y = 6$

### **Simulation: Data**

Red Points  $\sim U([0,1] \times [0,1])$ . Blue Points  $\sim \frac{1}{2} U([0.1,0.55] \times [0.1,0.55]) + \frac{1}{2} U([0.6,0.8] \times [0.6,0.8])$ 



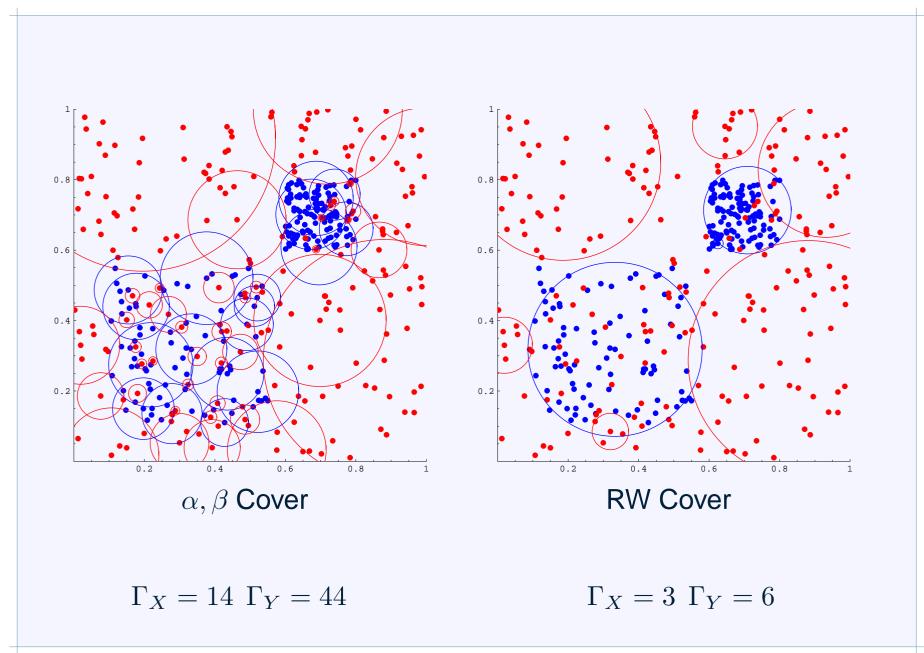
## **Simulation: Covers**

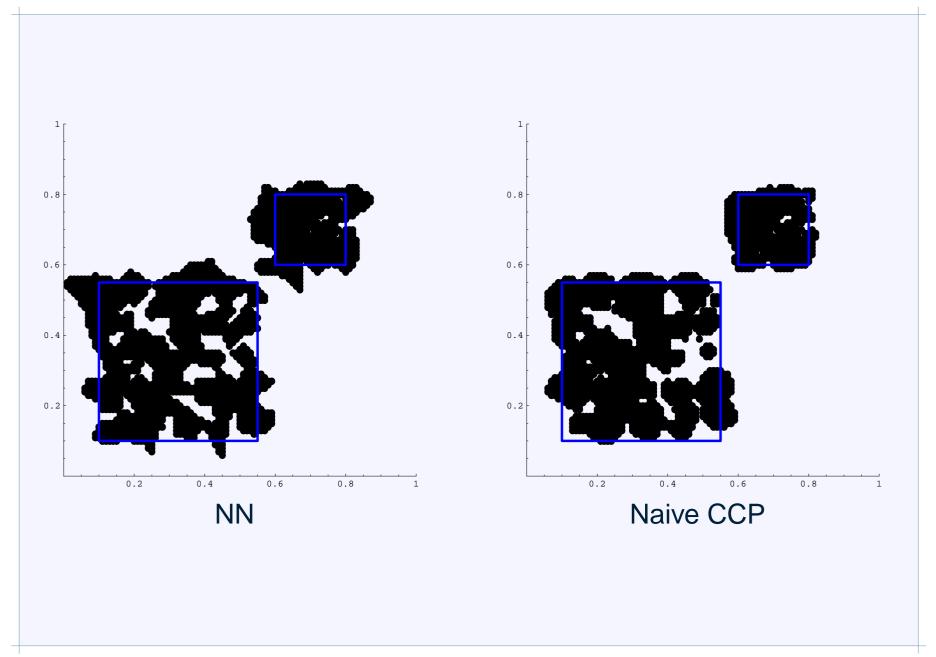


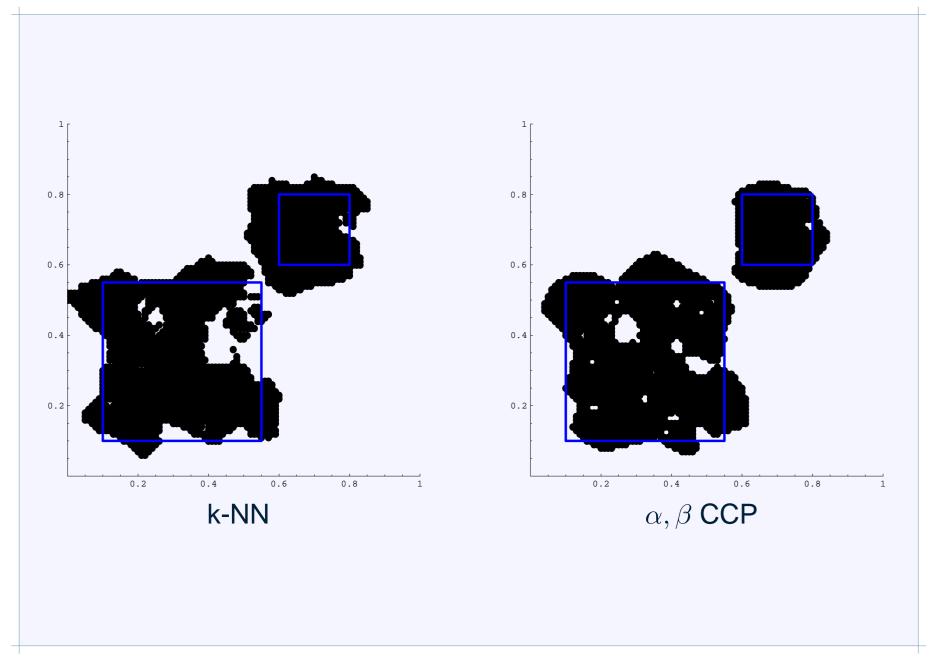
Pure Proper Cover

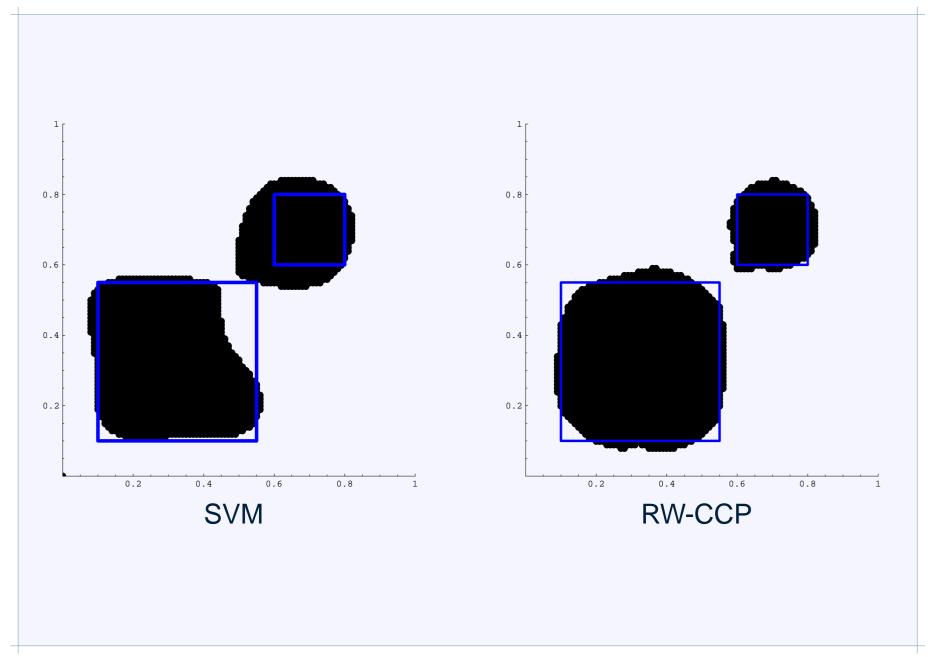
$$\Gamma_X = 46 \ \Gamma_Y = 44$$

## **Simulation: Covers**







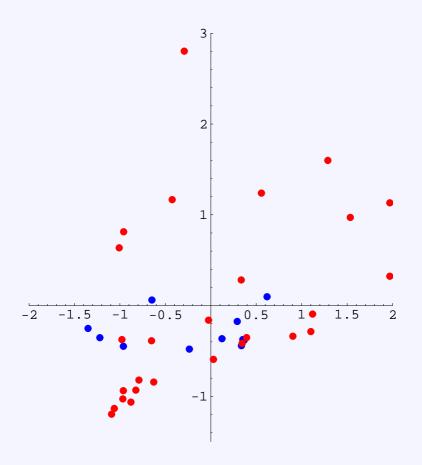


Here are the estimates for the misclassification rate  $\hat{L}$  as observed after 1000 trials.

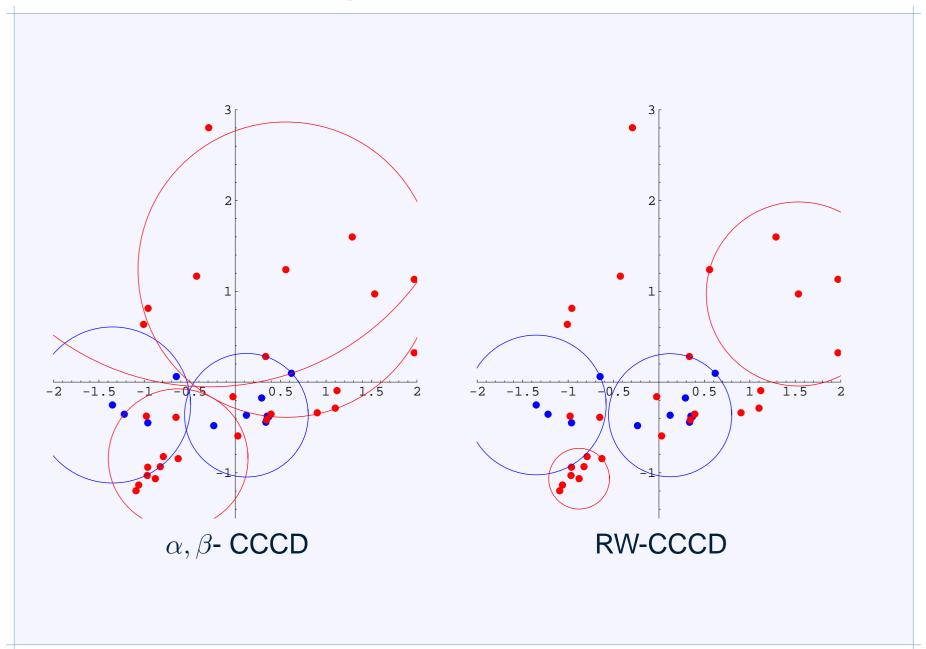
Tr. Size	NN	k-NN	SVM	Nv. CCP	$\alpha, \beta$ CCP	RW-CCP
50	0.242	.0.240	0.201	0.228	0.212	0.212
100	0.224	0.212	0.184	0.213	0.190	0.183
200	0.210	0.188	0.168	0.201	0.171	0.165
500	0.199	0.166	0.152	0.194	0.154	0.153

## **Experiment: Minefield Data**

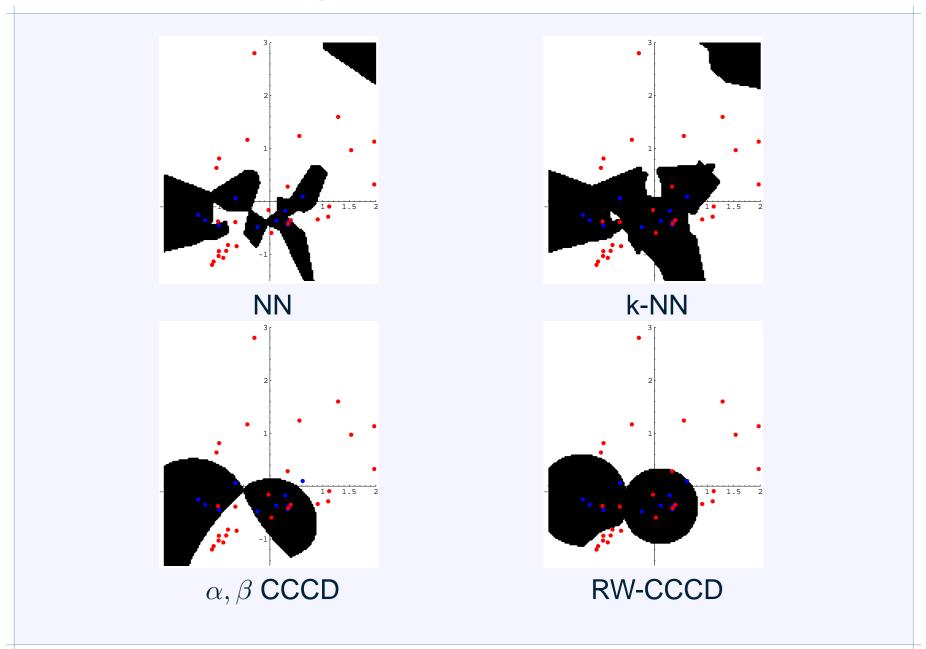
39 observations of multispectral images in a minefield.



## **Experiment: Covers**



# **Experiment: Classifiers**



#### The End

CCCD home page: http://www.mts.jhu.edu/~devinney/cccd

#### Acknowledgements:

- Carey Priebe Math/Sci Dept. Johns Hopkins U.
- Dave Marchette Naval Surface Warfare Center, Dalghren, VA.
- Diego Socolinsky Johns Hopkins U./ Equinox Inc.
- John Wierman Math/Sci Dept. Johns Hopkins U.
- Ed Scheinerman Math/Sci Dept. Johns Hopkins U.

SVM results courtesy of SVM\_light by T. Joachims (http://www-ai.cs.uni-dortmund.de/svm\_light)

Presentation prepared using LaTEX and Prosper.