Repeated Play of the SVM Game as a Means of Adaptive Classification

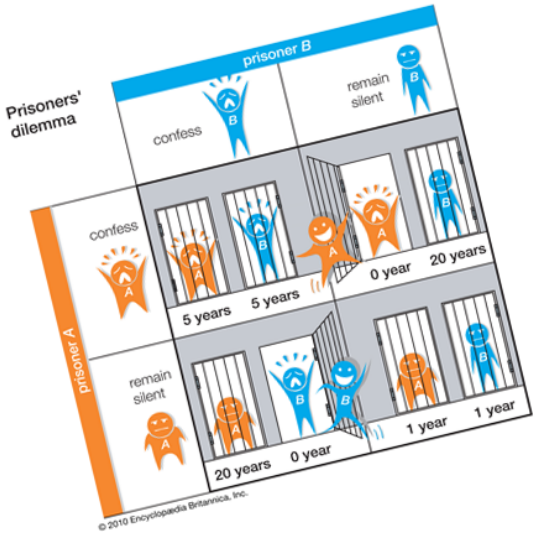
SAND2015-6422C

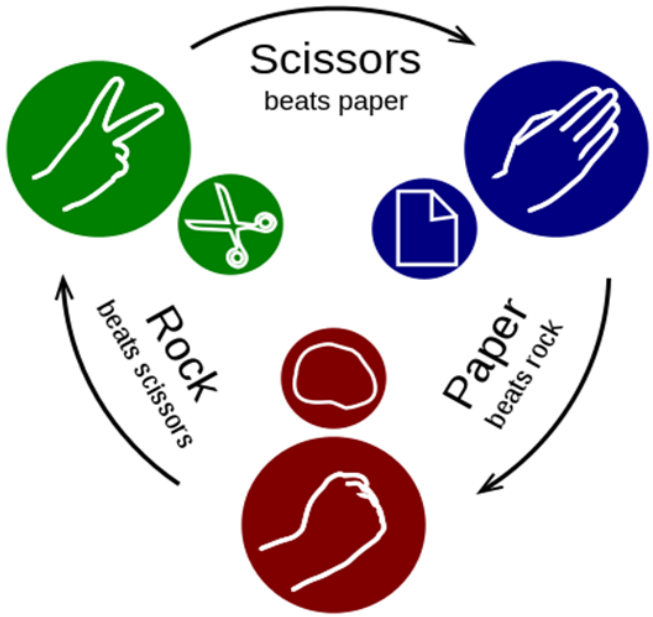
Craig M. Vineyard

Sandia National Laboratories

Game Theory

- Game theory is a branch of applied mathematics to formally analyze the strategic interaction between competing players





Algorithmic Game Theory

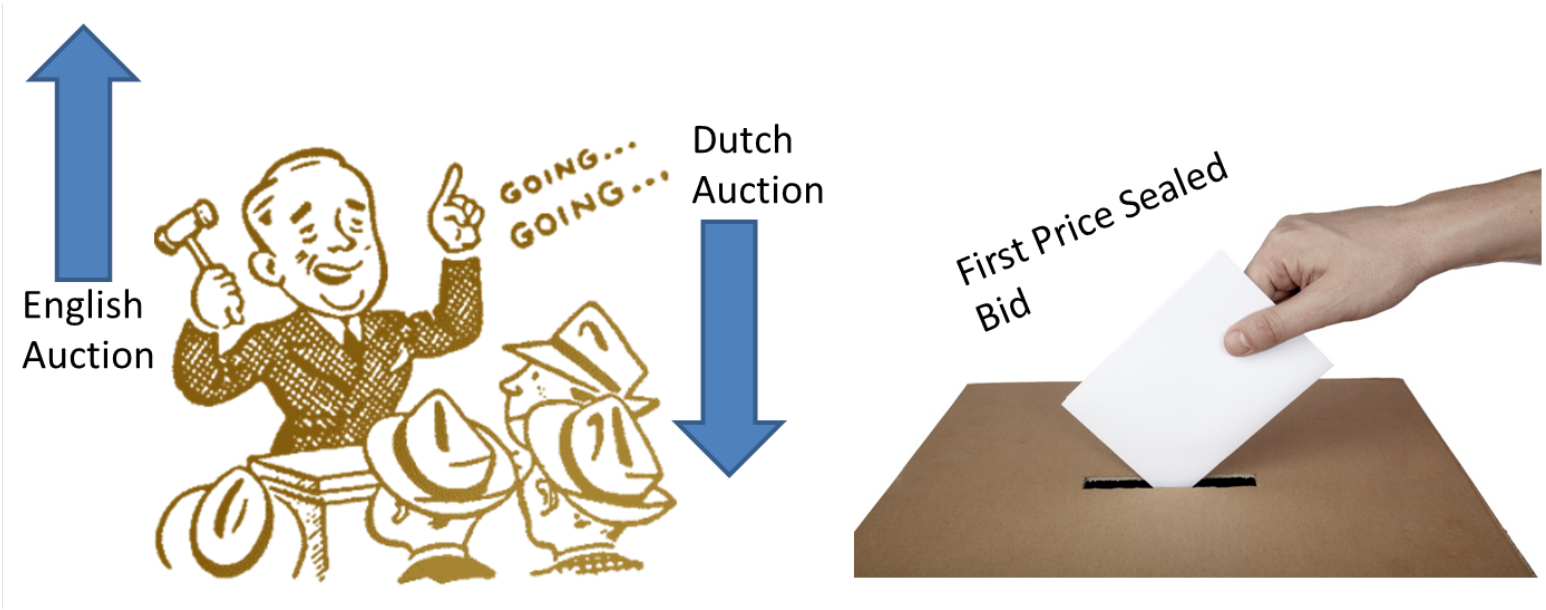
- Algorithmic Game Theory is the intersection of game theory and computer science research
  - Nissan & Ronan
- 2 Perspective:
  - Analysis - analyzes algorithms from game-theoretic perspective, focus on properties such as equilibria
  - Design - focuses upon development of algorithms with desirable theoretical properties
- Research areas: equilibria analysis, multi-agent systems, routing, algorithmic mechanism design

Mechanism Design

- Game theory primarily analyses the strategic actions between players, focusing upon concepts such as strategies and equilibria
- Mechanism design is a sub-field of game theory that takes a different perspective
  - With a desired outcome or goal in mind, mechanism design seeks to develop a framework defining player actions and the effect of these actions to try and attain the desired goal

Mechanism Design

- Commonly, a social choice function implements the desired outcome, operating upon player actions or preferences
  - Social choice theory studies collective decision making - how should we aggregate the preferences of the members of a group to obtain a social preference
  - i.e. Designing an auction and deciding if the bidding mechanism is first-price sealed-bid, vickrey, english, etc.

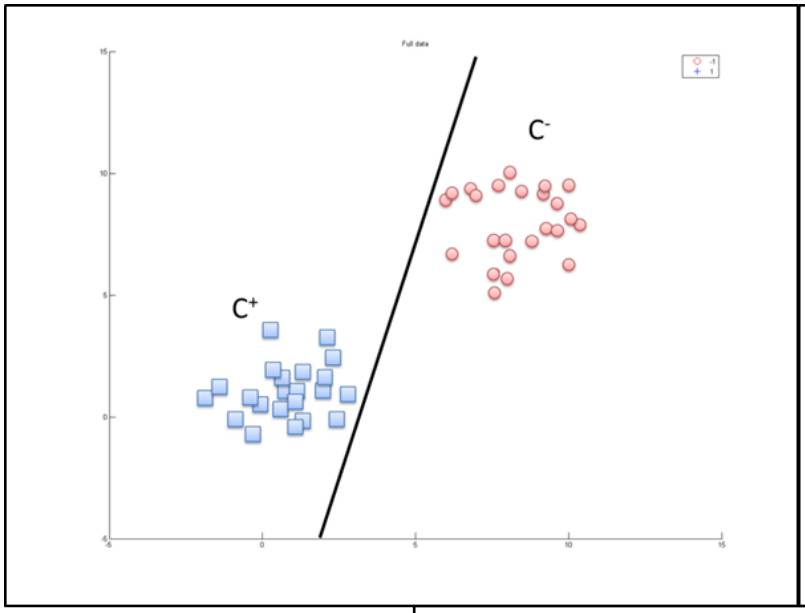


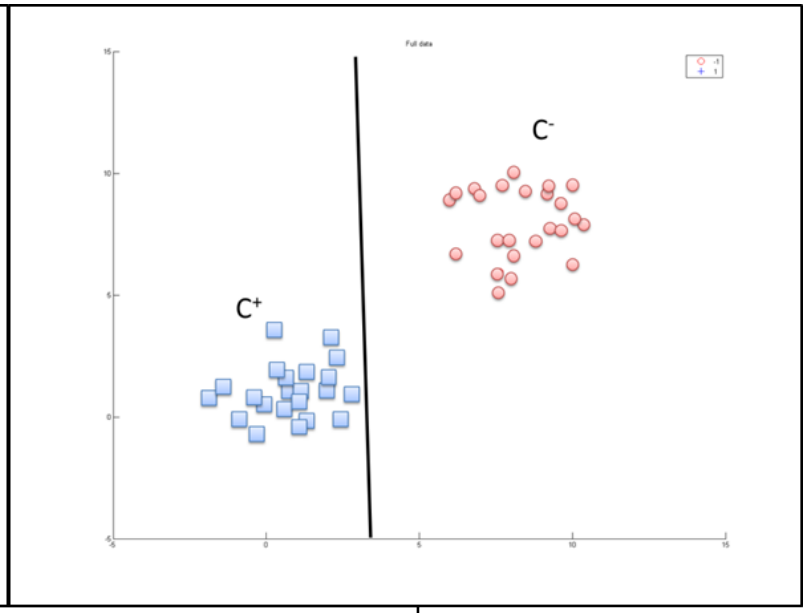
Mechanism Design Applied to Machine Learning

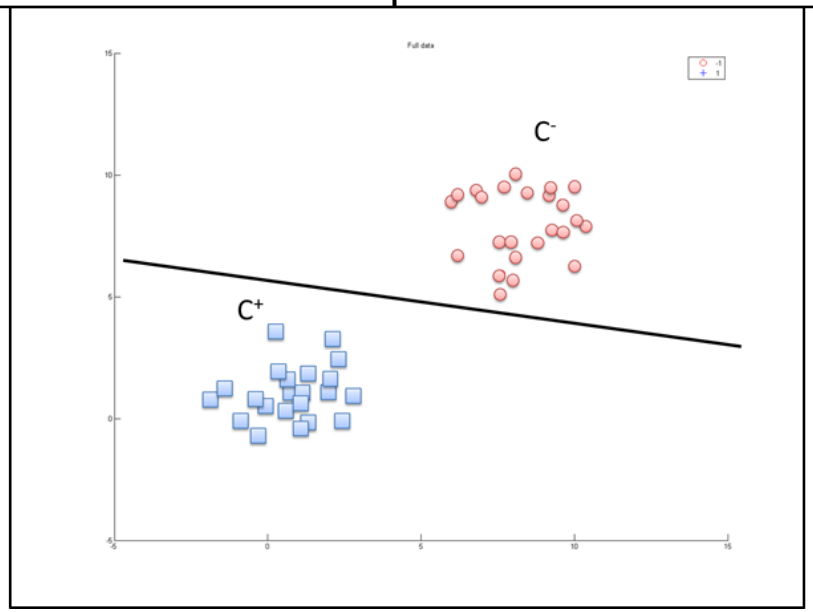
Two-class pattern classification problem

- A sample S of known example patterns are supplied from the classes $\mathcal{C}^+$ and $\mathcal{C}^-$,
- Based on S, one must produce a learning machine that will later classify unknown example patterns also drawn from $\mathcal{C}^+$ and $\mathcal{C}^-$
- Let $C^+ = \{c \in S \mid c \in \mathcal{C}^+\}$ and $C^- = \{c \in S \mid c \in \mathcal{C}^-\}$
  - Assume all example patterns are drawn from $\Re^n$
- From the perspective of designing a classifier (while inanimate objects like data points may not seem to have preferences), each and every data point innately imposes a constraint on the resulting discriminant
  - We want to aggregate representative features of data to design a good classifier

A good classifier discriminant?

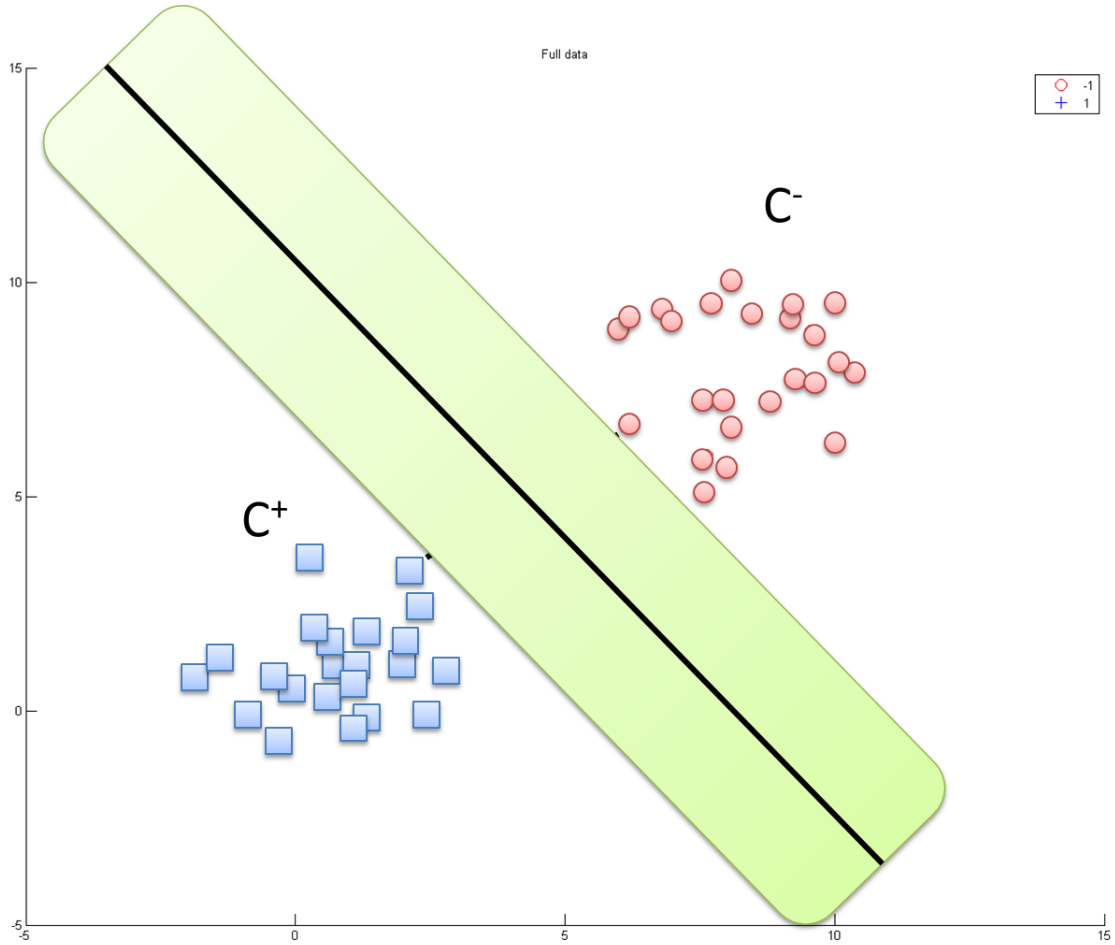






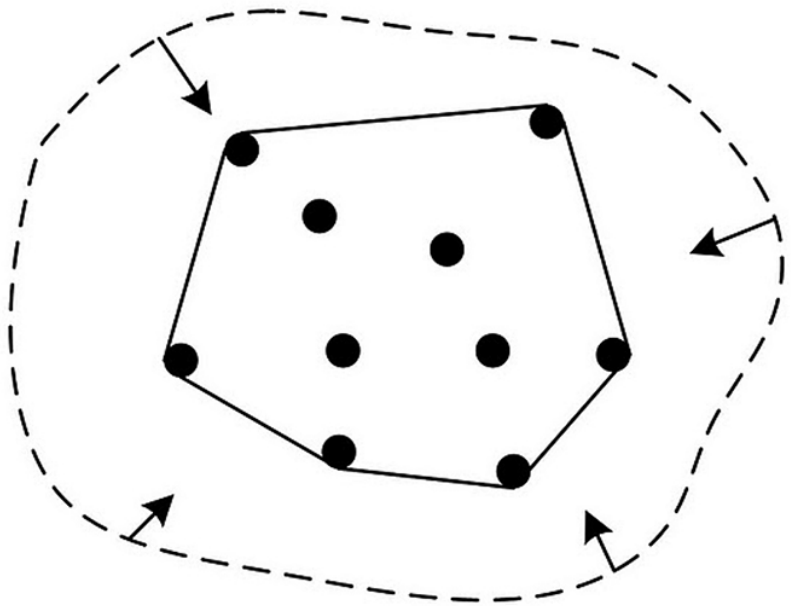
Support Vector Machine (SVM)

- In support vector machine (SVM) training, the goal is to find a separating hyperplane that maximizes the margin between the patterns in C^+ and C^- that serve as support vectors



Geometric SVM

- Geometric interpretation
  - Construct convex hulls around the sets C^+ and C^-
  - Find the closest points c^+ and c^- on each convex hull
  - Construct a line segment between these points and the discriminant is the perpendicular bisector of this line



Game Theoretic Classifier

The geometric SVM interpretation suggests a simple game:

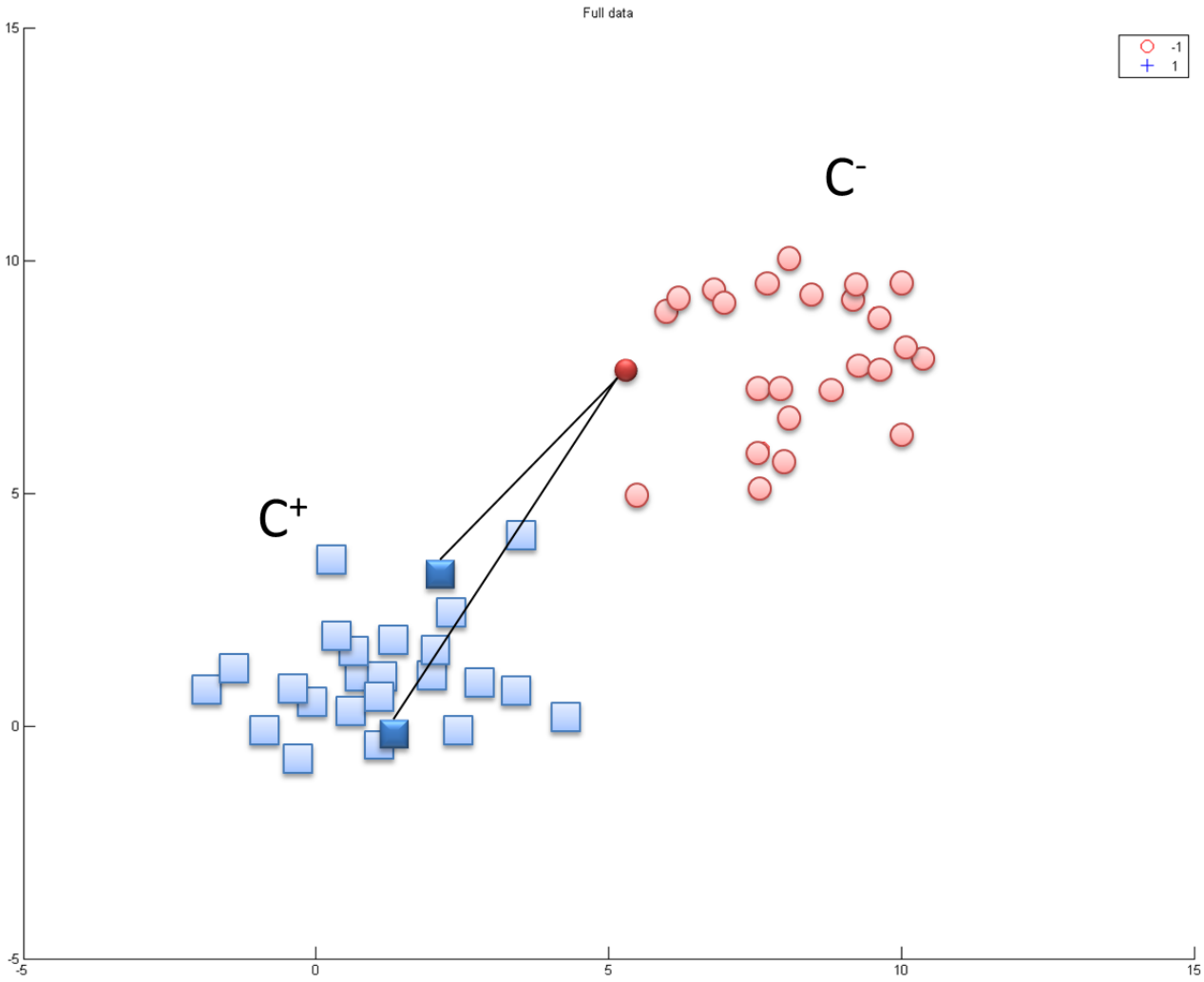
- This is a two player iterated game
  - Data patterns are the players
  - All players start with an initial equal quantity of α for each pattern class
- Each iteration of the game:
  - Randomly select two players from the same class and one data pattern from the opposing class
  - Each player has two possible actions, to pass (transferring some of its α to the other player), or to hold (keeping its own α)
  - A player must pass when it is further than its opponent from the other class data pattern
- As an iterated game, this same game procedure may be repeated for as many rounds as desired

Coalitional SVM Game

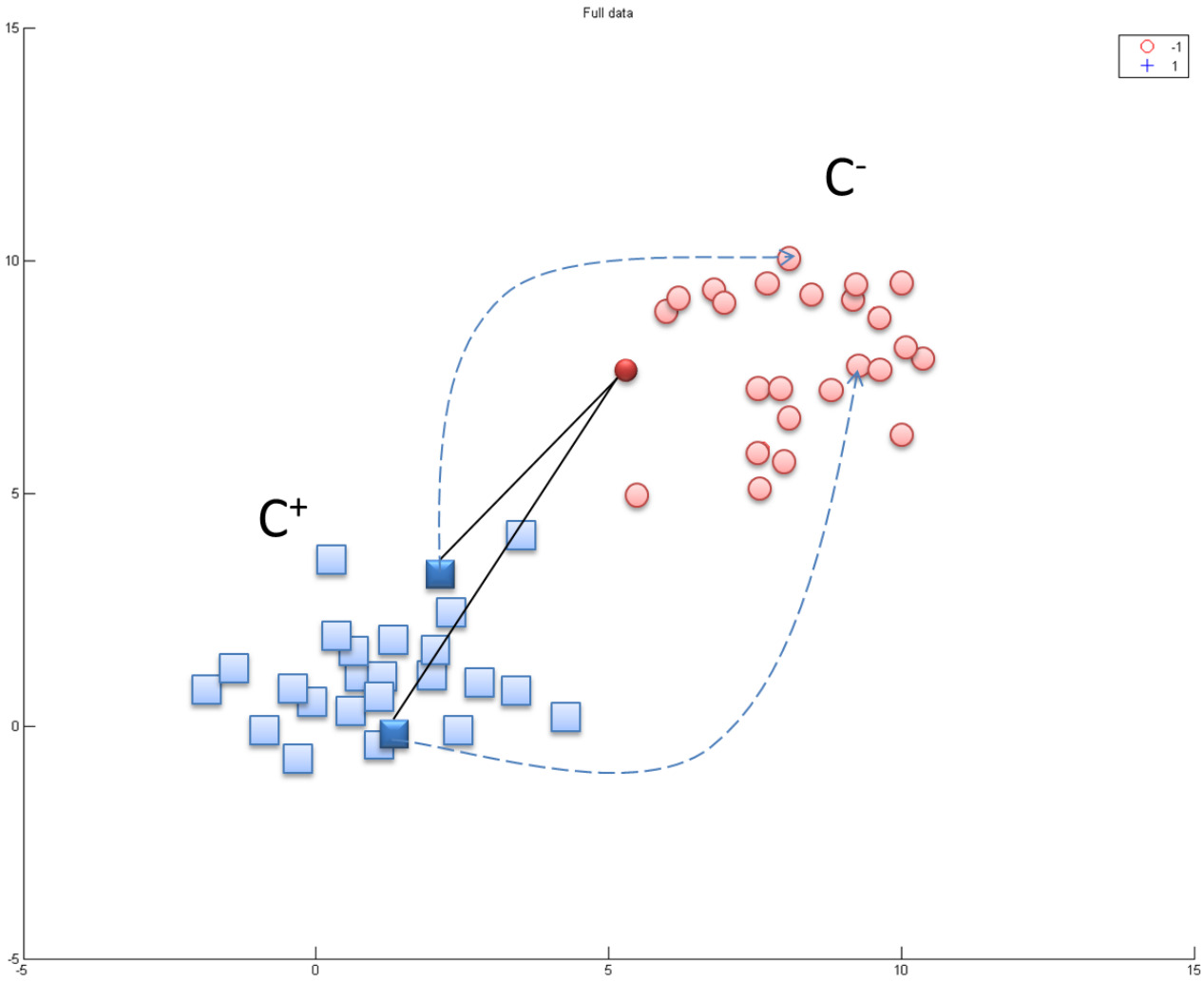
Adding coalitions to the basic SVM Game

- Each player (data point) has a coalition partner
  - Member of the opposing class believed to be closest to
- Many to one pairing
- Builds coalitions within a given class of the like-minded players who all agree upon the preferred (closest) player of the opposing class
- Each iteration of the Coalitional SVM, interacting players consider their relative distance to both the reference point from the opposing class as well as their coalition partner
- Provably adds stability to the SVM Game

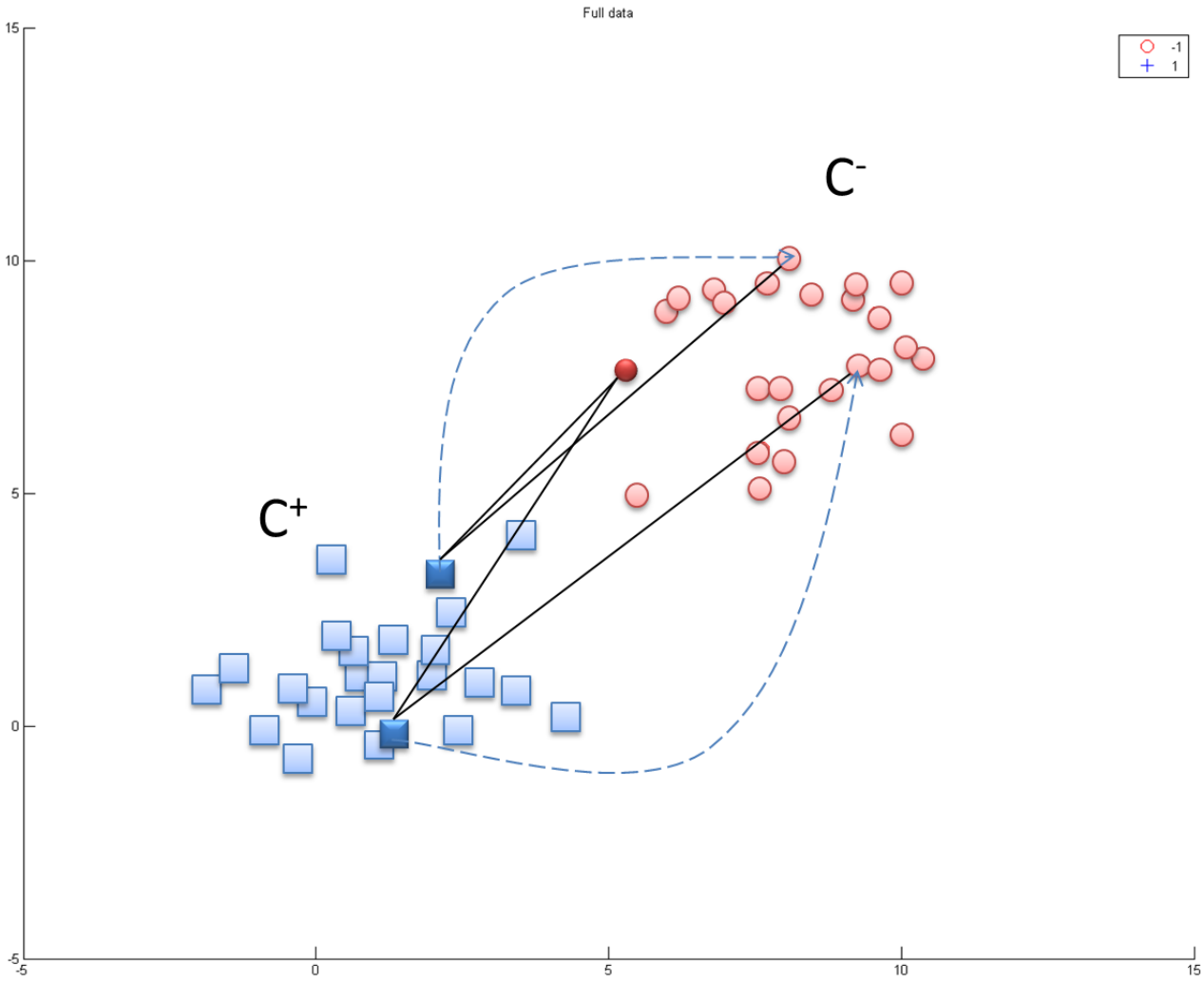
Coalitional SVM Game



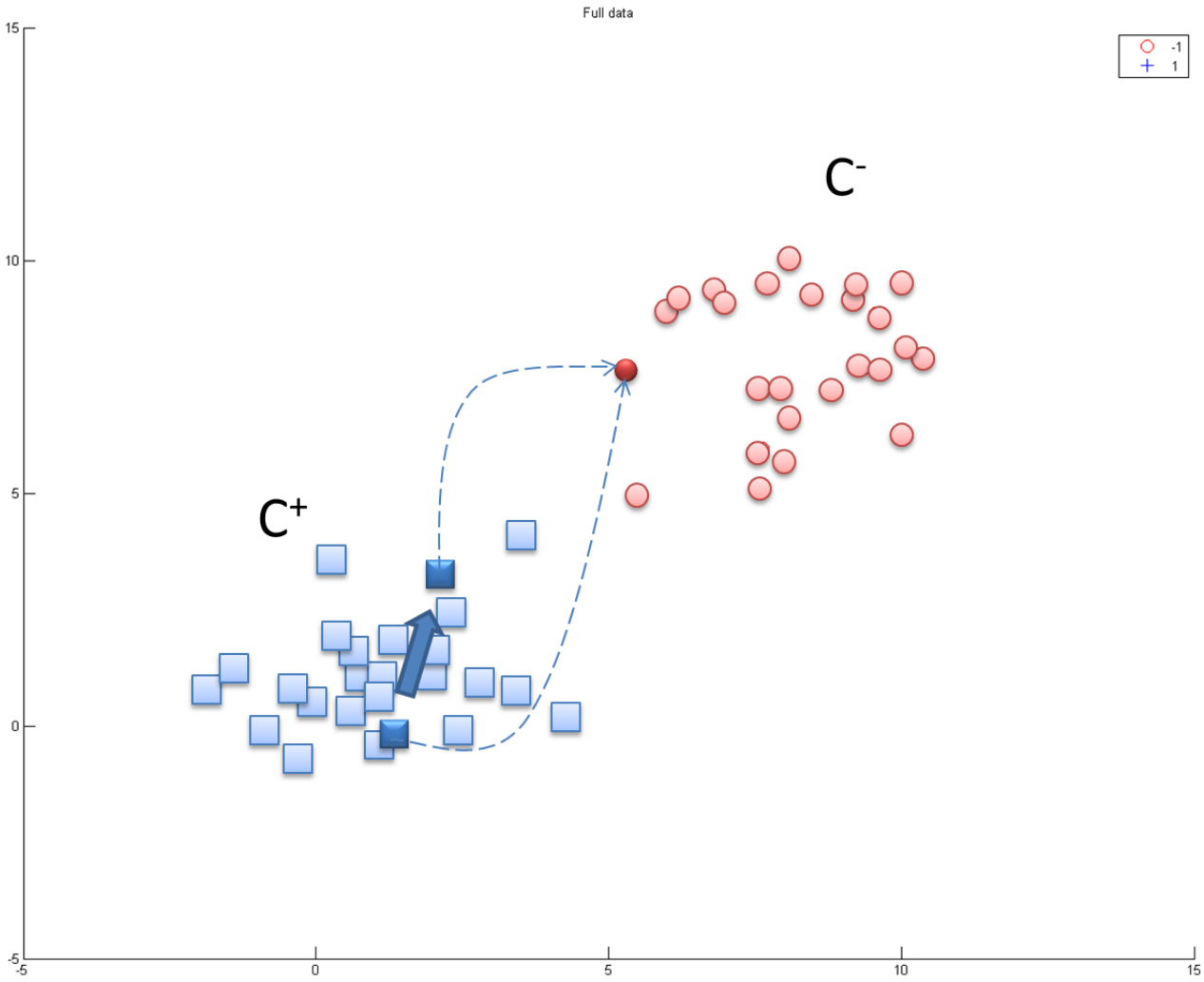
Coalitional SVM Game



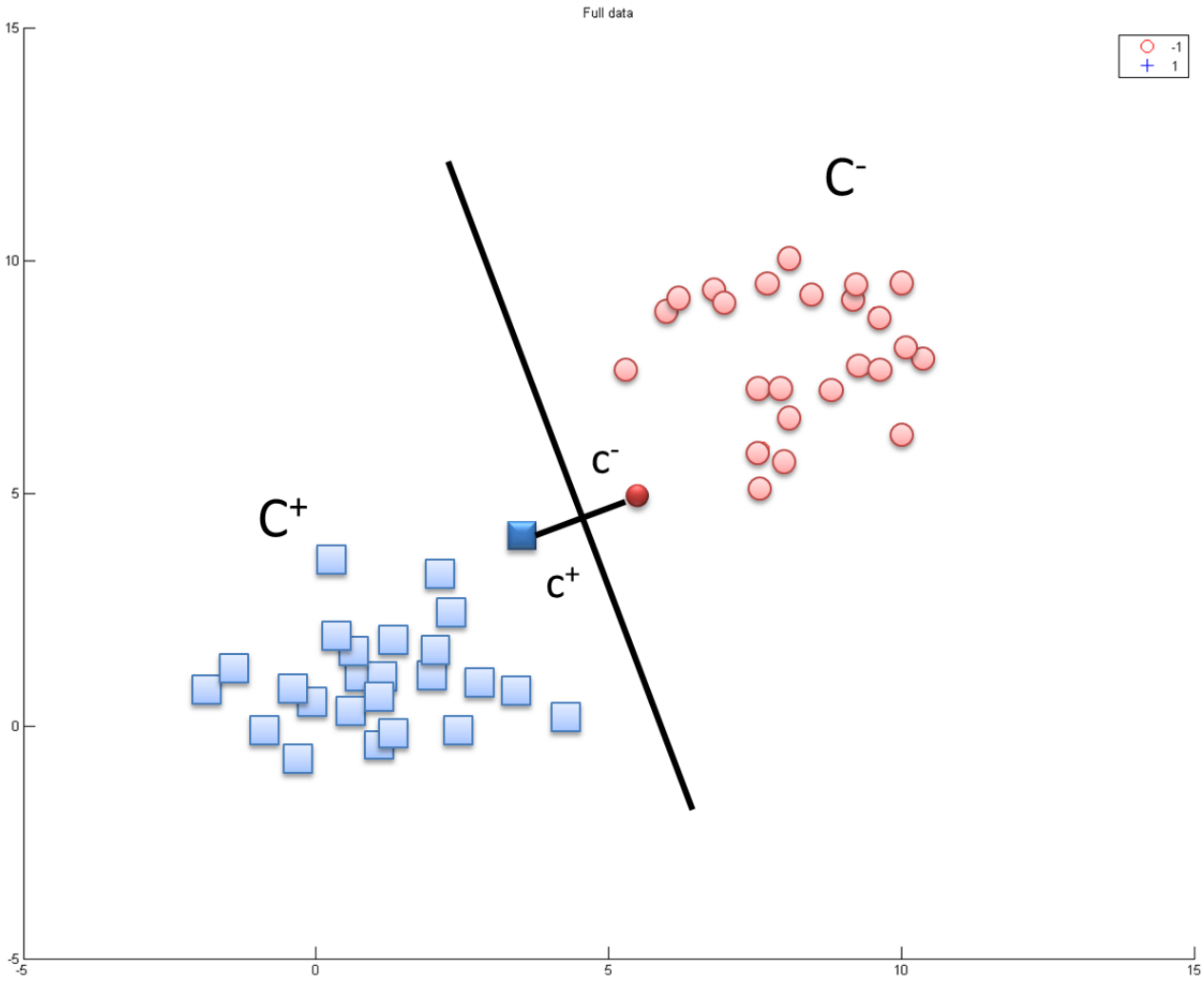
Coalitional SVM Game



Coalitional SVM Game



Coalitional SVM Game



Coalitional SVM Game Algorithm

Algorithm 3 Coalitional SVM Game

procedure SVM$_C(S)$ ▷ Sets $C^+, C^- \in S$
  for all $p_i \in S$ **do** ▷ Initialization
    $p_i \leftarrow$ Initial Coalition Parter from Opposite Class
  end for
  while iterations≠desired iterations **do**
    for each class C^+ and C^- **do**
      $p_x \leftarrow random(C^1)$
      $p_y \leftarrow random(C^1)$
      $p_r \leftarrow random(C^2)$
      Coalition$(p_x) \leftarrow \min(\mathrm{d}(p_x, p_r), \mathrm{d}(p_x, Coalitiona(p_x)), \mathrm{d}(p_x, Coalition(p_y)))$
      Coalition$(p_y) \leftarrow \min(\mathrm{d}(p_y, p_r), \mathrm{d}(p_y, Coalitiona(p_y)), \mathrm{d}(p_y, Coalition(p_x)))$
      if $\mathrm{d}(p_x, \text{Coalition}(p_x)) \leq \mathrm{d}(p_y, \text{Coalition}(p_y))$ **then**
        $p_x \leftarrow \alpha$ from p_y
      else
        $p_y \leftarrow \alpha$ from p_x
      end if
    end for
  end while
  return Sets A^+, A^-
    s.t. $\forall\, p_i \in A^+ \;\exists\, p_m \in C^-$ s.t. Coalition$(p_m) = p_i$, $p_i \in C^+$
    and
    s.t. $\forall\, p_j \in A^- \;\exists\, p_n \in C^+$ s.t. Coalition$(p_n) = p_j$, $p_j \in C^-$
end procedure

Non-linear Discriminant

- Smith Set: the smallest non-empty set of candidates in a particular election such that each member defeats every other candidate outside the set in a pairwise election

- When there is not a single unanimous Condorcet winner, the resulting partitions are the Smith Set
  - Each of these is a locally linear optimal region
  - A non-linear discriminant may be generated as a composition of these local segments
    - Provides a means of addressing non-linearly separable problems

Non-stationary Data

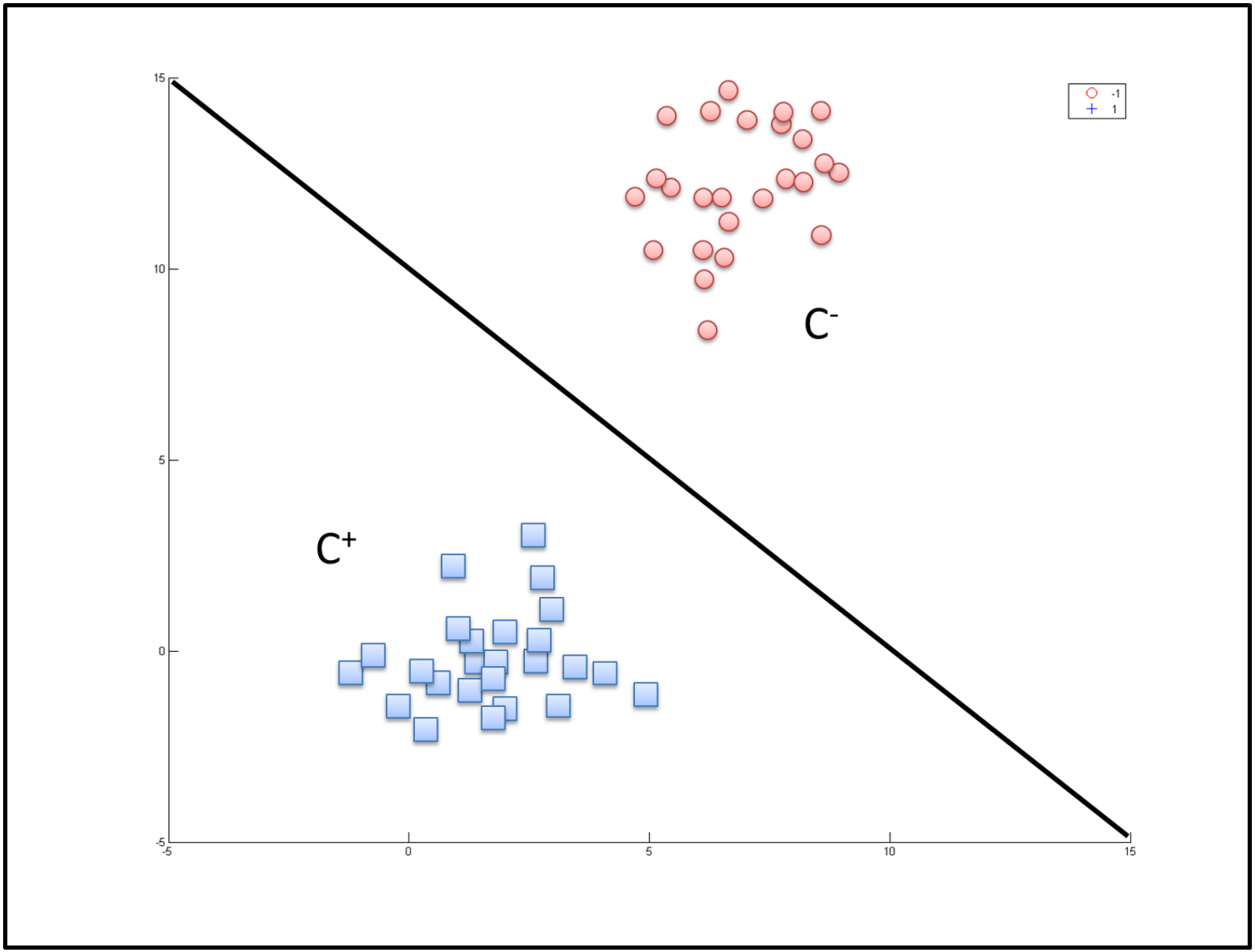
- Sometimes the data a machine learning algorithm is attempting to learn is Non-stationary $\rightarrow$ Concept Drift
- 2 categories of concept drift:
  - Virtual - changes in underlying data distribution drive need to update the model
  - Real - concepts themselves are changing
- However, it is not always feasible to generate a new machine learning model as data is updated or changes
  - Repeated play in game theory
- Repeated SVM Game
  - Game iterations are independent of one another
  - Able to update and continue learning

Repeated SVM Game Algorithm

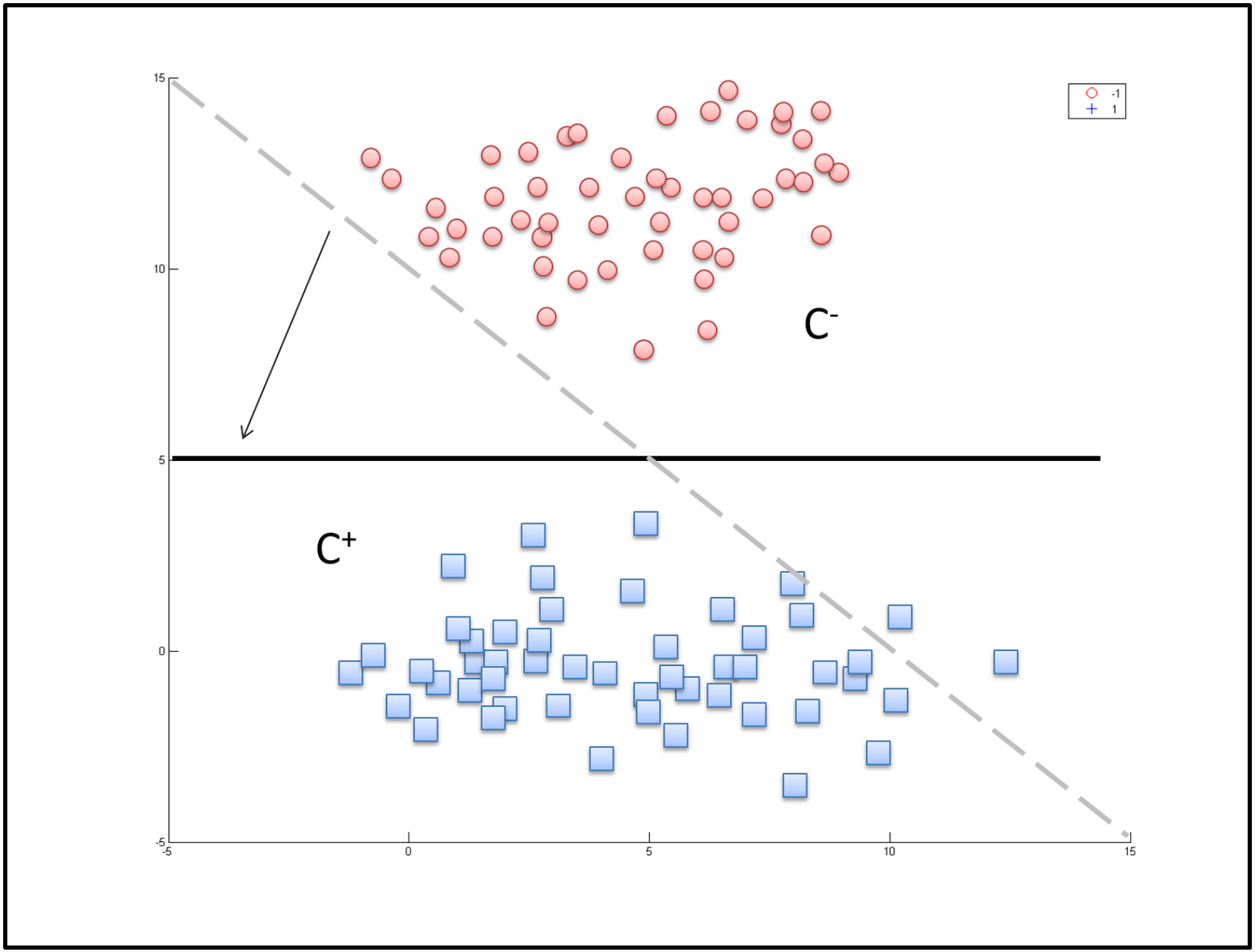
Algorithm 6 Repeated SVM Game

procedure SVM_GAME$_R(s_1, s_2, s_3, ...)$
▷ Subsets $s_1, s_2, s_3, ... \in S$ consisting of $C^+, C^- \in S$
  while Receiving s_i **do**
    $C^+ \leftarrow$ Update$(s_i \in C^+)$
    $C^- \leftarrow$ Update$(s_i \in C^-)$
    for new p_k **do** ▷ Initialization
      $p_k \leftarrow$ Initial Coalition Parter from Opposite Class
    end for
    while iterations≠desired iterations **do**
      for each class C^+ and C^- **do**
        $p_x \leftarrow random(C^1)$
        $p_y \leftarrow random(C^1)$
        $p_r \leftarrow random(C^2)$
        Coalition$(p_x) \leftarrow \min(\mathrm{d}(p_x, p_r), \mathrm{d}(p_x, Coalitiona(p_x)), \mathrm{d}(p_x, Coalition(p_y)))$
        Coalition$(p_y) \leftarrow \min(\mathrm{d}(p_y, p_r), \mathrm{d}(p_y, Coalitiona(p_y)), \mathrm{d}(p_y, Coalition(p_x)))$
        if $\mathrm{d}(p_x, \text{Coalition}(p_x)) \leq \mathrm{d}(p_y, \text{Coalition}(p_y))$ **then**
          $p_x \leftarrow \alpha$ from p_y
        else
          $p_y \leftarrow \alpha$ from p_x
        end if
      end for
    end while
    return Sets A^+, A^-
      s.t. $\forall\ p_i \in A^+\ \exists\ p_m \in C^-$ s.t. Coalition$(p_m) = p_i$, $p_i \in C^+$
      and
      s.t. $\forall\ p_j \in A^-\ \exists\ p_n \in C^+$ s.t. Coalition$(p_n) = p_j$, $p_j \in C^-$
  end while
end procedure

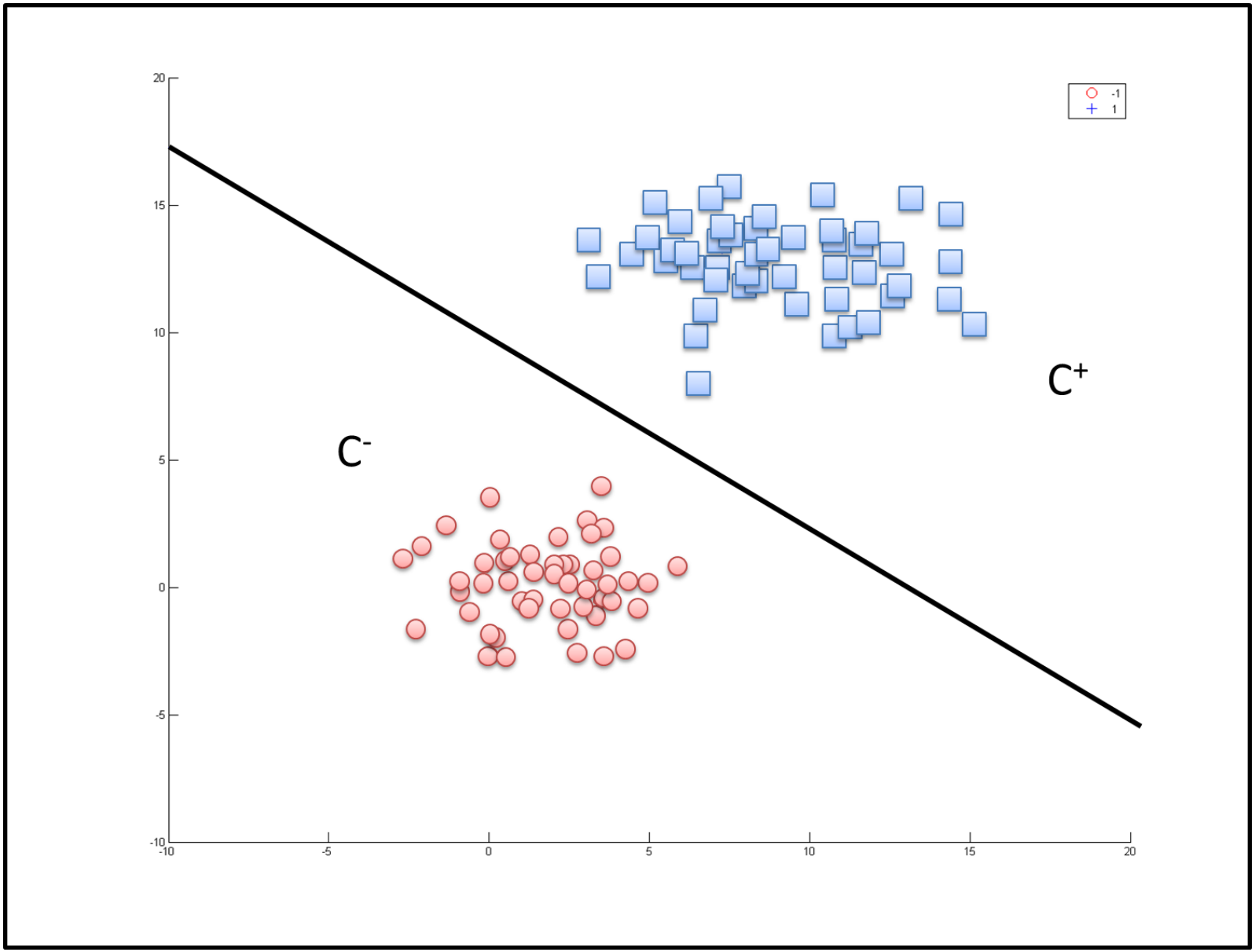
Repeated SVM Game - Linear Example



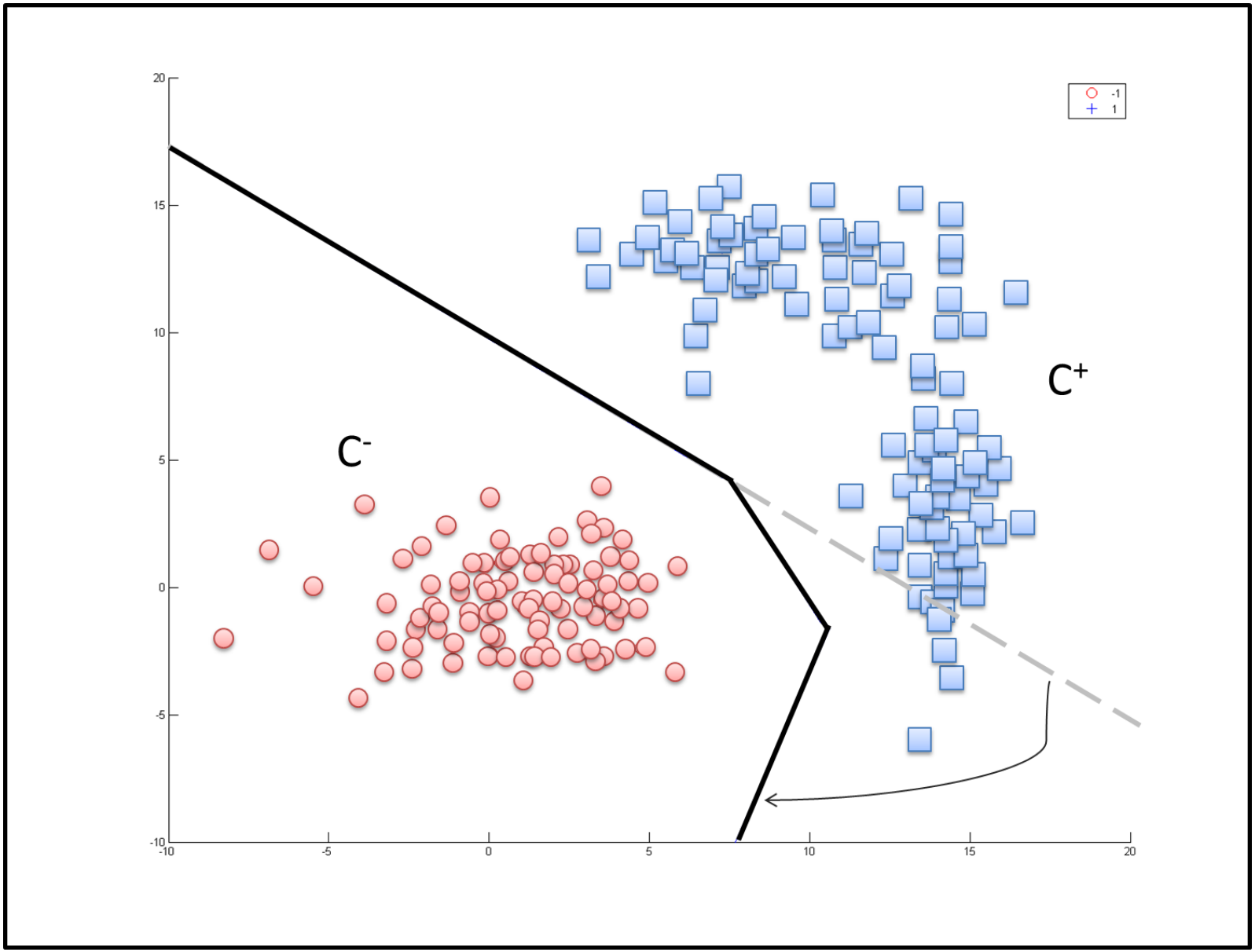
Repeated SVM Game - Linear Example



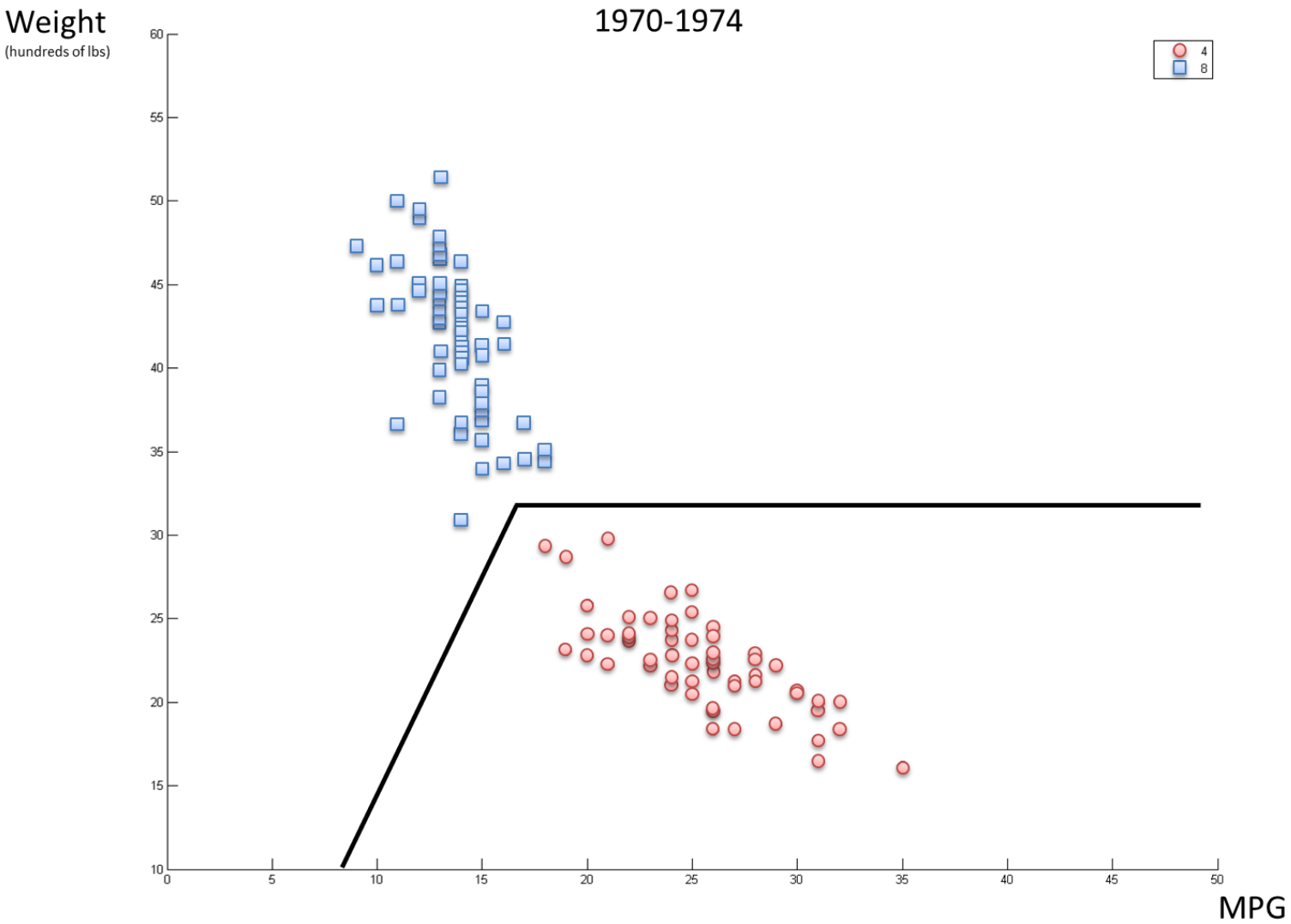
Repeated SVM Game - Curved Example



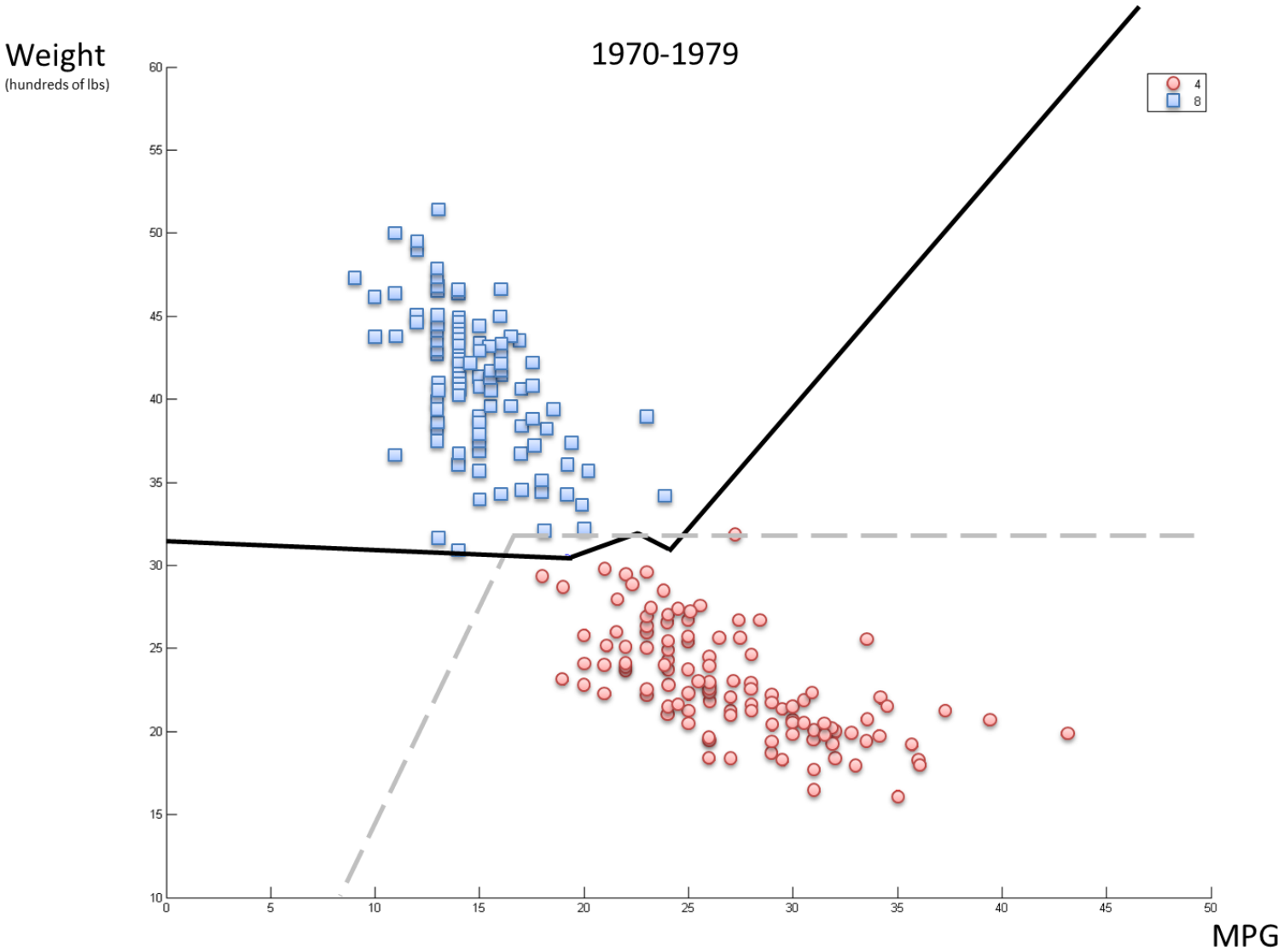
Repeated SVM Game - Curved Example



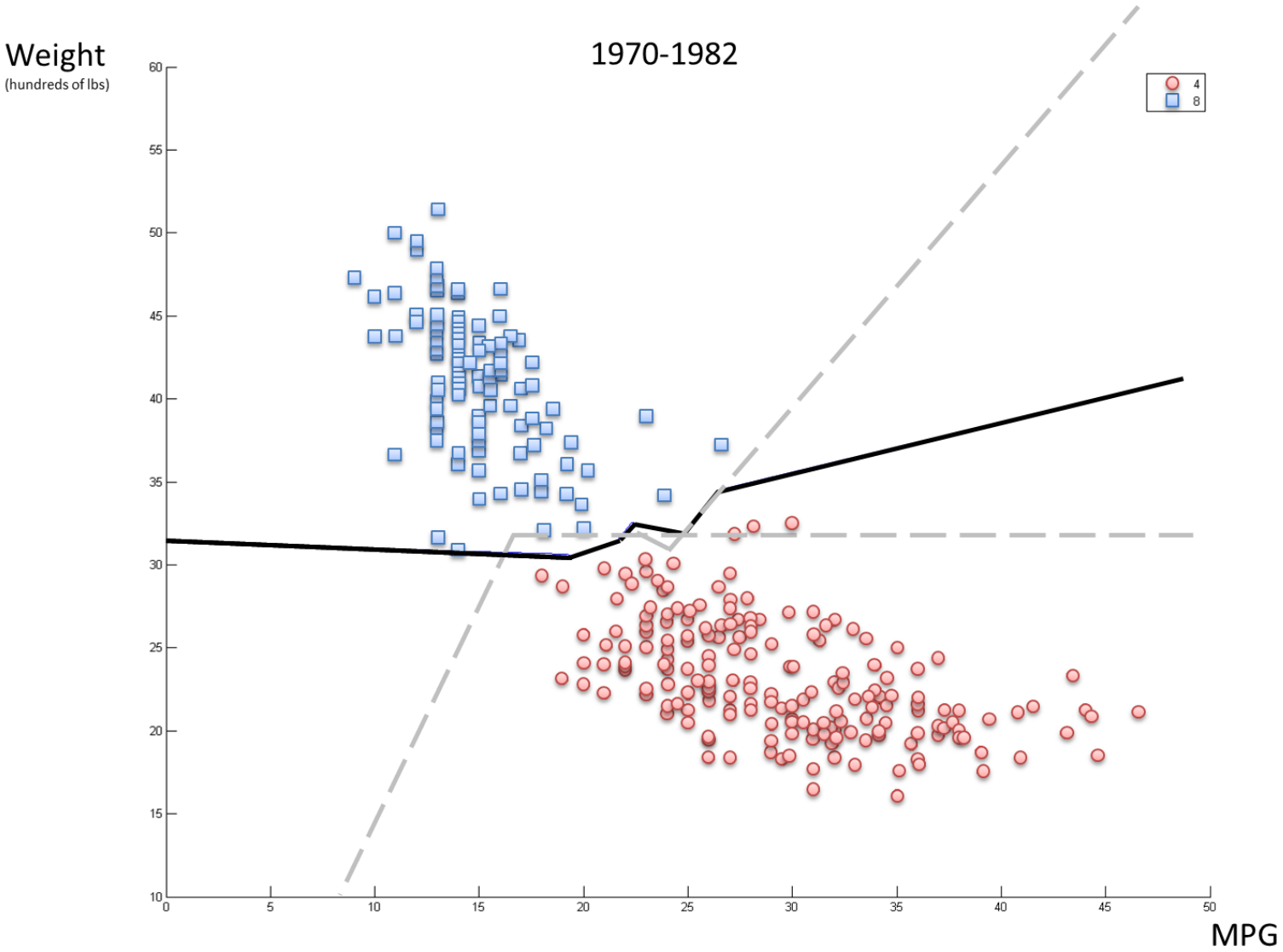
Repeated SVM Game - Automotive Example



Repeated SVM Game - Automotive Example



Repeated SVM Game - Automotive Example



Summary

- Presented the use of algorithmic game theory in relation to machine learning problem of classification
  - Descriptive as well as functional benefit
- SVM Game: game-theoretic Condorcet method classifier able to address linear, piecewise-linear, overlapping, and non-stationary data distributions through online learning
  - Distributed and Adaptive machine learning algorithms
- Repeated SVM Game: operating in an online learning and online training paradigm, through repeated game play is applicable for adaptive classification such as concept drift

Questions?

SVM Game Algorithm - with Preprocessing

Algorithm 4 Coalitional SVM Game

procedure SVM$_C(S)$ ▷ Sets $C^+, C^- \in S$
  Transform(S)
  for all $p_i \in S$ **do** ▷ Initialization
    $p_i \leftarrow$ Initial Coalition Parter from Opposite Class
  end for
  while iterations≠desired iterations **do**
    for each class C^+ and C^- **do**
      $p_x \leftarrow random(C^1)$
      $p_y \leftarrow random(C^1)$
      $p_r \leftarrow random(C^2)$
      Coalition$(p_x) \leftarrow \min(\mathrm{d}(p_x, p_r), \mathrm{d}(p_x, Coalitiona(p_x)), \mathrm{d}(p_x, Coalition(p_y)))$
      Coalition$(p_y) \leftarrow \min(\mathrm{d}(p_y, p_r), \mathrm{d}(p_y, Coalitiona(p_y)), \mathrm{d}(p_y, Coalition(p_x)))$
      if d$(p_x$,Coalition$(p_x)) \leq$ d$(p_y$,Coalition$(p_y))$ **then**
        $p_x \leftarrow \alpha$ from p_y
      else
        $p_y \leftarrow \alpha$ from p_x
      end if
    end for
  end while
  return Sets A^+, A^-
    s.t. $\forall\ p_i \in A^+\ \exists\ p_m \in C^-$ s.t. Coalition$(p_m) = p_i$, $p_i \in C^+$
    and
    s.t. $\forall\ p_j \in A^-\ \exists\ p_n \in C^+$ s.t. Coalition$(p_n) = p_j$, $p_j \in C^-$
end procedure

SVM Game Algorithm - with Kernel Trick

Algorithm 5 Coalitional SVM Game

procedure SVM_GAME$_C(S)$ ▷ Sets $C^+, C^- \in S$
  for all $p_i \in S$ **do** ▷ Initialization
    $p_i \leftarrow$ Initial Coalition Parter from Opposite Class
  end for
  while iterations$\neq$desired iterations **do**
    for each class C^+ and C^- **do**
      $p_x \leftarrow random(C^1)$
      $p_y \leftarrow random(C^1)$
      $p_r \leftarrow random(C^2)$
      $\text{Coalition}(p_x) \leftarrow \min(K(p_x, p_r), K(p_x, Coalitiona(p_x)), K(p_x, Coalition(p_y)))$
      $\text{Coalition}(p_y) \leftarrow \min(K(p_y, p_r), K(p_y, Coalitiona(p_y)), K(p_y, Coalition(p_x)))$
      if $\text{d}(p_x, \text{Coalition}(p_x)) \leq \text{d}(p_y, \text{Coalition}(p_y))$ **then**
        $p_x \leftarrow \alpha$ from p_y
      else
        $p_y \leftarrow \alpha$ from p_x
      end if
    end for
  end while
  return Sets A^+, A^-
    s.t. $\forall\, p_i \in A^+ \;\exists\, p_m \in C^-$ s.t. $\text{Coalition}(p_m) = p_i$, $p_i \in C^+$
    and
    s.t. $\forall\, p_j \in A^- \;\exists\, p_n \in C^+$ s.t. $\text{Coalition}(p_n) = p_j$, $p_j \in C^-$
end procedure

- This may be represented as a quadratic optimzation problem as follows:

$$\min_{\alpha} \quad \frac{1}{2}\left\|C^{+}\alpha^{+} - C^{-}v\right\|^{2} \tag{1}$$

$$\text{s.t.} \quad e^{T}u = 1,\ e^{T}v = 1,\ u \geq 0,\ v \geq 0, \tag{2}$$

where a linear discriminant x'w $= \gamma$ is constructed using the results of (1) and (2) by choosing

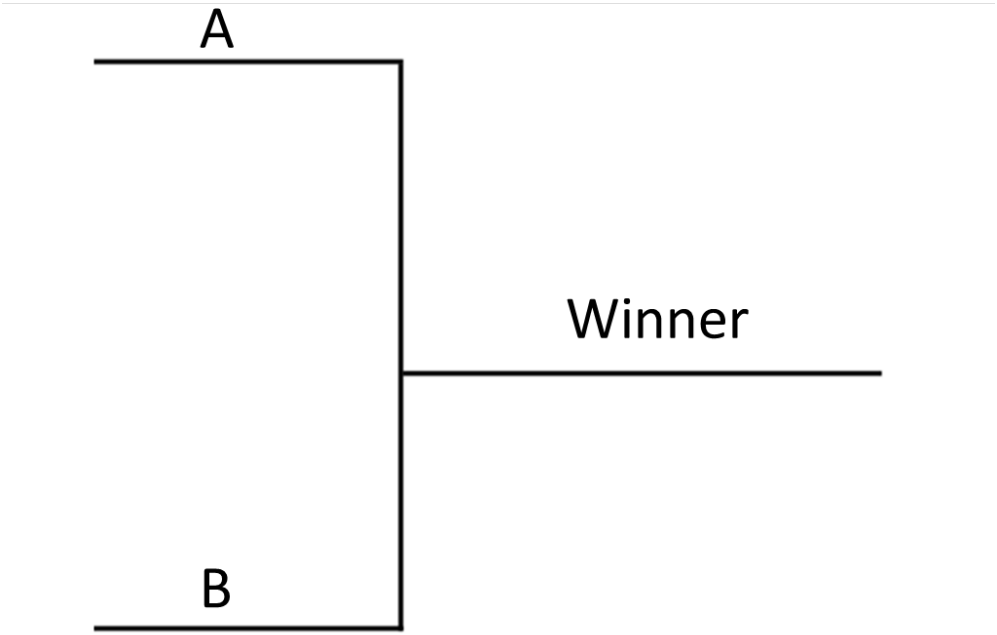
$$w = c^{+} - c^{-} = C^{+}\alpha^{+}{}_{opt} - C^{-}\alpha^{-}opt \tag{3}$$

and

$$\gamma = (\frac{c^{+} + c^{-}}{2})'w \tag{4}$$

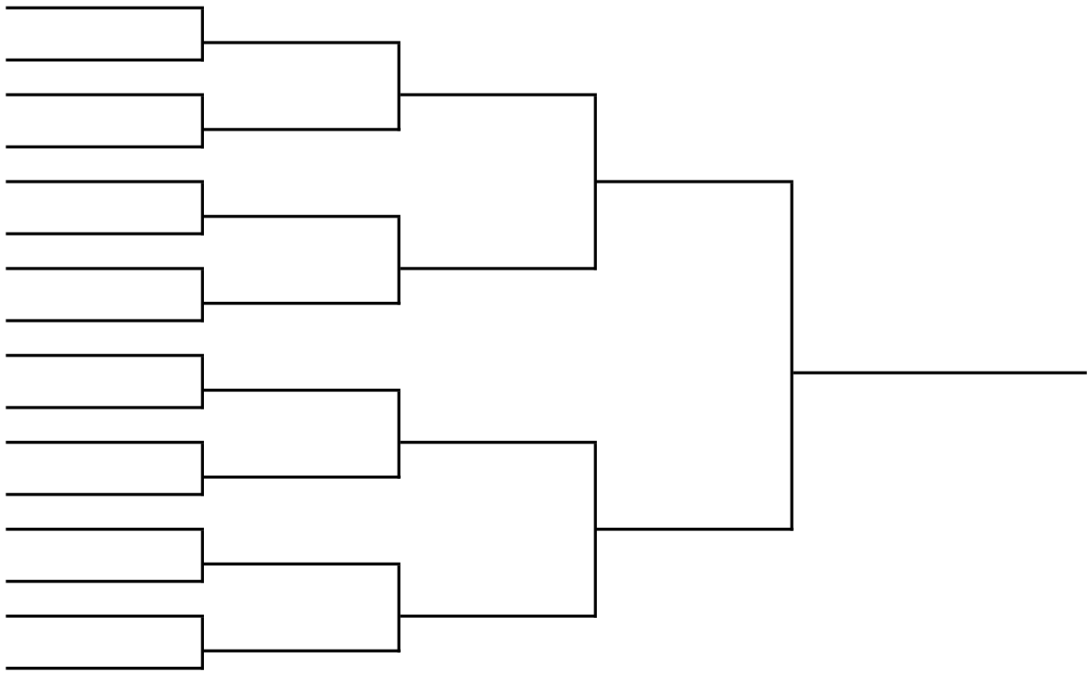
Social Choice Functions

- Selecting between two options or candidates it is straightforward to directly evaluate the two choices against one another



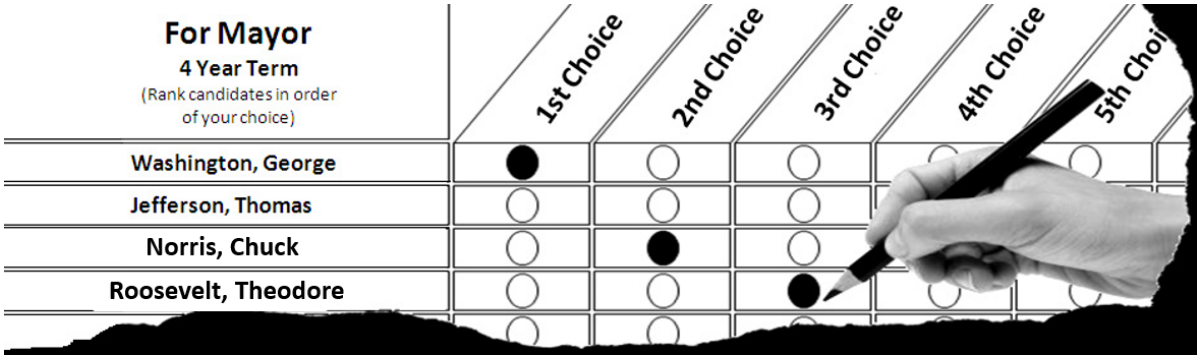
Social Choice Functions

- However, for three (or more choices) it is a more complicated matter



Social Choice Functions

- Various approaches have been developed to elect a single winner.
  - Most well known approach is to elect the candidate or option with the greatest amount of votes (plurality voting)
  - Majority vote
  - Other variants (such as Borda Count)



Condorcet Method

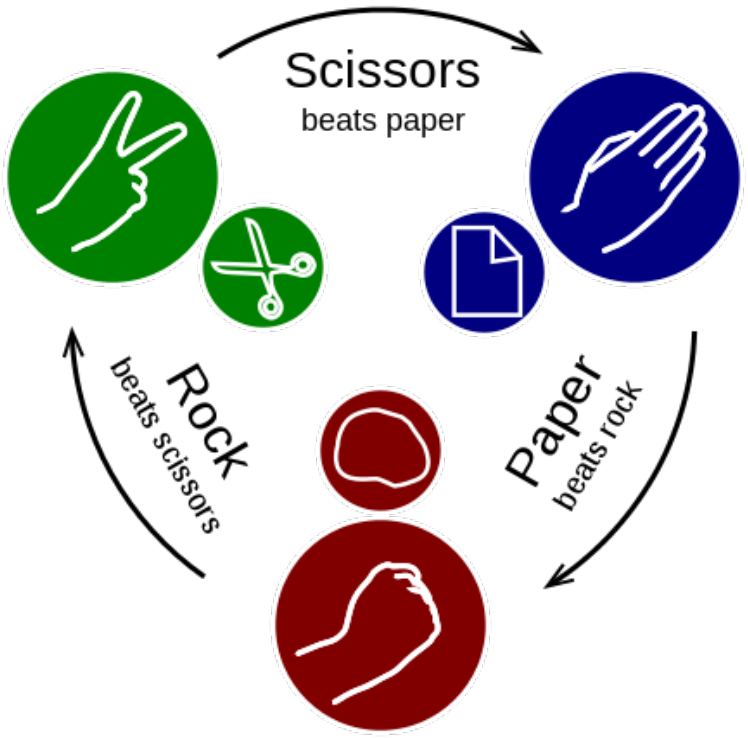
- Preferences of voters are ranked for all candidates
- Then all possible pairings of candidates are evaluated to determine the candidate preferred over all others
- For example:
  - voter 1 $x > y > z$
  - voter 2 $x > z > y$
  - voter 3 $z > x > y$
  - x vs. y $\rightarrow$ all three voters prefer x over y
  - y vs. z $\rightarrow$ two of the three voters prefer z over y (only voter 1 prefers y to z)
  - x vs. z $\rightarrow$ two of the three voters prefer x to z.
  - Thus, x beats every other candidate in pairwise competition and is the Condorcet winner

Convergence - Condorcet Paradox

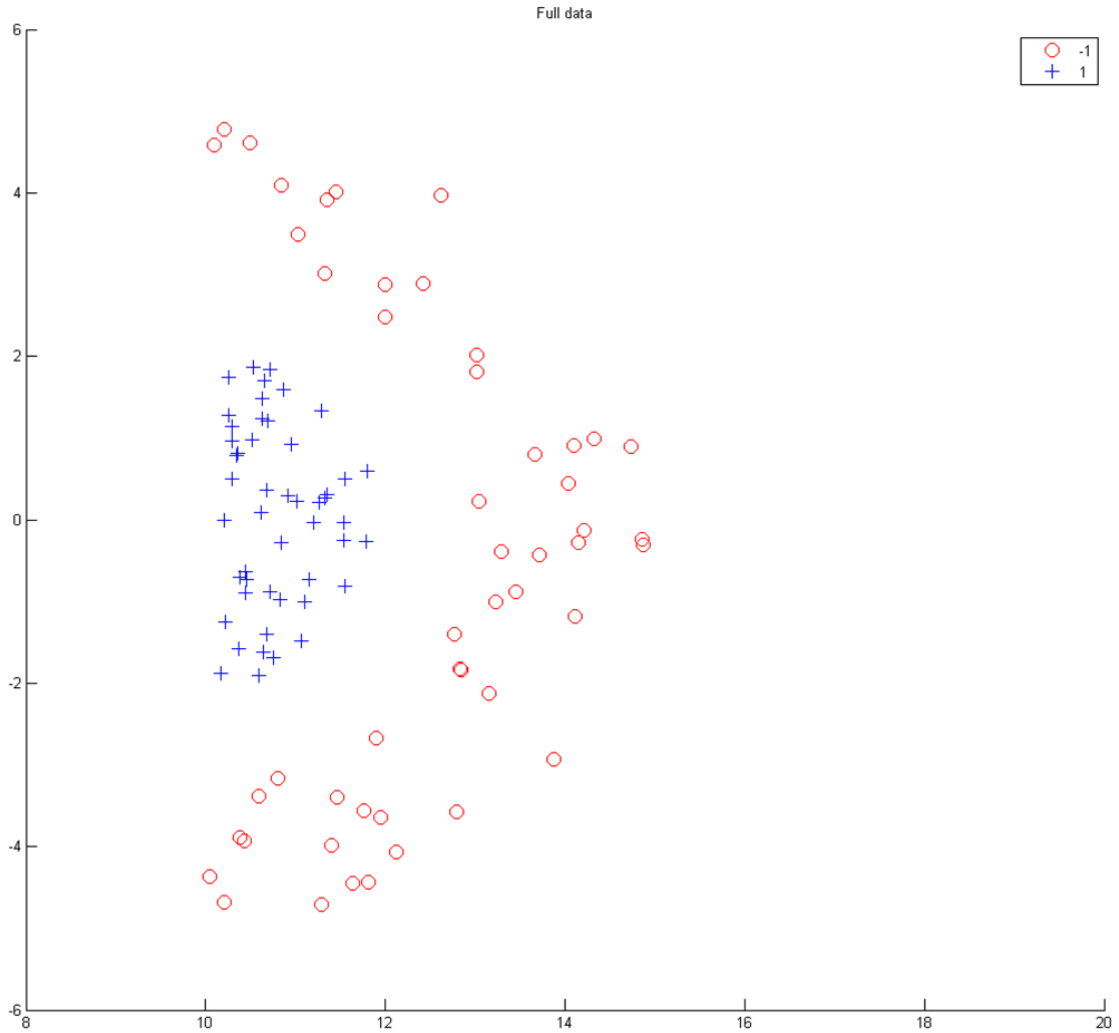
- There is not always a Condorcet winner
  - Condorcet Paradox - a non-transitive group preference can arise from transitive individual preferences
- Example:
  - voter 1 $x > y > z$
  - voter 2 $y > z > x$
  - voter 3 $z > x > y$
- Evaluating the pairwise comparisons amongst all possible candidates there is no longer a candidate which beats all others
  - x beats y according to two of the three voters
  - likewise y is preferred to z by two of the three voters
  - z is preferred to x by two of the three

Condorcet Paradox

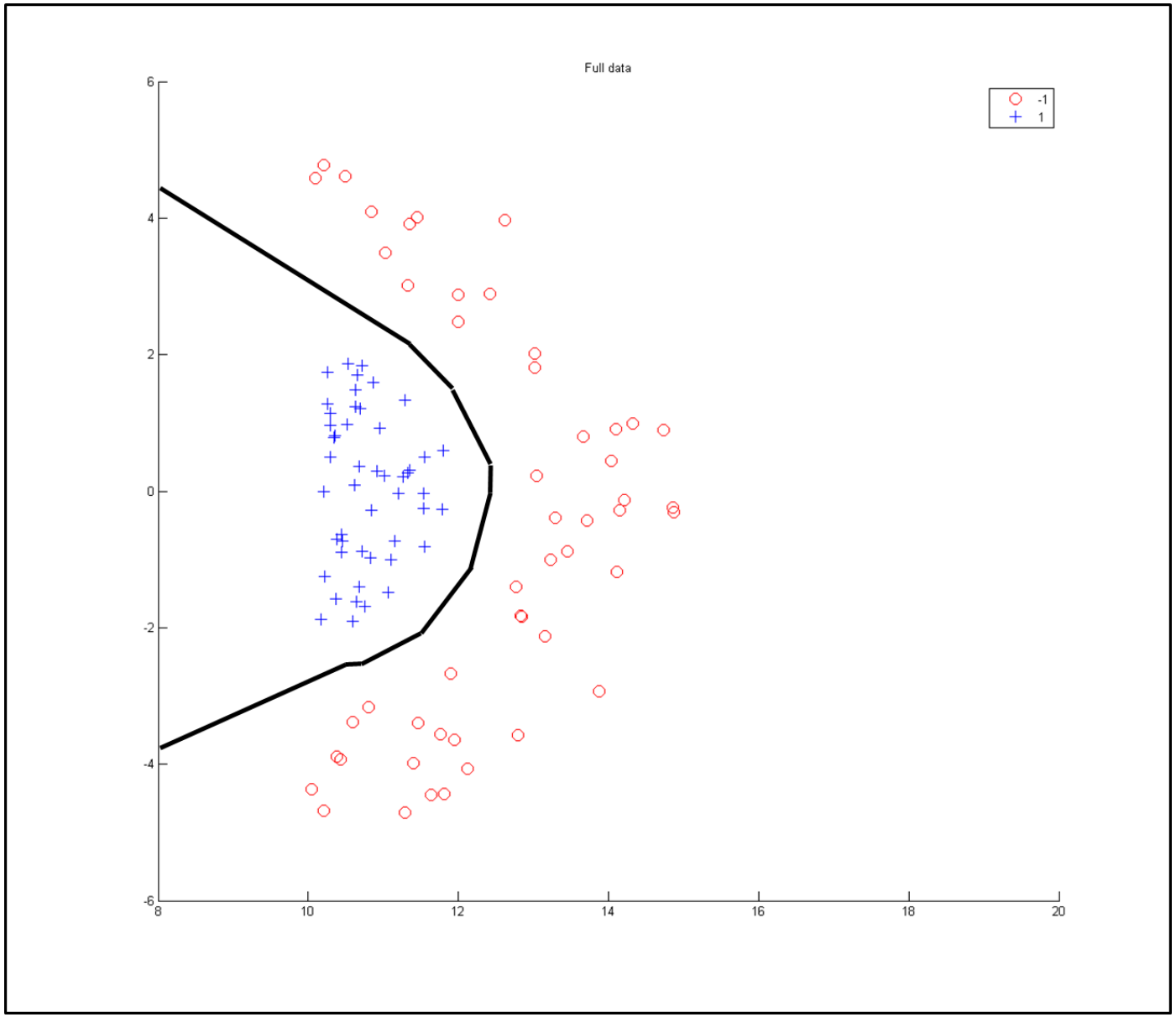
Rock Paper Scissors also demonstrates this notion that comparing all possible pairs does not necessarily provide a globally agreed upon winner



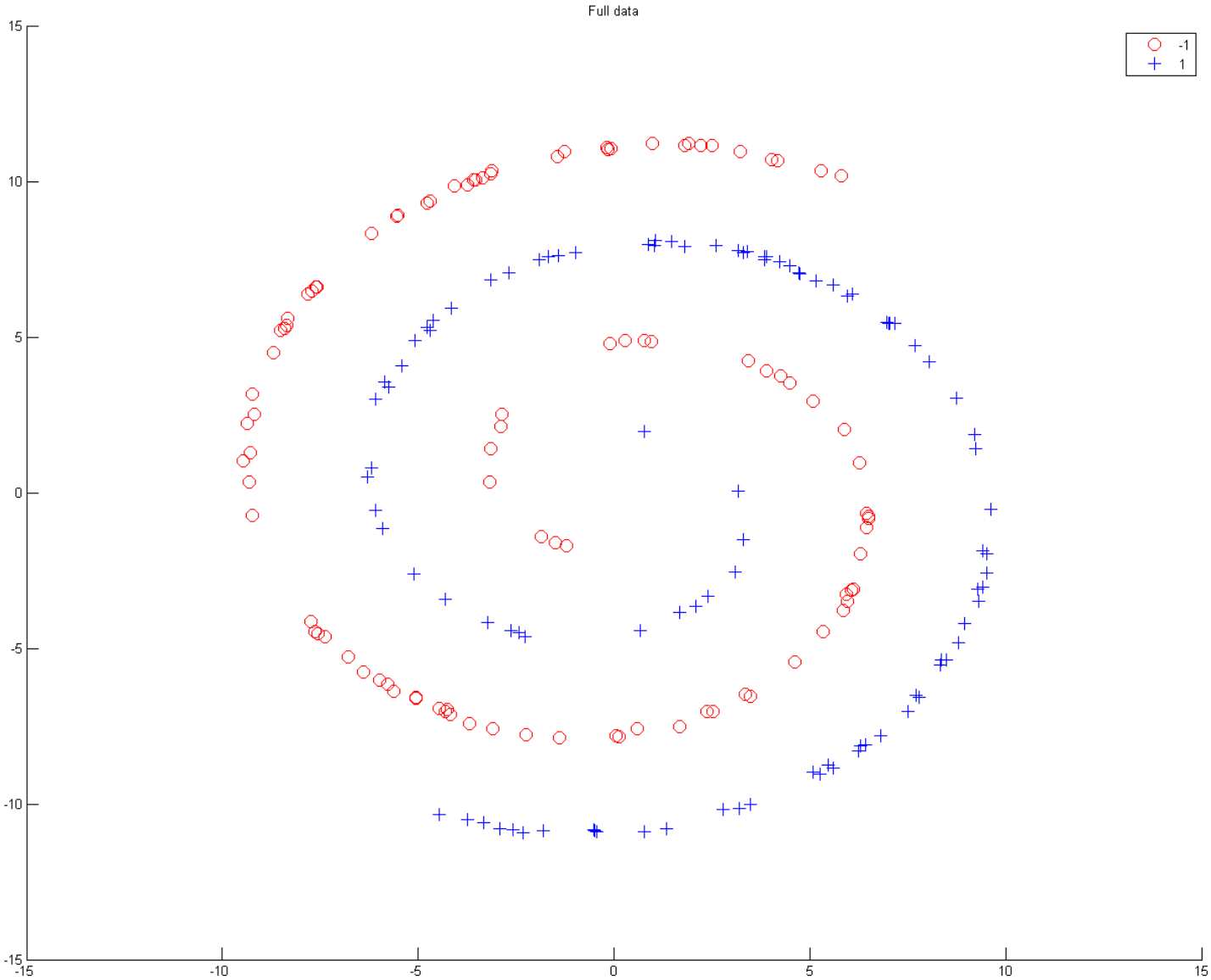
Piecewise Linearly Seperable



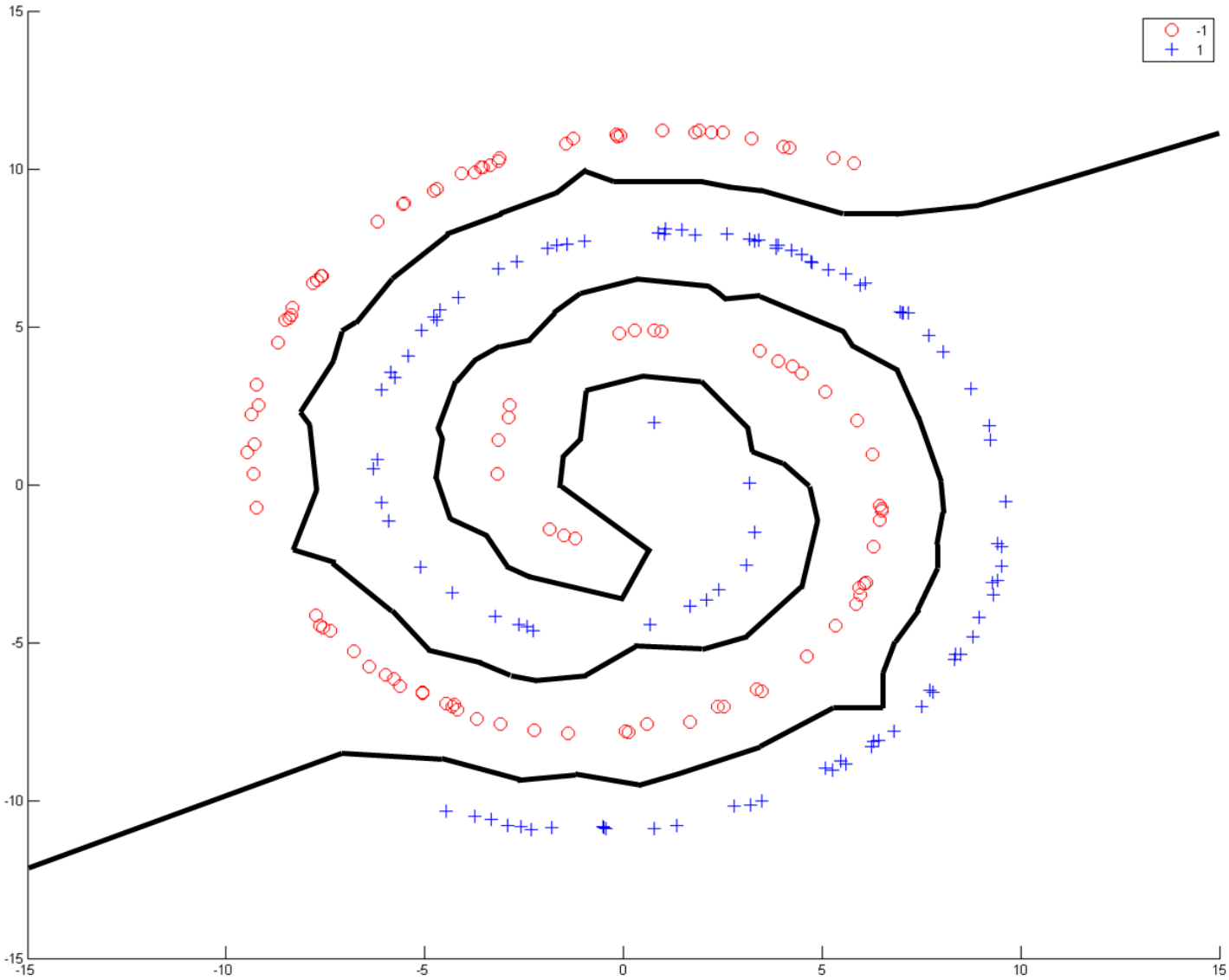
Piecewise Linearly Seperable



Piecewise Linearly Seperable - Spirals (100 Pts)



Piecewise Linearly Seperable - Spirals (100 Pts)



Non-separable Data

- It is not always the case that the data distributions being classified are comprised of disjoint data classes
- Preprocessing approaches
- Wilson's editing algorithm eliminates misleading examples from a training set
  - Various techniques may be used such as k-Nearest Neighbor or K-Means Clustering
- Kernel trick method uses an appropriate kernel mapping to cast both non-linear and non-separable data to a higher dimensionality in which the data is separable