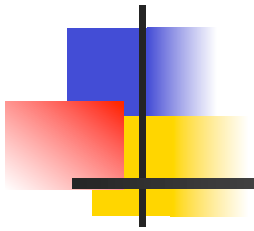


Classification as Clustering:
A Pareto Cooperative-Competitive GP Approach

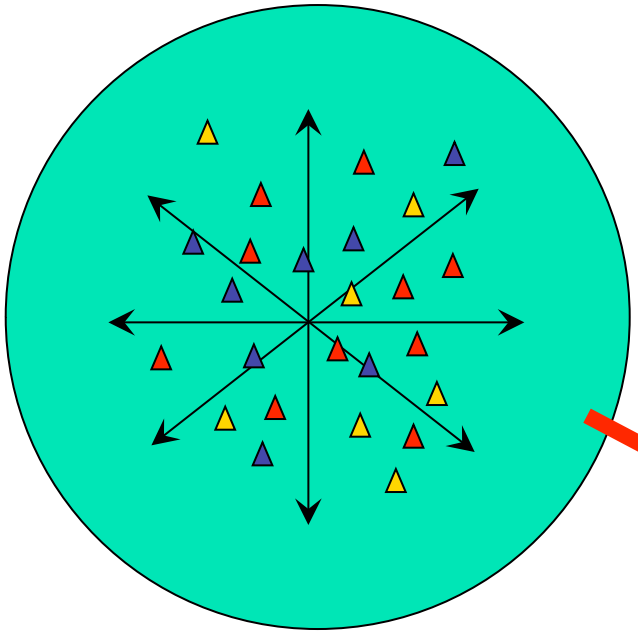
Andrew McIntyre, Malcolm Heywood

Evolutionary Computation Journal

(to appear)

MIT Press

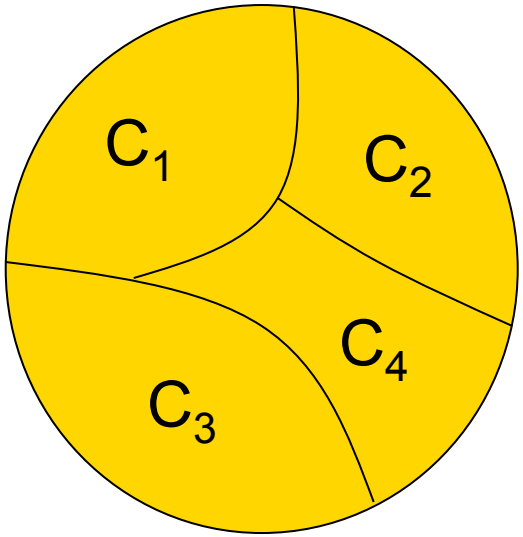
Classification Problem Domain



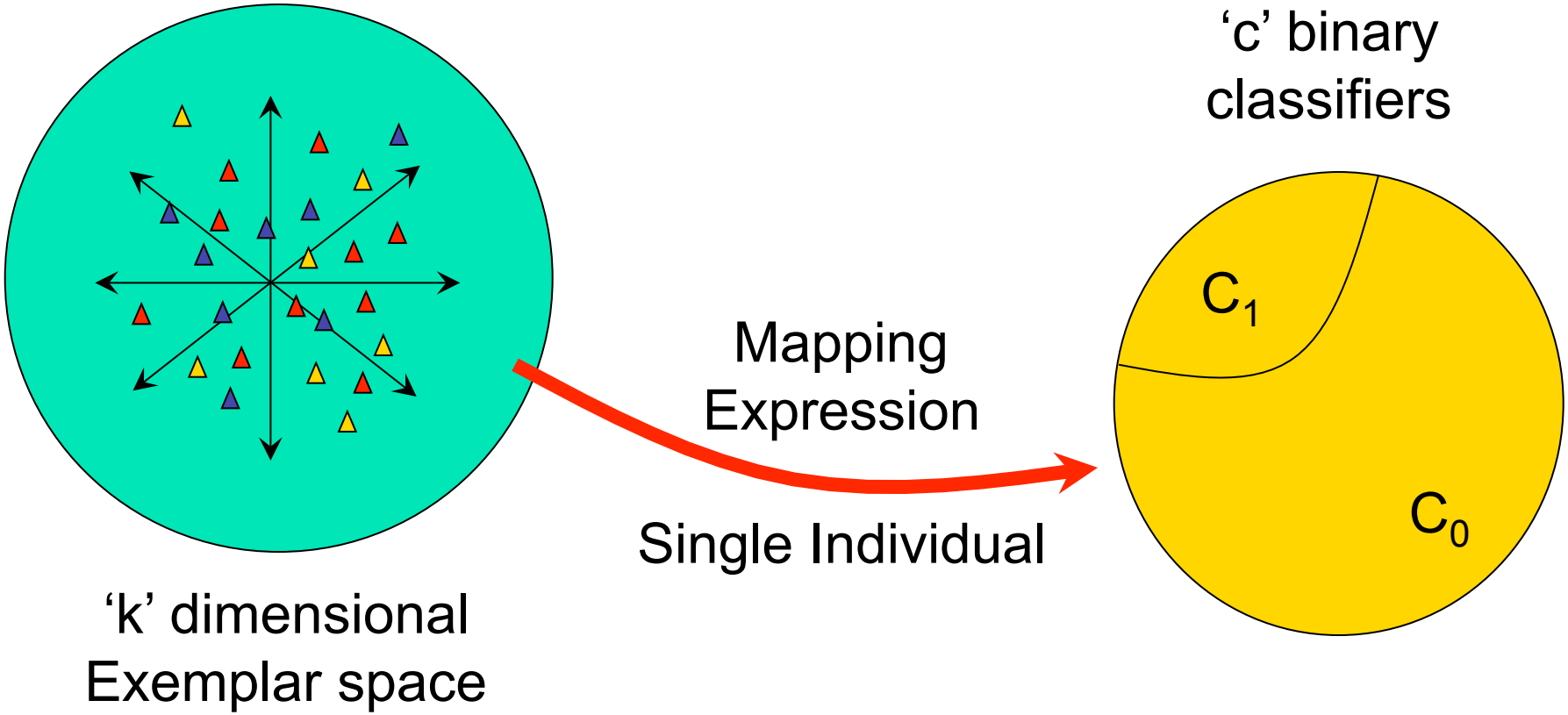
‘k’ dimensional
Exemplar space

Initially
Unknown mapping
expression

‘c’ dimensional class
Consistent space



Canonical GP Classifier Methodology

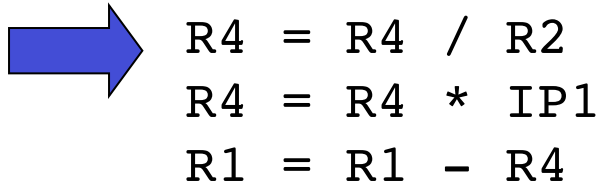


Canonical GP Classifier Pragmatics

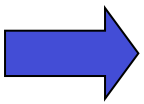
GP individual → Mapping

Set of Training Exemplars

<Exemplar → vector of 'k' features>



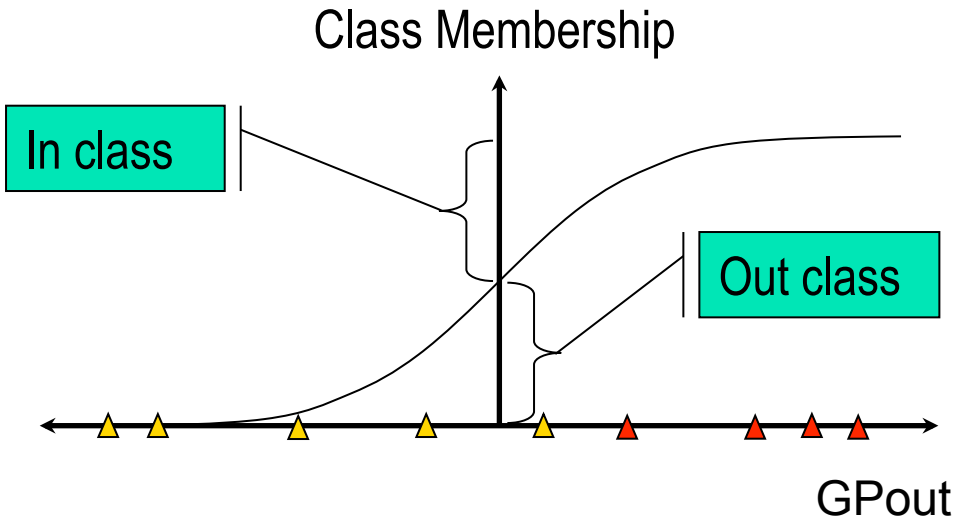
```
R1 = R1 - IP4
R4 = R4 / R2
R4 = R4 * IP1
R1 = R1 - R4
```



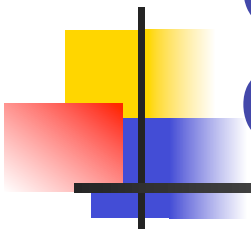
Wrapper Operator

→ Global Membership function

→ Denote 'R1' as 'GPout'

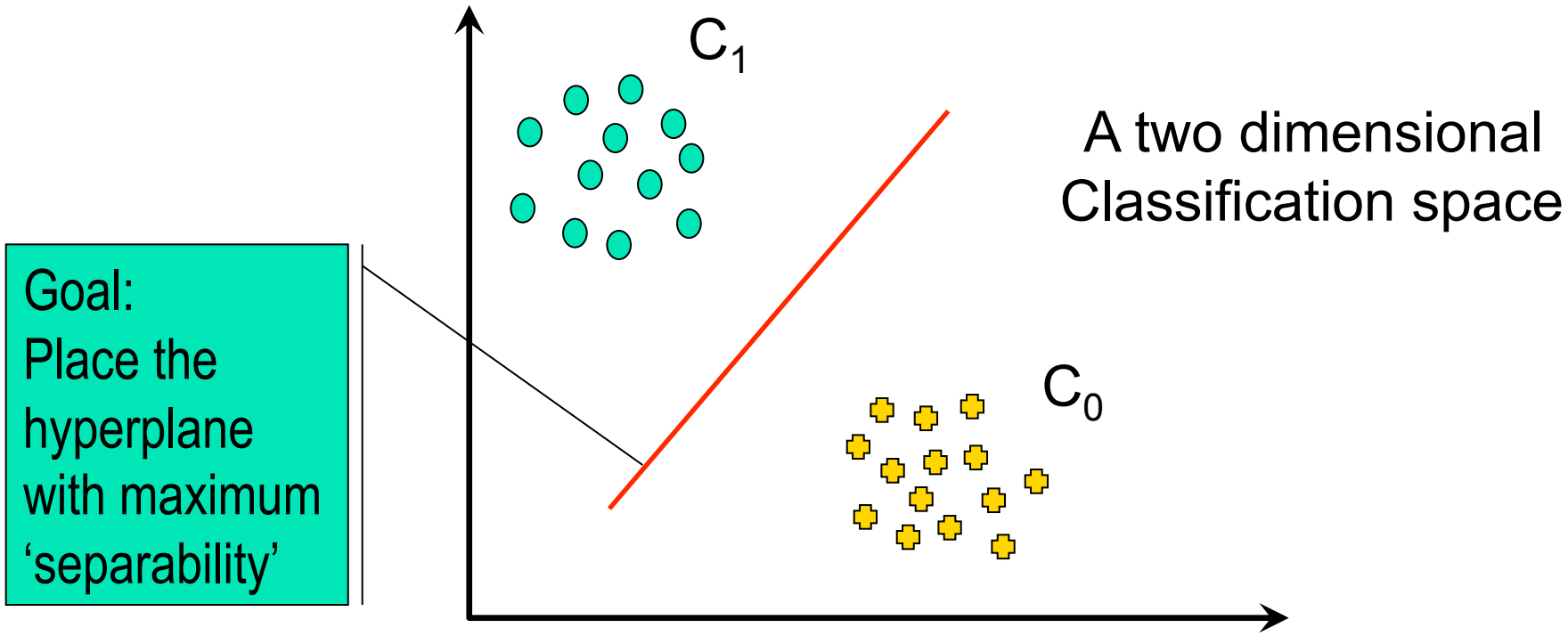


Comparison of Ideal versus Canonical GP Classifier Properties

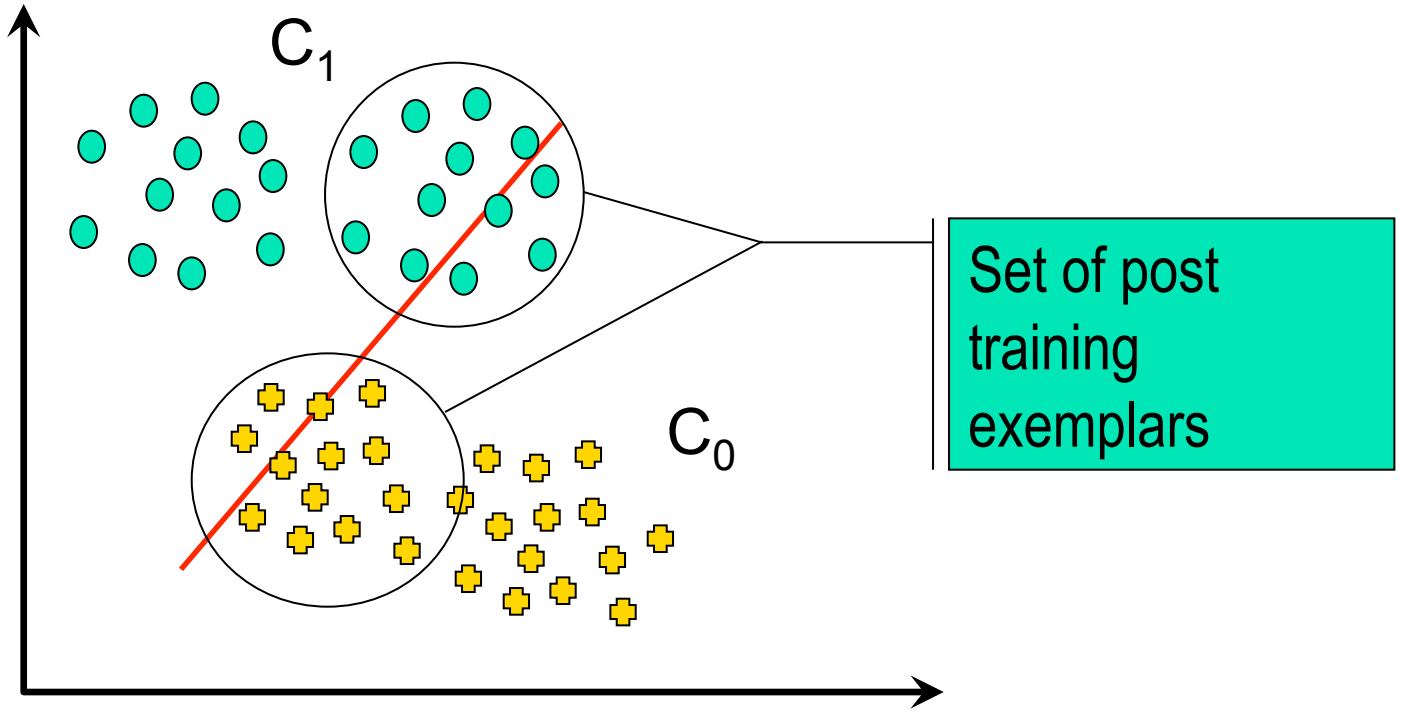


- Ideal
  - Problem Decomposition
  - Predicable behavior
  - Scalable training
  - Minimal model complexity
- Canonical GP
  - Limited to single '**super**' individual
  - '**discrimination**' based classification assumed
  - Exhaustive iteration over **entire** data set
  - Supported

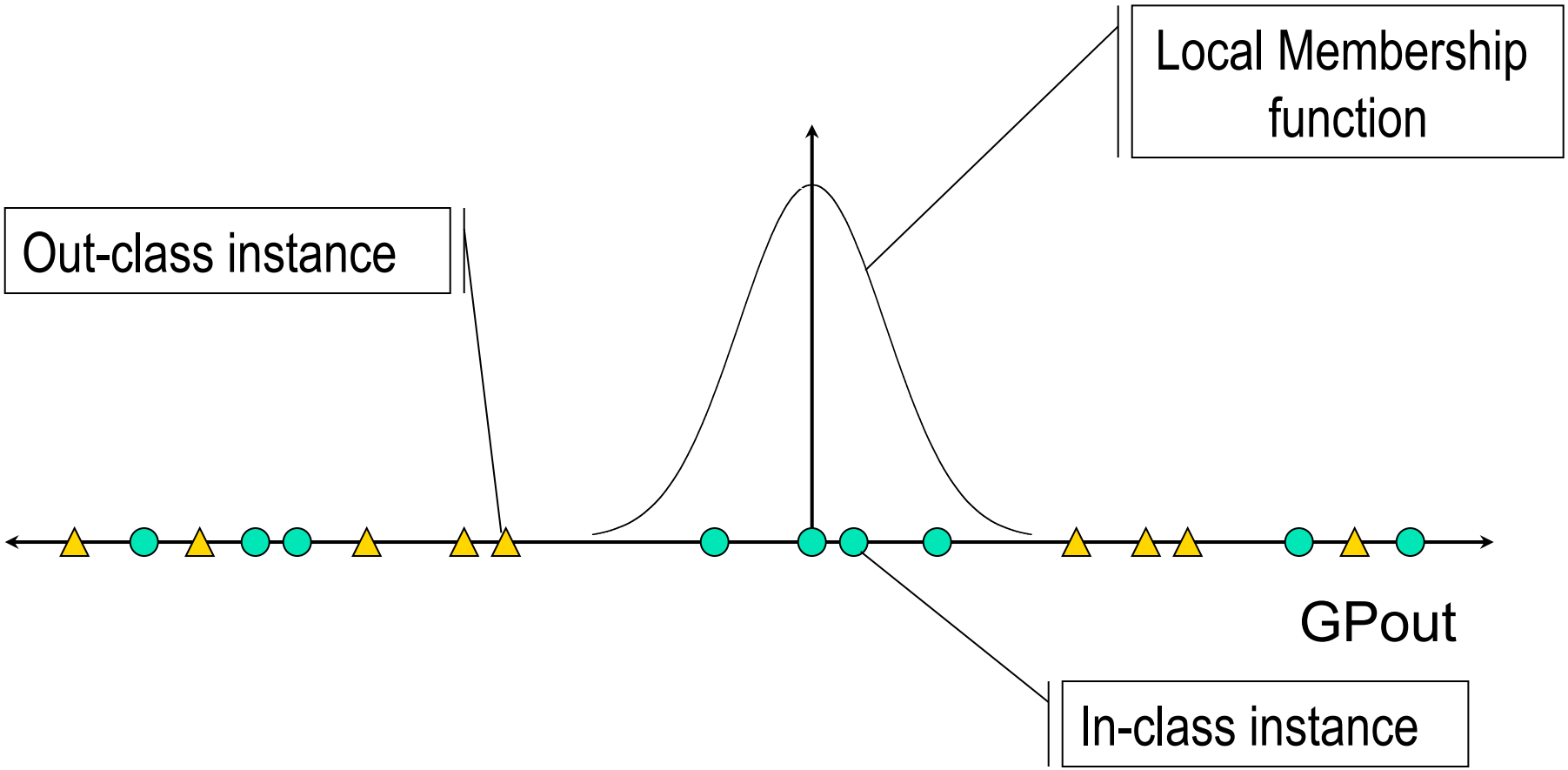
Discriminator based Classification



Disadvantage of Discriminatory Classification model: Unpredictable behaviour



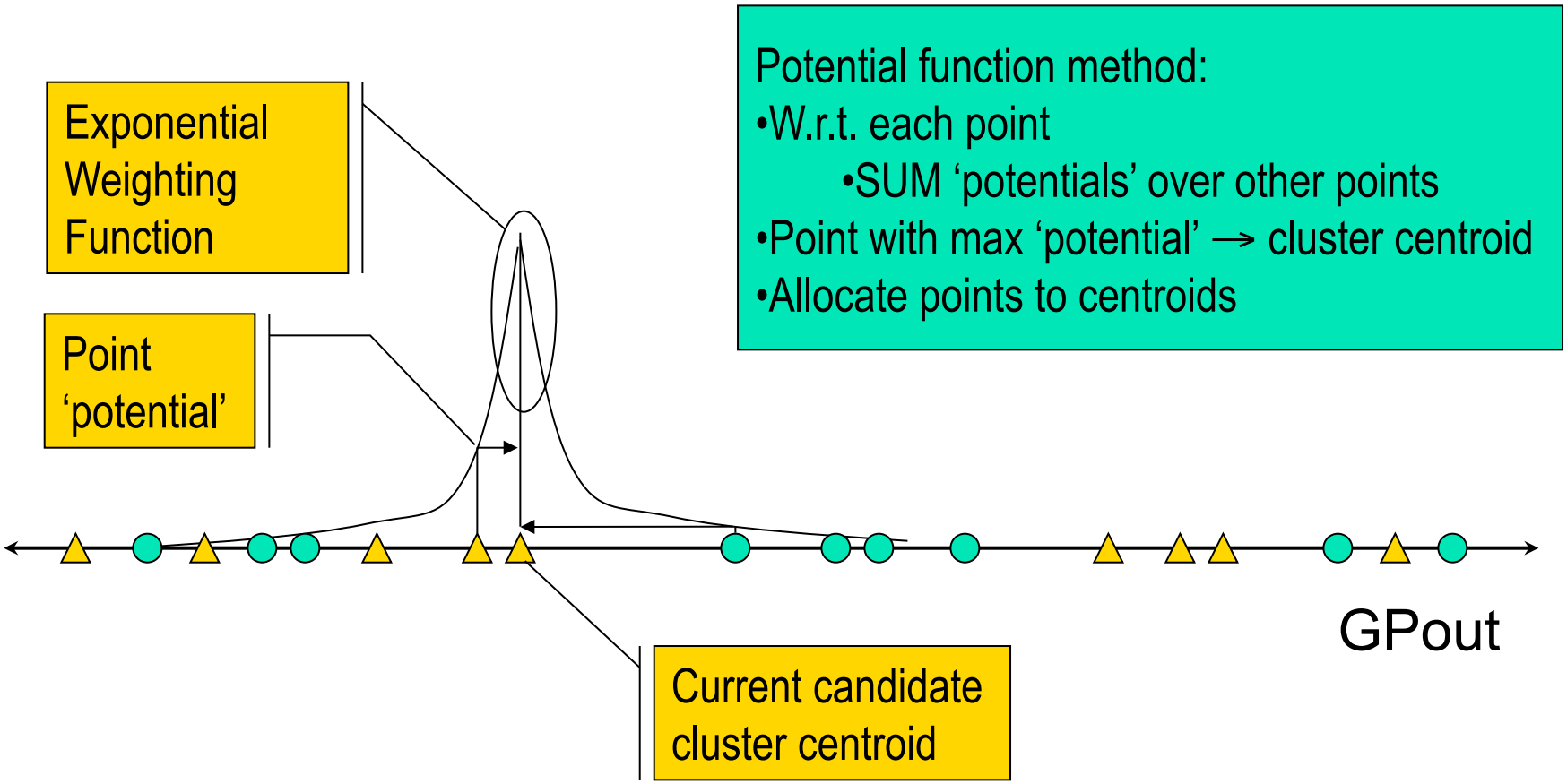
Novelty Detection:
Local Membership Function



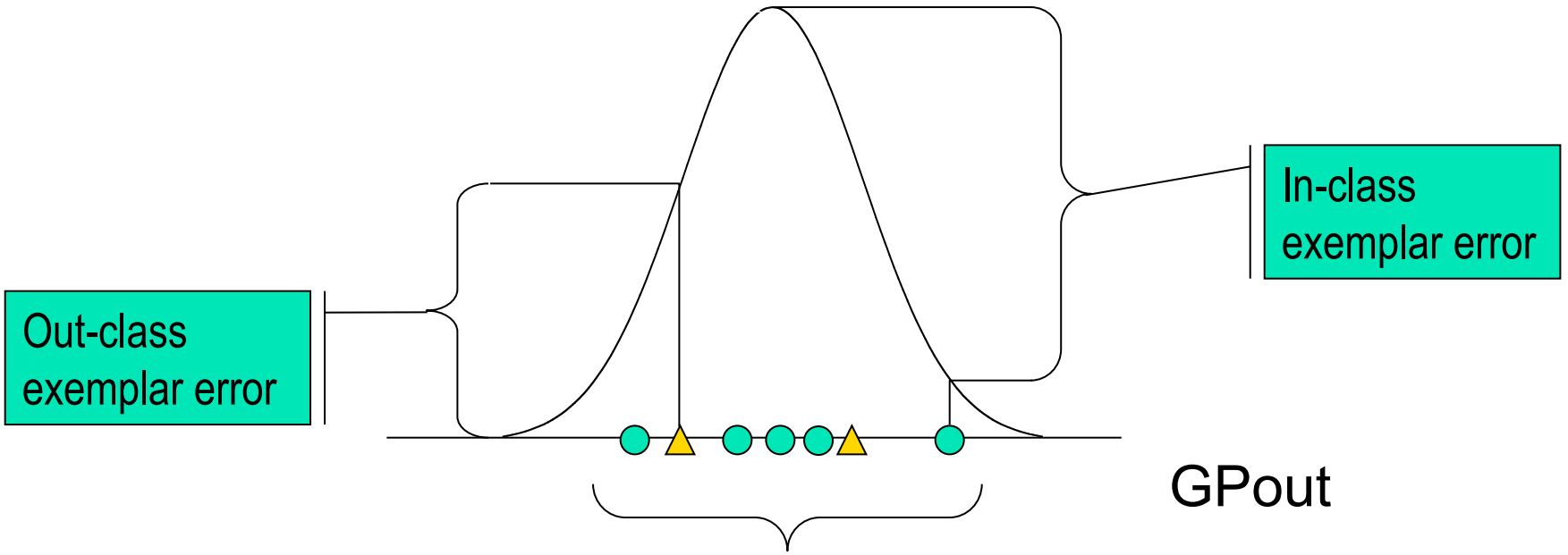
Novelty Detection:
Design Challenges

- What subset of 'GPout' should the LMF be associated with?
- How is total exemplar 'coverage' maximized?
  - How are individuals encouraged to map to different exemplar subsets?
- Clustering
  - Density heuristic
  - No a priori target cluster count
- Cooperative Coevolution
  - Cost function → Pareto Multi-criteria Optimization

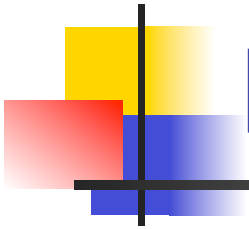
Identify 'GPout' Region with Max Density: Class independent



Cluster Characterization

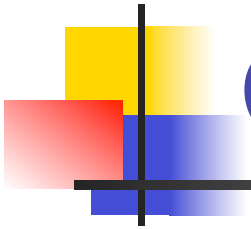


$\sigma \rightarrow$ exemplar subset assigned to LMF



Recap

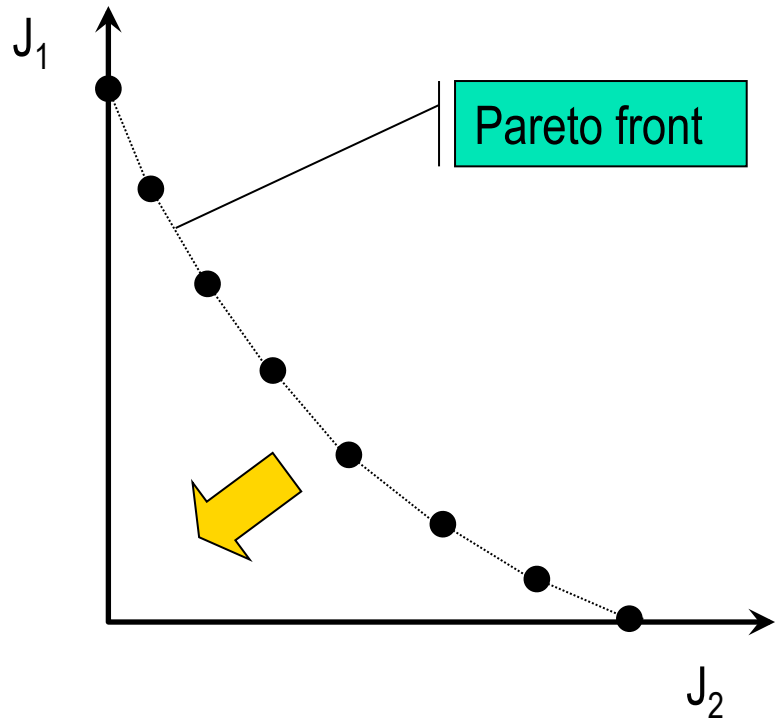
- A GP individual maps
  - Multi-dimensional Input space TO
    - 1D output space
- Cluster Identifies region of ‘GPout’ axis
  - Returns exemplars associated with most dense region
  - Build corresponding local membership function (LMF)
- Next Goal
  - Establish credit assignment model
  - Establish mechanism for cooperative behaviour



Cost function

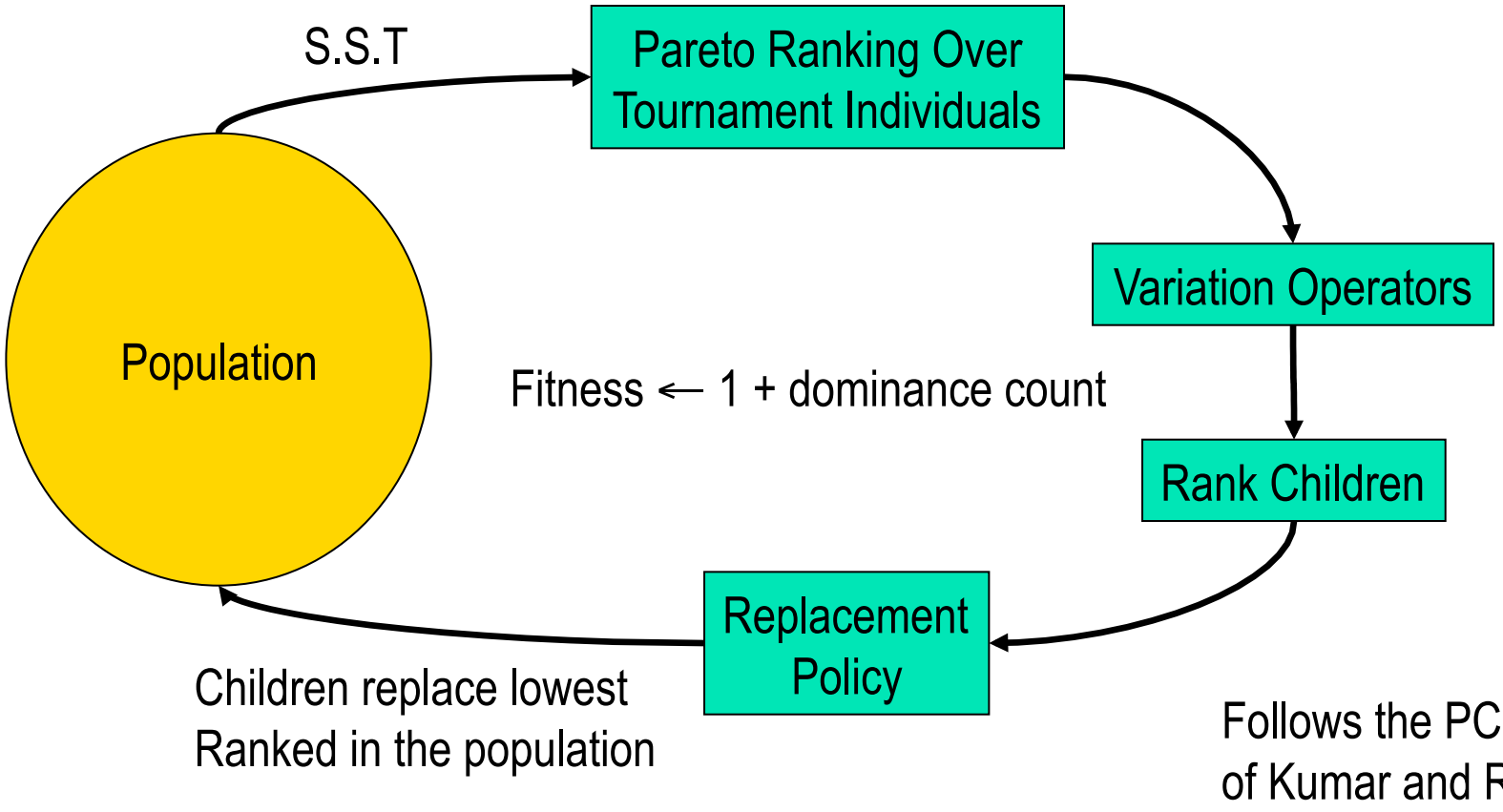
- Relative to exemplar subset
  - Minimize sum square error (SSE)
    - Penalizes out of class membership
  - Maximize in-class exemplar count
    - Penalizes degenerate behaviours relative to 1st objective
  - Minimize overlap in exemplars mapped by different individuals
- Case for
  - Multi-criteria Optimization
  - Cooperative Coevolution

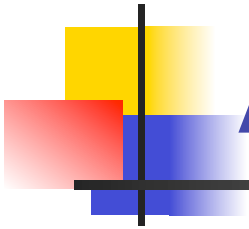
Pareto Evolutionary Multi-criterion Optimization



- 'A' dominates 'B'
  - IFF
    - A and B equivalent on objective $J_i \in \mathbf{J}$
    - A better than B on at least one objective, Ji
- Refinements,
  - 'ε' dominance
    - Establishes the degree of differentiation necessary for dominance to be satisfied.

Evolutionary Multi-criteria Optimization (EMO)

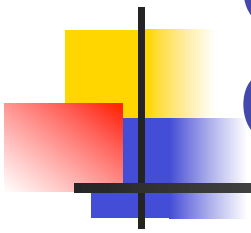




Additional Properties of EMO

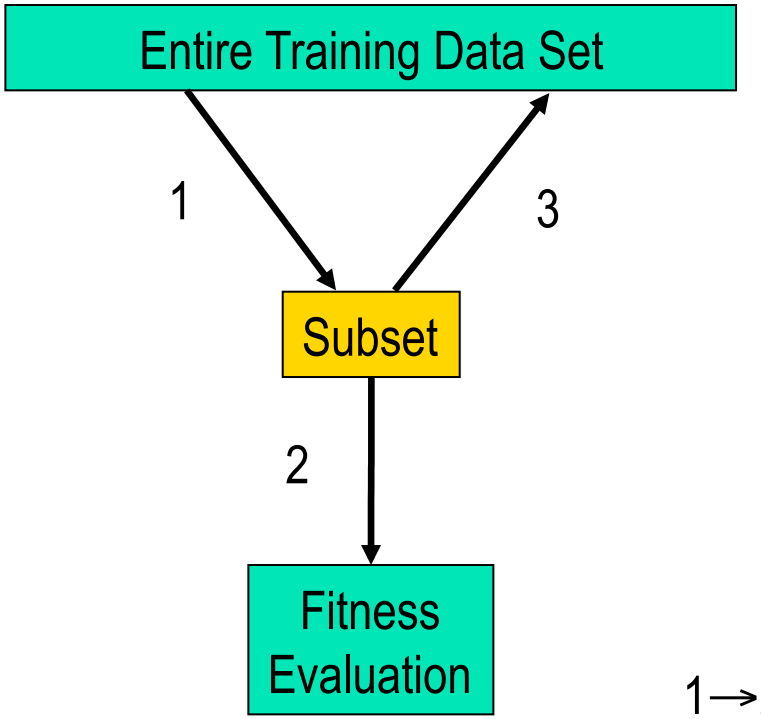
- S.S. tournament provides
  - Elitism
  - Tides broken with rank + 1 assigned to one individual
  - Incremental modification of population
- Problem independent stop criterion
  - IF content of the ‘rank 1’ Pareto front is static
  - THEN assume model converged

Comparison of Ideal versus EMO GP Classifier Properties



- Ideal
  - Problem Decomposition
  - Predicable behavior
  - Scalable training
  - Minimal model complexity
- EMP GP Classifier
  - Natural attribute of evolution
  - Novelty detection paradigm enforced
  - More expensive than Canonical GP
  - Explicitly encouraged

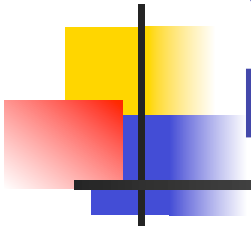
Reducing the Computational Overhead: Active Learning



- Sampling (Weiss & Provost)
  - Uniform sampling (US)
  - US + Balance Enforcing
- Competitive Coevolution
  - Host Parasite Model
  - Pareto EMO Model

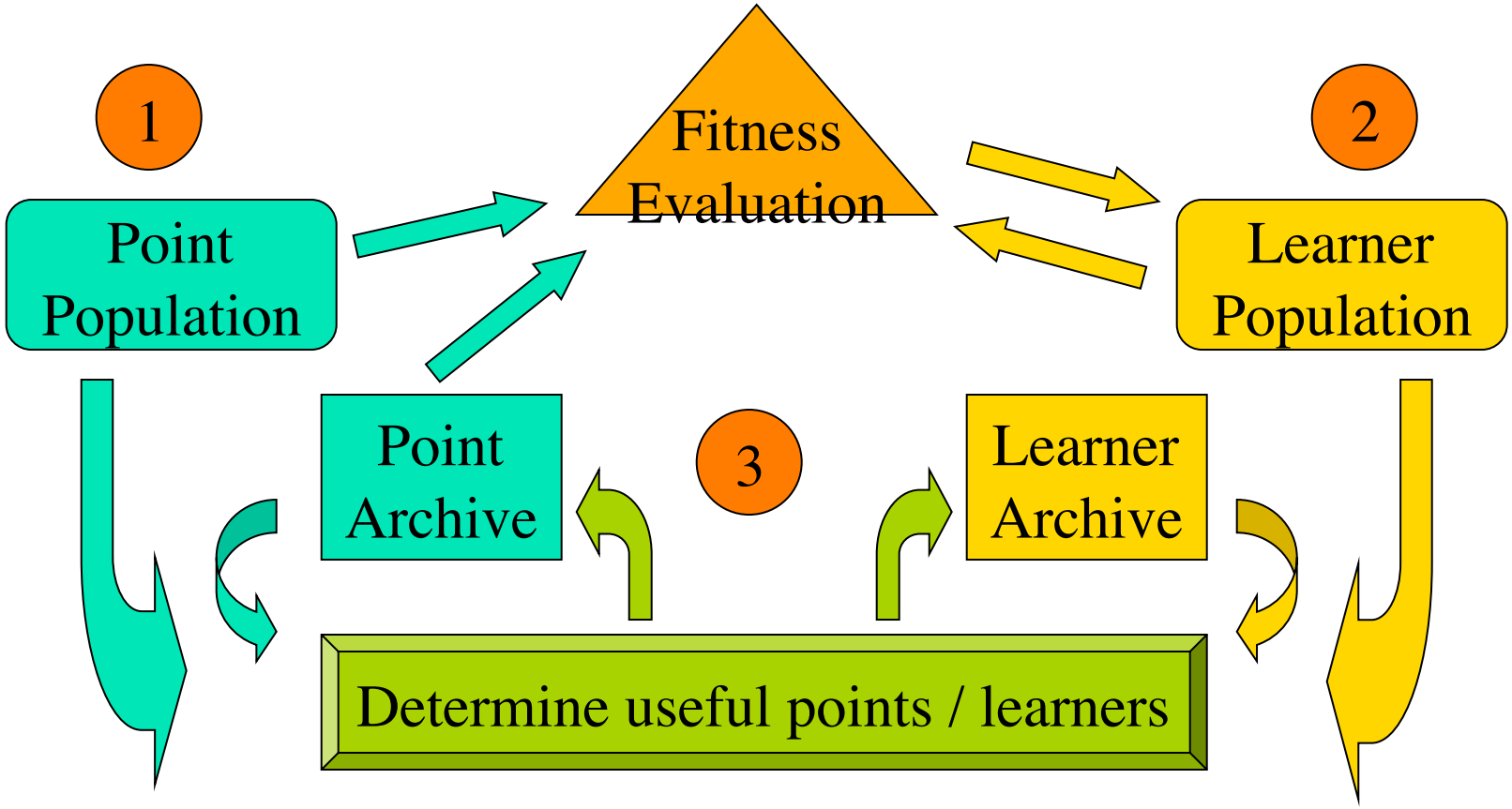
1→2 →3 →1 →2 →3 →....

Competitive Coevolution: Pareto EMO

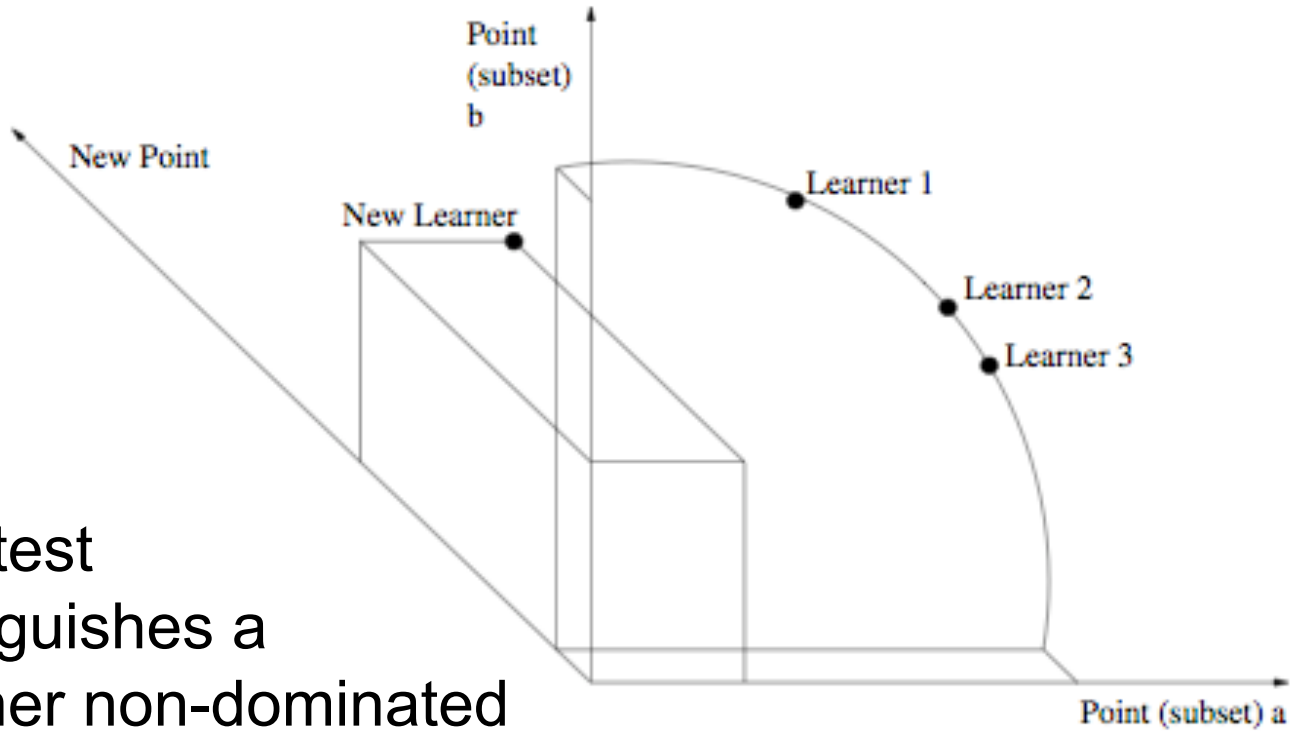


- Two Population Model with Archiving (de Jong)
  - Exemplar (subset)
  - Classifiers
- Exemplar ranking:
  - Ability to distinguish between <highest ranked> Learners (Ficici & Pollock)
- Learner ranking:
  - Ability to maximize classifier objectives
- Caveats:
  - Disengagement and forgetting

Competitive Coevolution:
Archive model

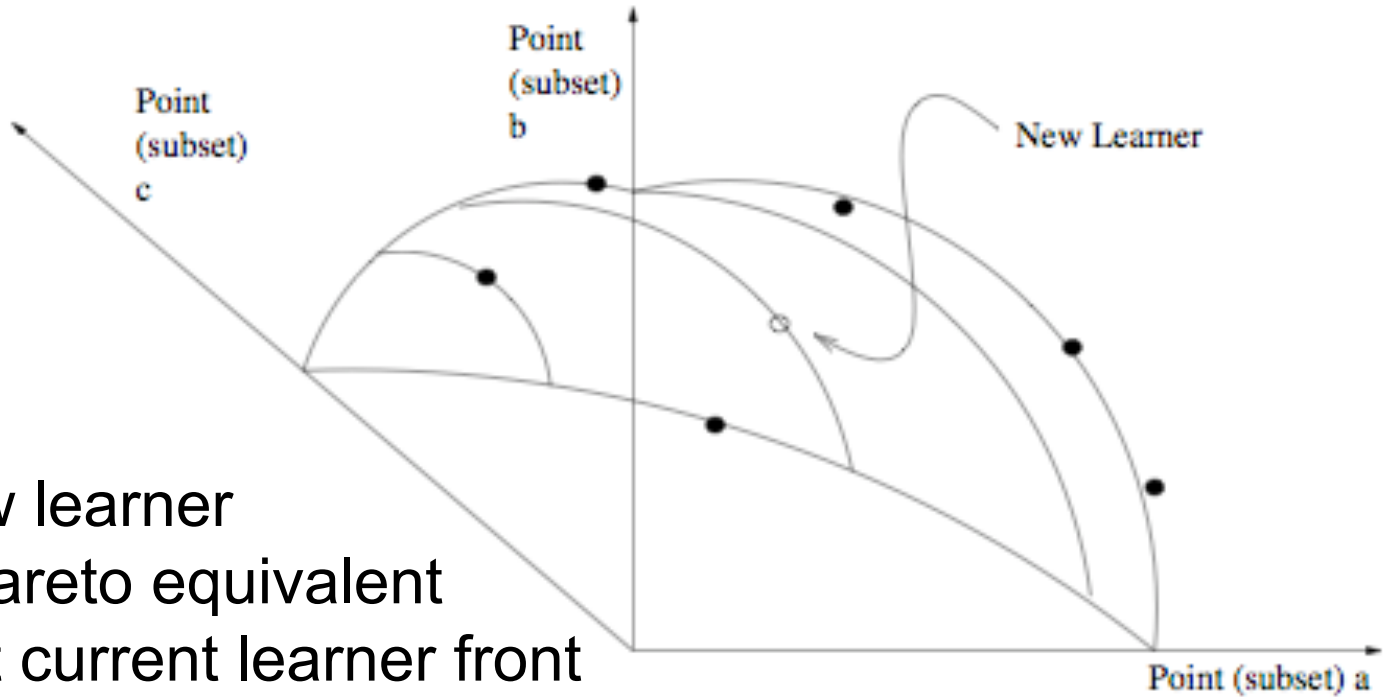


Step 3a: Update Point Archive



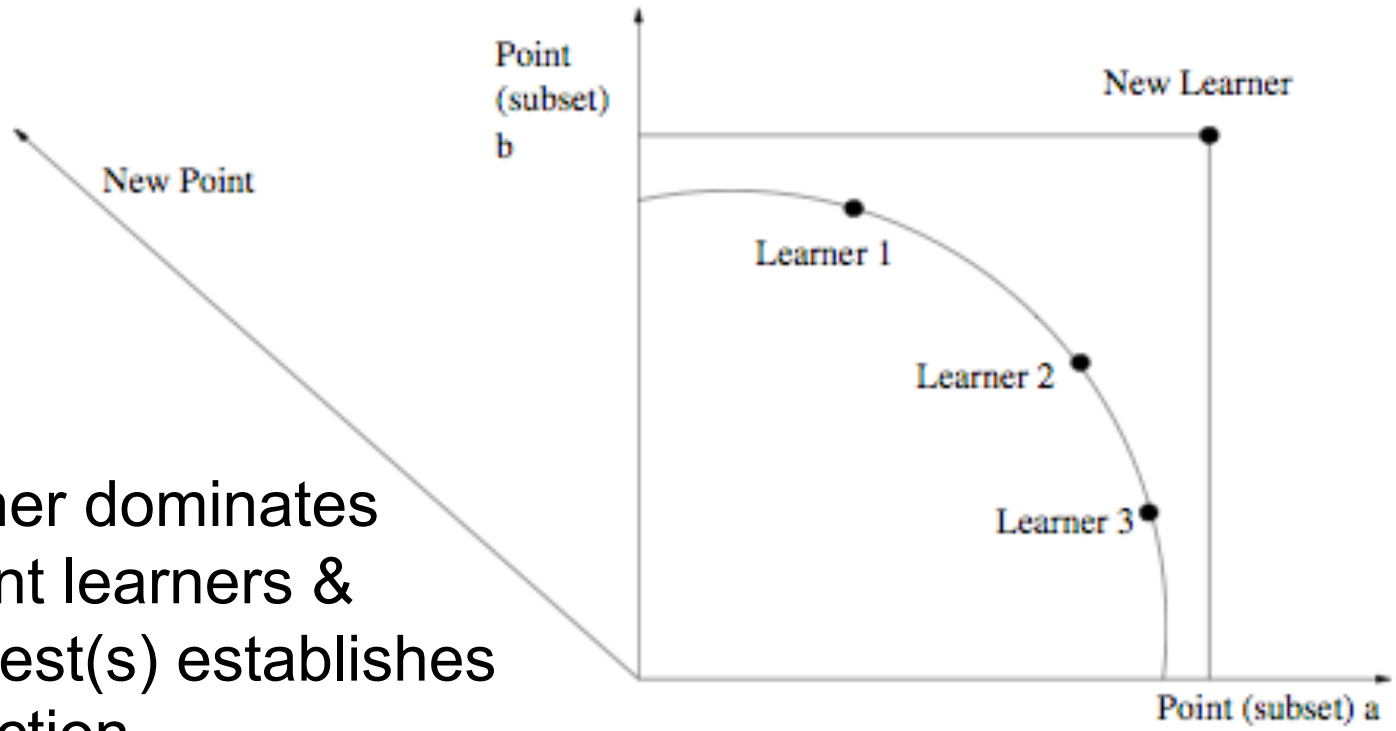
New test distinguishes a Learner non-dominated

Step 3b: Update Learner Archive - Case 1



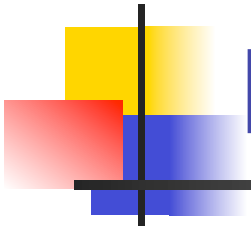
New learner
is pareto equivalent
w.r.t current learner front

Step 3b: Update Learner Archive - Case 2



Learner dominates current learners & new test(s) establishes distinction

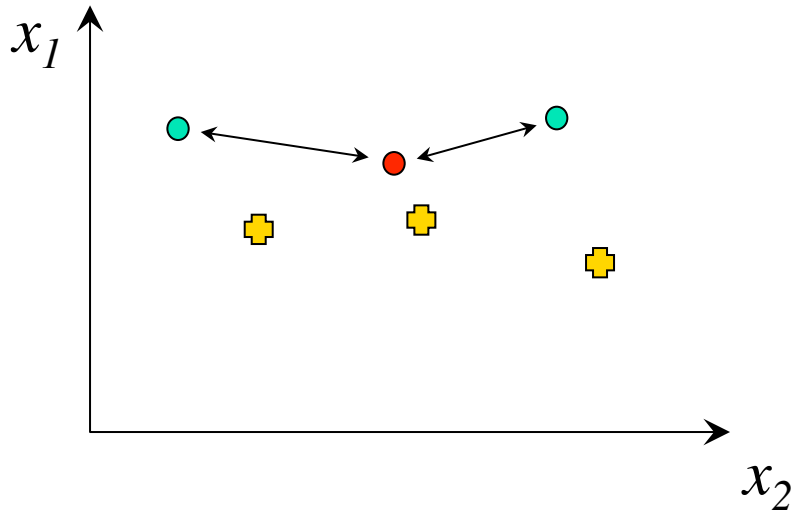
Enforcing archive size limits - Part 1



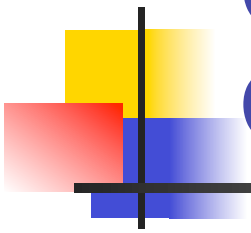
- Learner Archive
  - Greedy replacement bias
    - Replace individuals with highest miss classification count first (w.r.t point archive)
      - Tends to replace individuals with lower correct classification counts

Enforcing archive size limits - Part 2

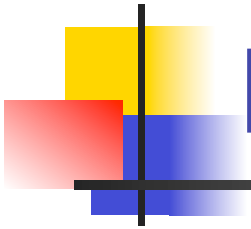
- Point Archive
  - Genotypic similarity, class consistent, balance enforcing



Comparison of Ideal versus Cooperative-Competitive Coevolution

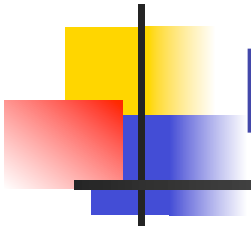


- Ideal
  - Problem Decomposition
  - Predicable behavior
  - Scalable training
  - Minimal model complexity
- CCC Classifier
  - Natural attribute of evolution
  - Novelty detection paradigm enforced
  - Significantly faster than canonical GP
  - Explicitly encouraged



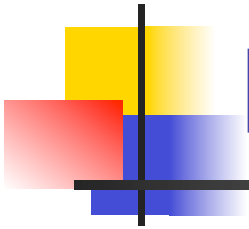
Benchmark Design

- Parameterization
  - Common GP
  - Scalable GP variants
  - Common Evaluation Limit
- Data Sets
  - Multi-class, variable class distributions
- Performance metrics
  - Accuracy versus ‘score’
- Models Compared
  - EC models
  - Deterministic baseline
- Behavioural Assessment
  - Assess learner to exemplar relationships



Parameterization: GP model

- Common GP model
  - Grammatical Evolution
  - Crossover & Mutation
    - Std and context aware variants
  - Arg 2 → arithmetic operators
  - Arg 1 → cosine, square root, log, exp.
  - Arg 0 → domain features

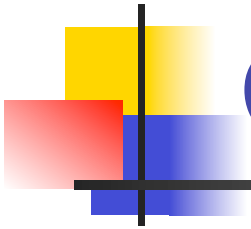


Parameterization: GP variants

- (Canonical) GE
  - Fitness est. over all exemplars (no eval. Limit)
- Std GE
  - Fitness est. over all exemplars*
- Rss GE
  - Fitness est. over Balanced uniform subset selection*
- PGEC
  - Fitness est. over Balanced uniform subset + archiving*
- * implies “common evaluation limit”

Parameterization:
Evaluation Limits

Model	Fitness Eval.	Stop Criterion
GE	All exemplars	Max. Generations
Std GE	All exemplars	Evaluation Limit
RSS GE	Subset	
PGEC	Subset + Archive	
CMGE		Evaluation Limit × Classes



CMGE variants

- CMGE1
  - Greedy learner archive pruning
  - W.r.t learner error and point subset
- CMGE2
  - Greedy learner archive pruning
  - W.r.t learner out of class error and point subset

Parameterization:
Miscellaneous

- LP size: 50
- LA, PP, PA size: 30
- # Runs: 50
- Post training evaluation:

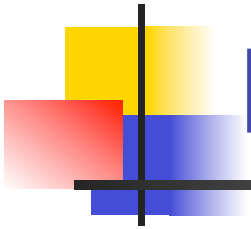
$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn}$$

$$\text{Score} = \frac{DR_0 + DR_1 + \ldots + DR_C}{C}$$

$\rightarrow$ Scores of < 1/C are degenerate

Data Sets

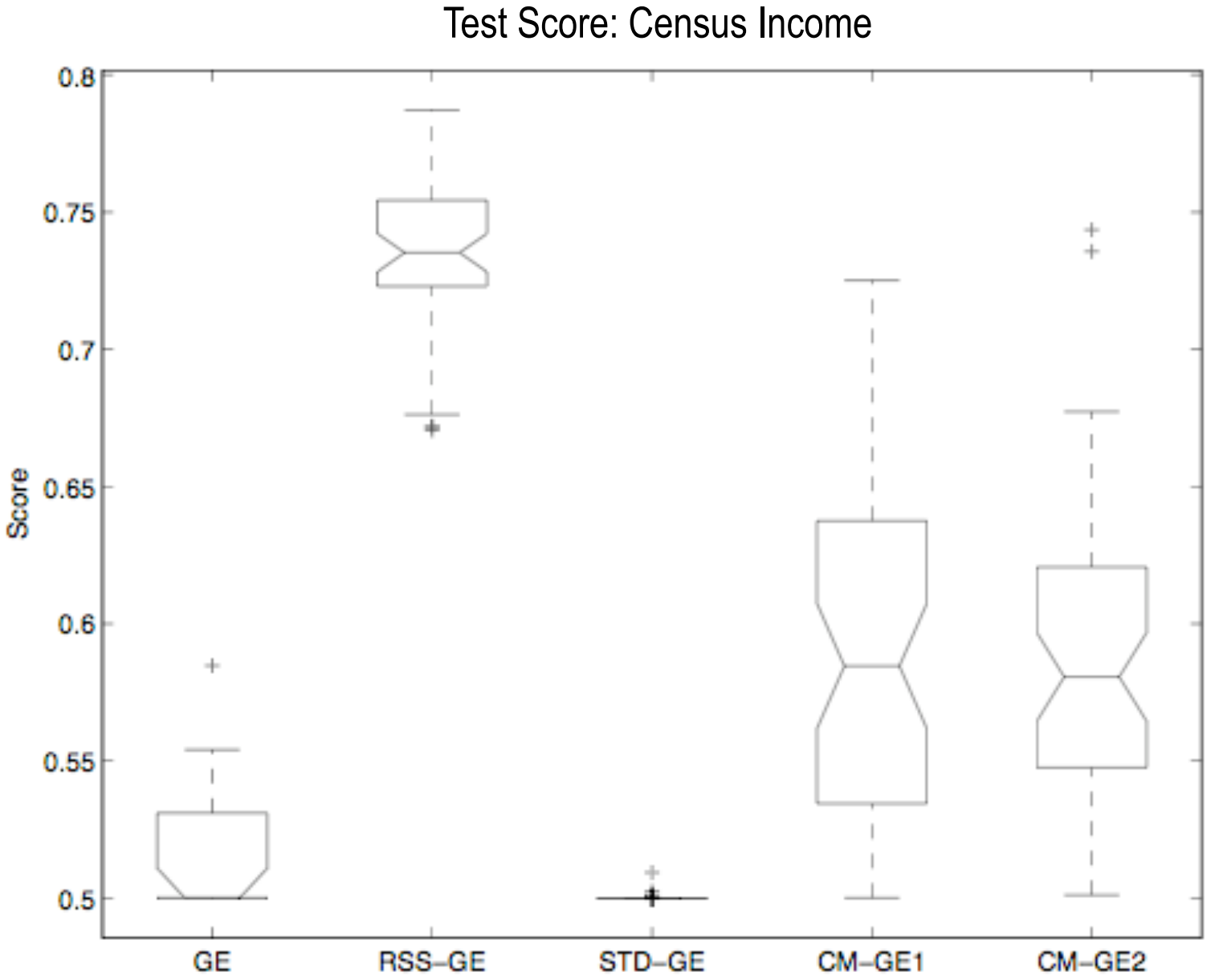
Data Set	Total Data	Percent Training	Percent Test	Training Data Class Distribution (%)
BOST	506	10 fold		Equal
BUPA	341	10 fold		42 / 58
CENS	295,173	66.5	33.5	94 / 6
CONT	1,425	10 fold		43 / 22 / 35
IMAG	2,310	9.1	90.9	Equal
IRIS	147	10 fold		Equal
KD99	222,871	65	35	60 / 0.8 / 37.5 / 1.5 / >.1
PIMA	768	10 fold		65 / 35
SHUT	58,000	75	25	78 / >.1 / 0.3 / 15 / 5.6 / >.1 / >.1
THYD	7,129	52	48	2 / 5 / 93
WINE	178	10 fold		33 / 40 / 27
WISC	675	10 fold		65 / 35



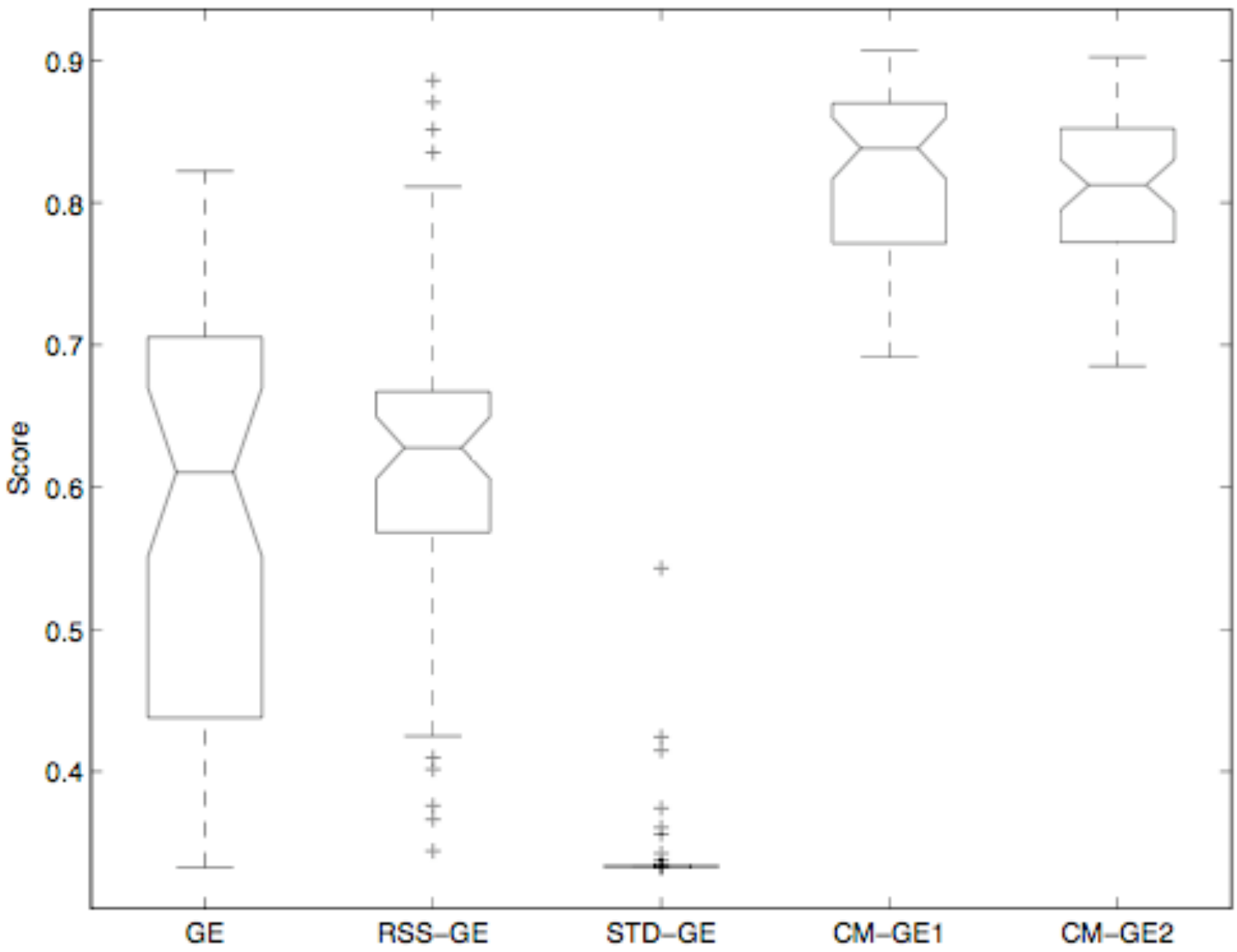
Benchmarking Study

- Results for all EC algorithms
  - Subset of ‘interesting’ data sets
  - Overall Ranking
    - Parametric / Non parametric confidence tests
- Results for EC subset versus C4.5
  - Establish methodology for comparison

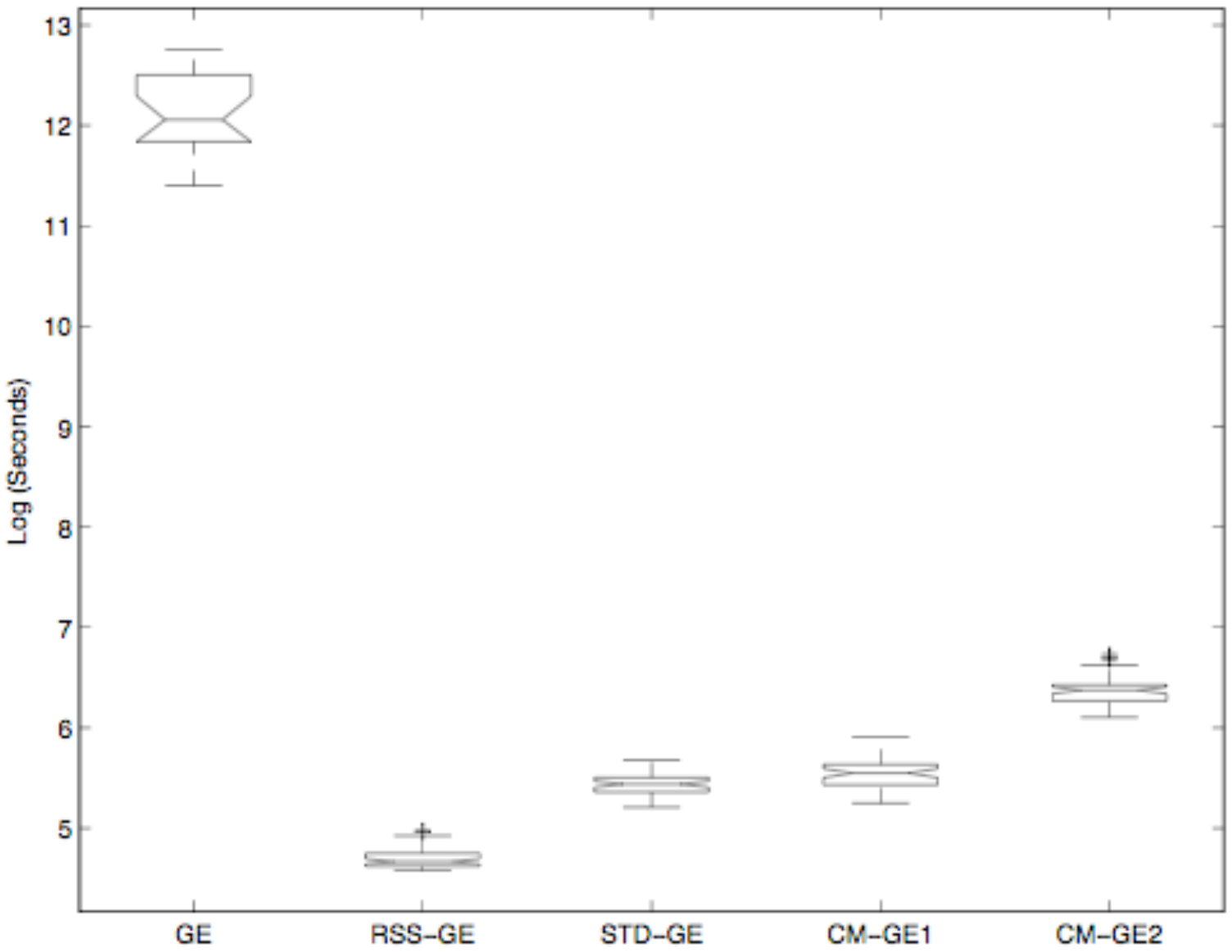
GE Comparison



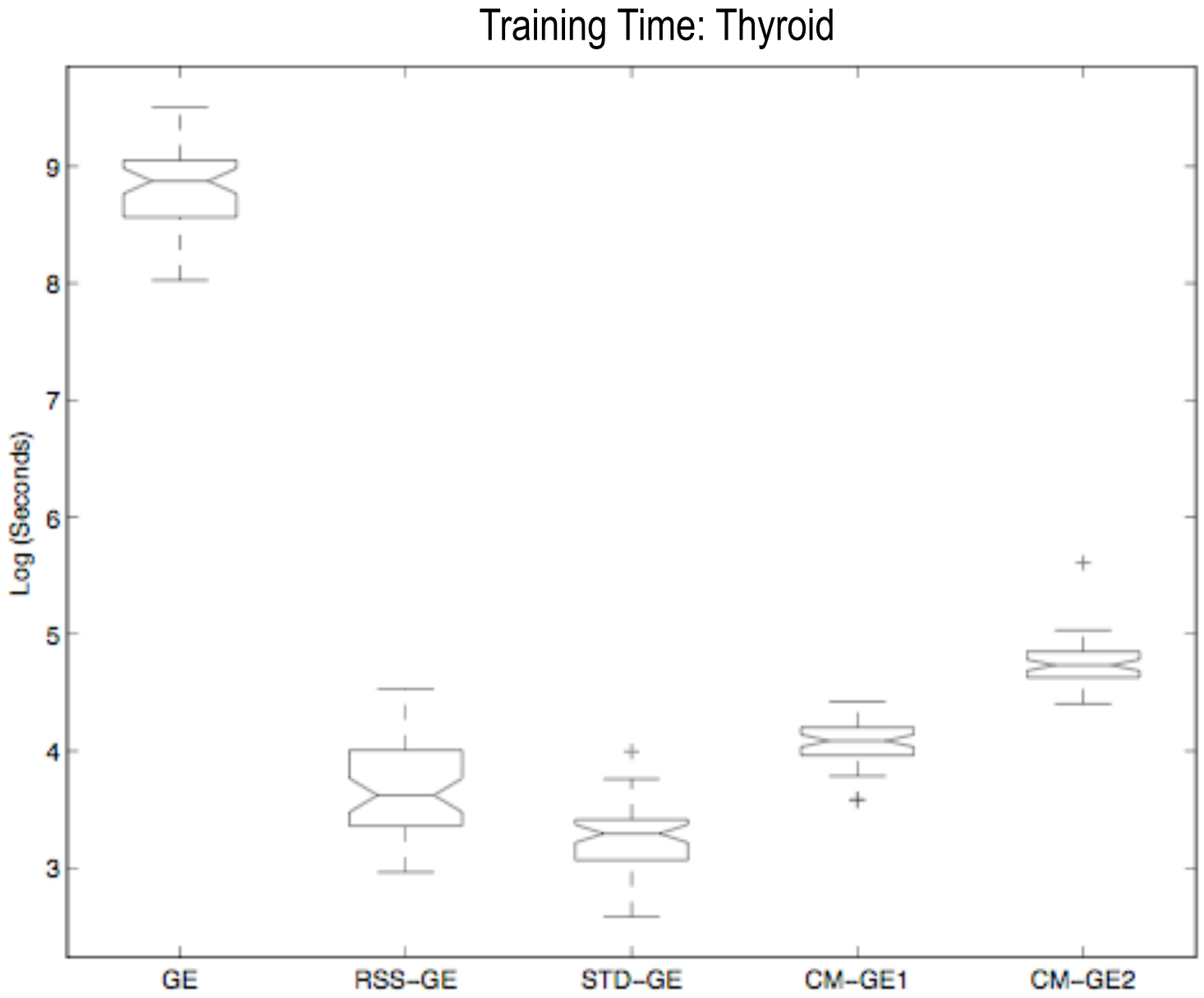
GE Comparison



GE Comparison



GE Comparison



CMGE versus Rss and Std GE

Ranked Accuracy Metric Training (Test if different)

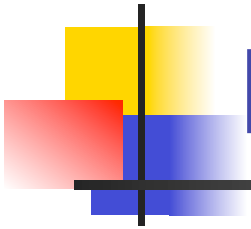
Data set	RssGE	StdGE	CMGE1	CMGE2
BOST	3	4	2 (1)	1 (2)
BUPA	4	3	2	1
CENS	4	1	3	2
CONT	4	3	2	1
IMAG	4	3	1	2
IRIS	4	3	2 (1)	1 (2)
KD99	3	4	2 (1)	1 (2)
PIMA	4	5	1	2
SHUT	4	5	2	1*
THYD	4	3 (1)	2 (3)	1 (2)
WINE	4	3	1*	2
WISC	4	3	2	1*

CMGE versus RssGE and StdGE

Ranked Score Metric
Training (Test if different)

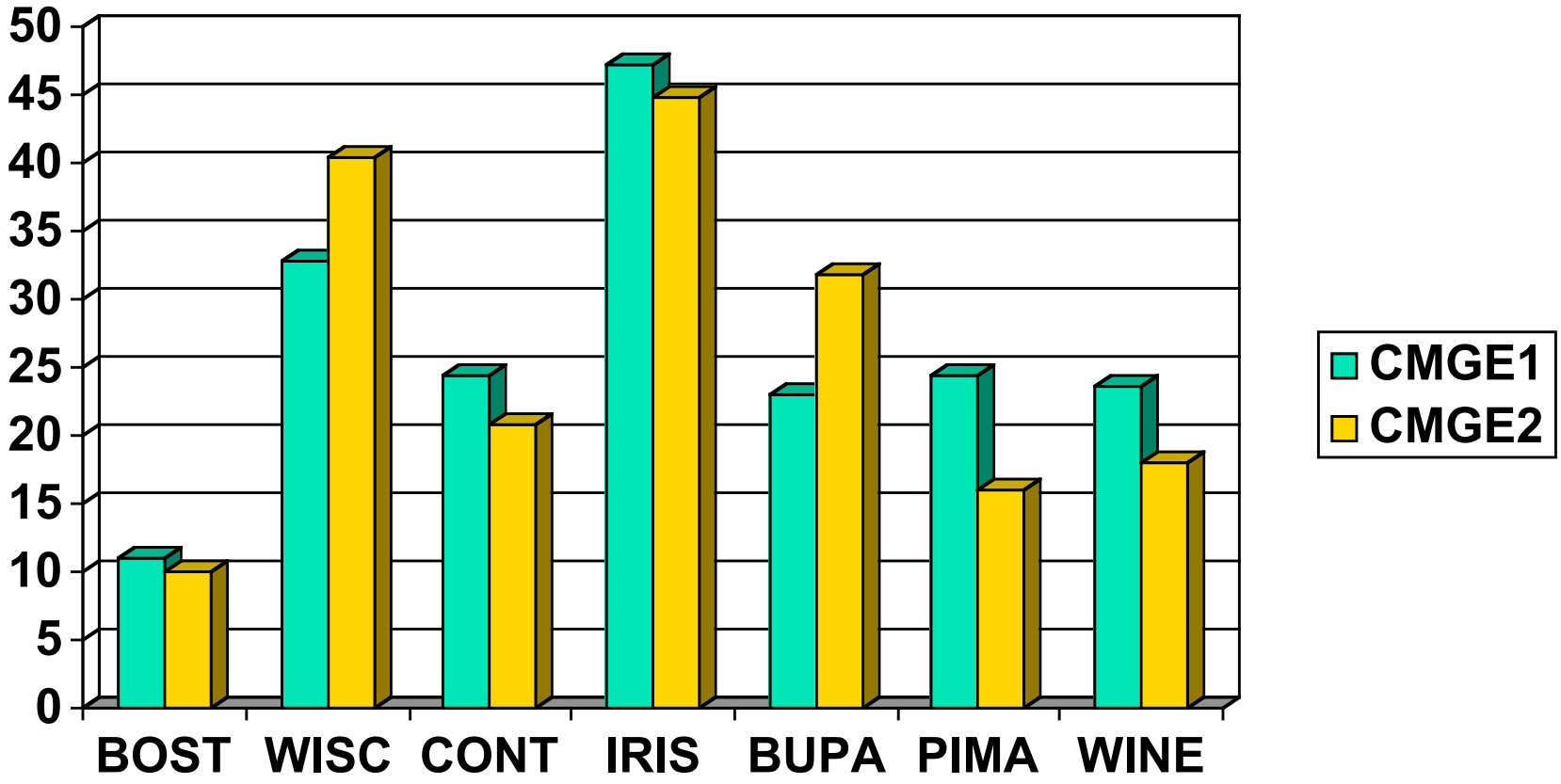
Data set	RssGE	StdGE	CMGE1	CMGE2
BOST	3	4	2 (1)	1 (2)
BUPA	3	4	2	1*
CENS	1*	4	2	3
CONT	1*	4	2	3
IMAG	4	3	1	2
IRIS	4	3	2 (1)	1 (2)
KD99	3	4	1	2
PIMA	3	4	2	1*
SHUT	3	4	1 (2)	2 (1)
THYD	3	4	1	2
WINE	4	3	1*	2
WISC	4	3	2	1*

Methodology for Deterministic Learning comparison

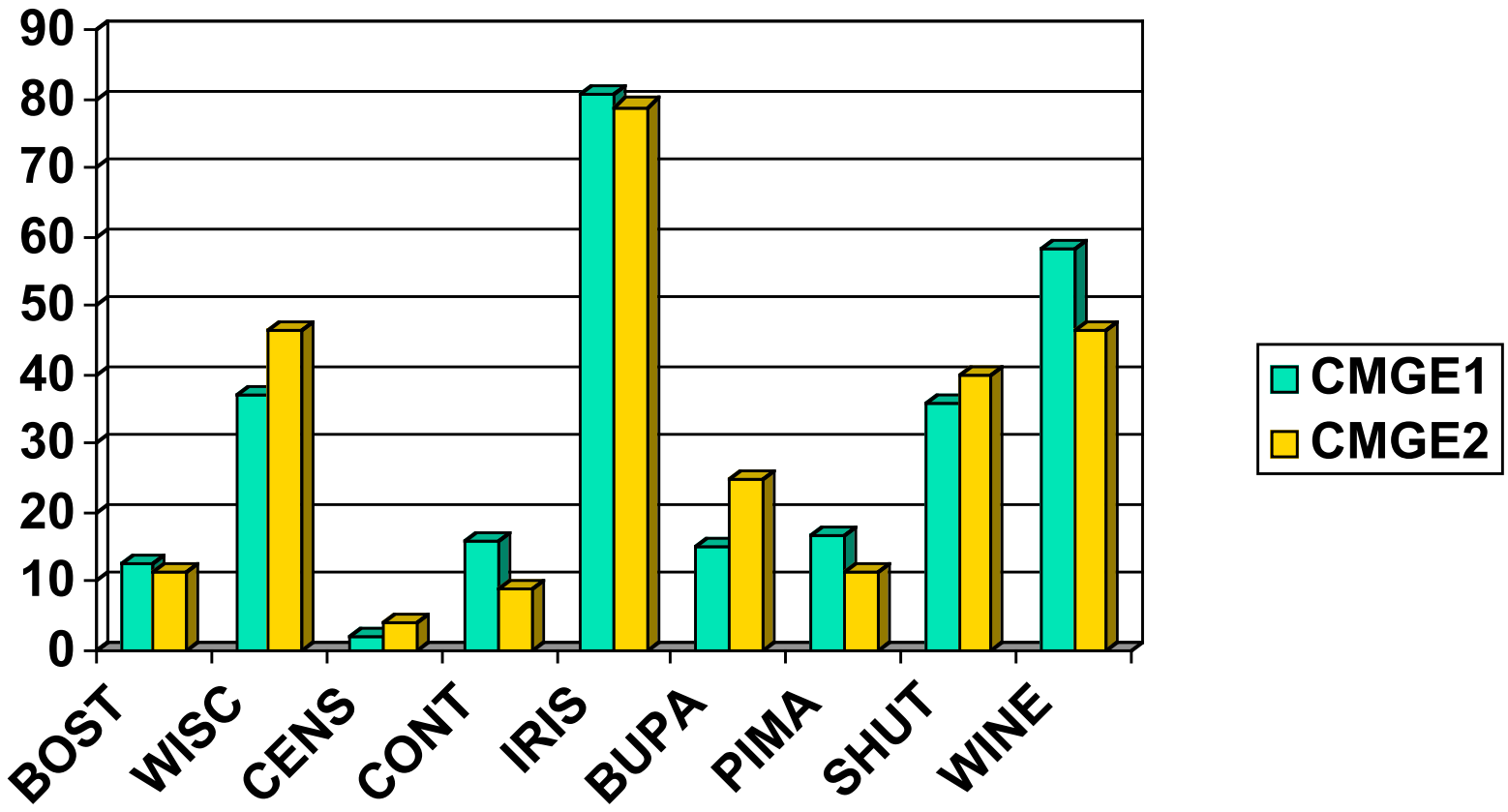


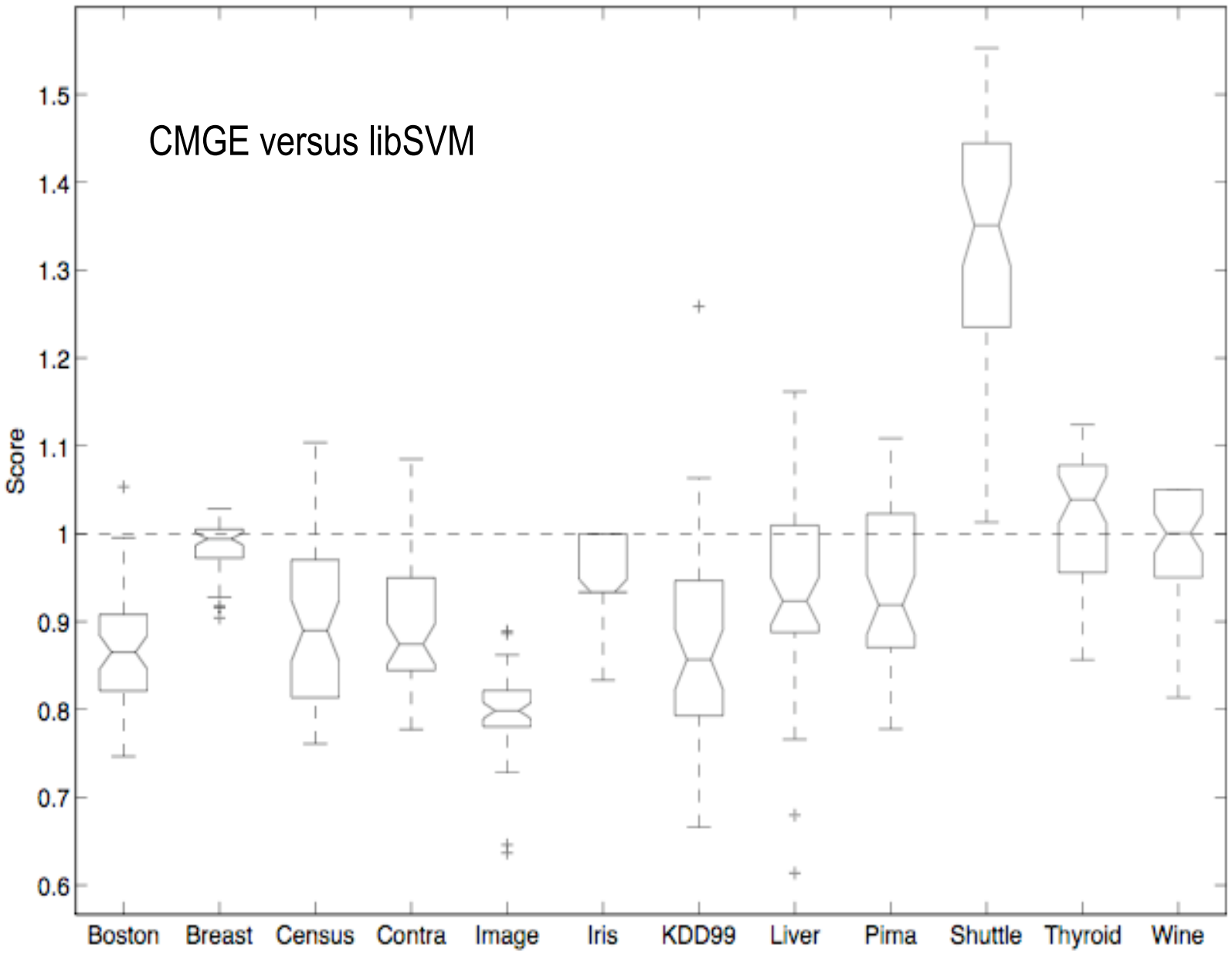
- EC based on the opportunity to make non greedy decisions
  - Multiple initializations
  - 10 fold → 10 points (deterministic) versus 500 points (EC)
- Deterministic → performance threshold
- EC → P(EC better than Deterministic)

CMGE versus C4.5: % better or equivalent test ACCURACY

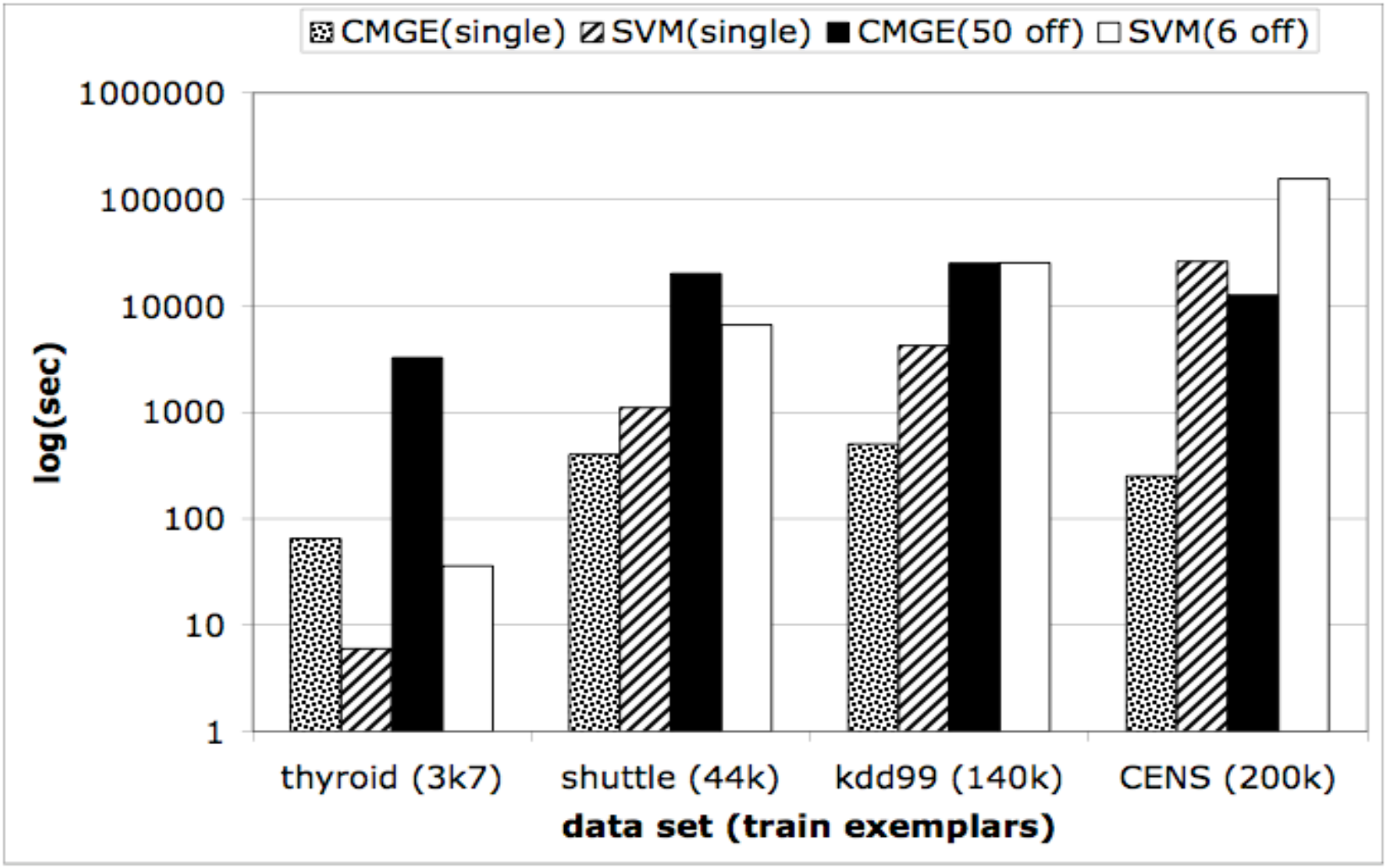


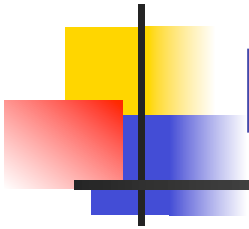
CMGE versus C4.5:
% better or equivalent test SCORE




CMGE versus libSVM
Score
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
Boston
Breast
Census
Contra
Image
Iris
KDD99
Liver
Pima
Shuttle
Thyroid
Wine

CMGE versus libSVM: Training time

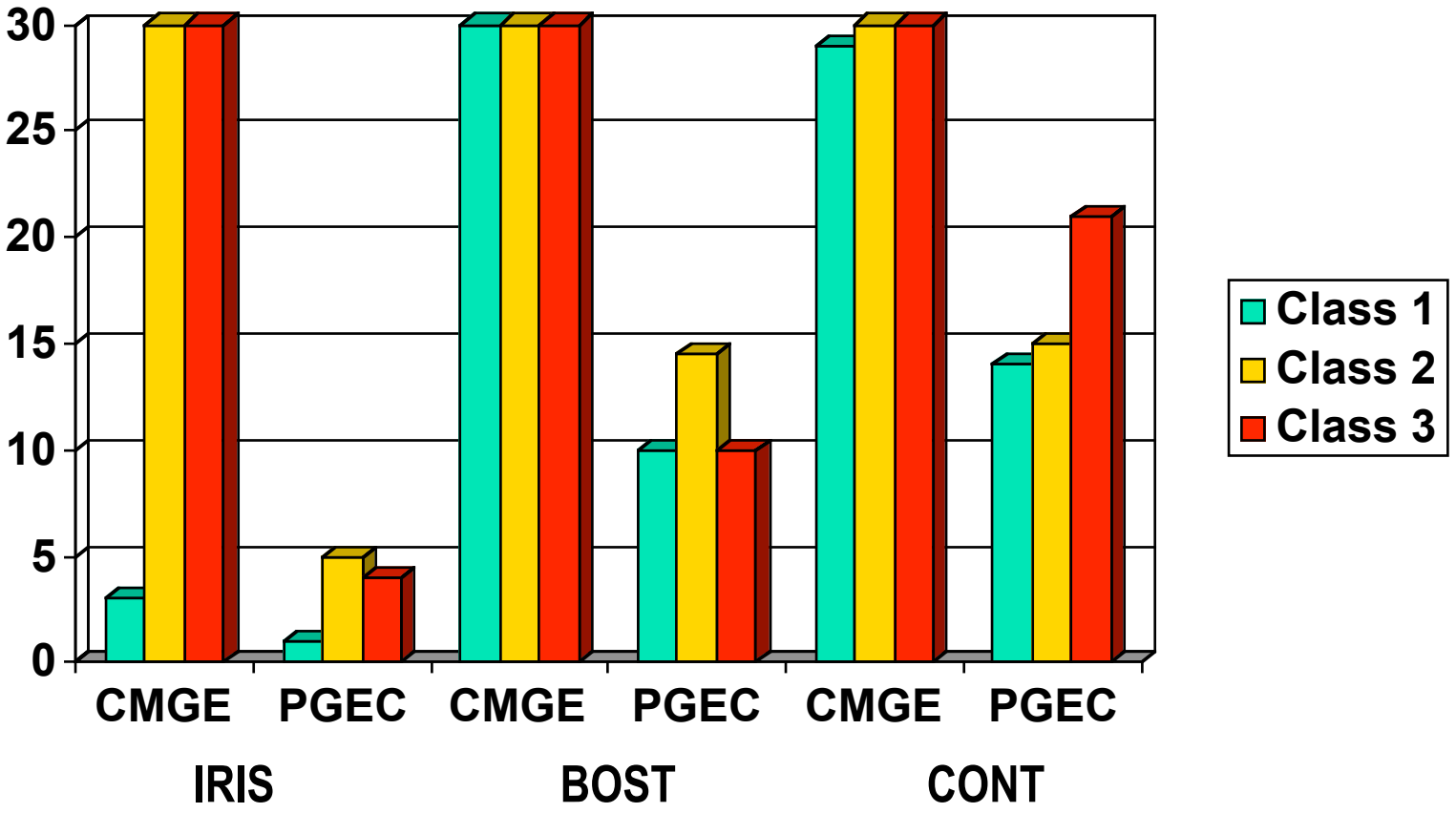




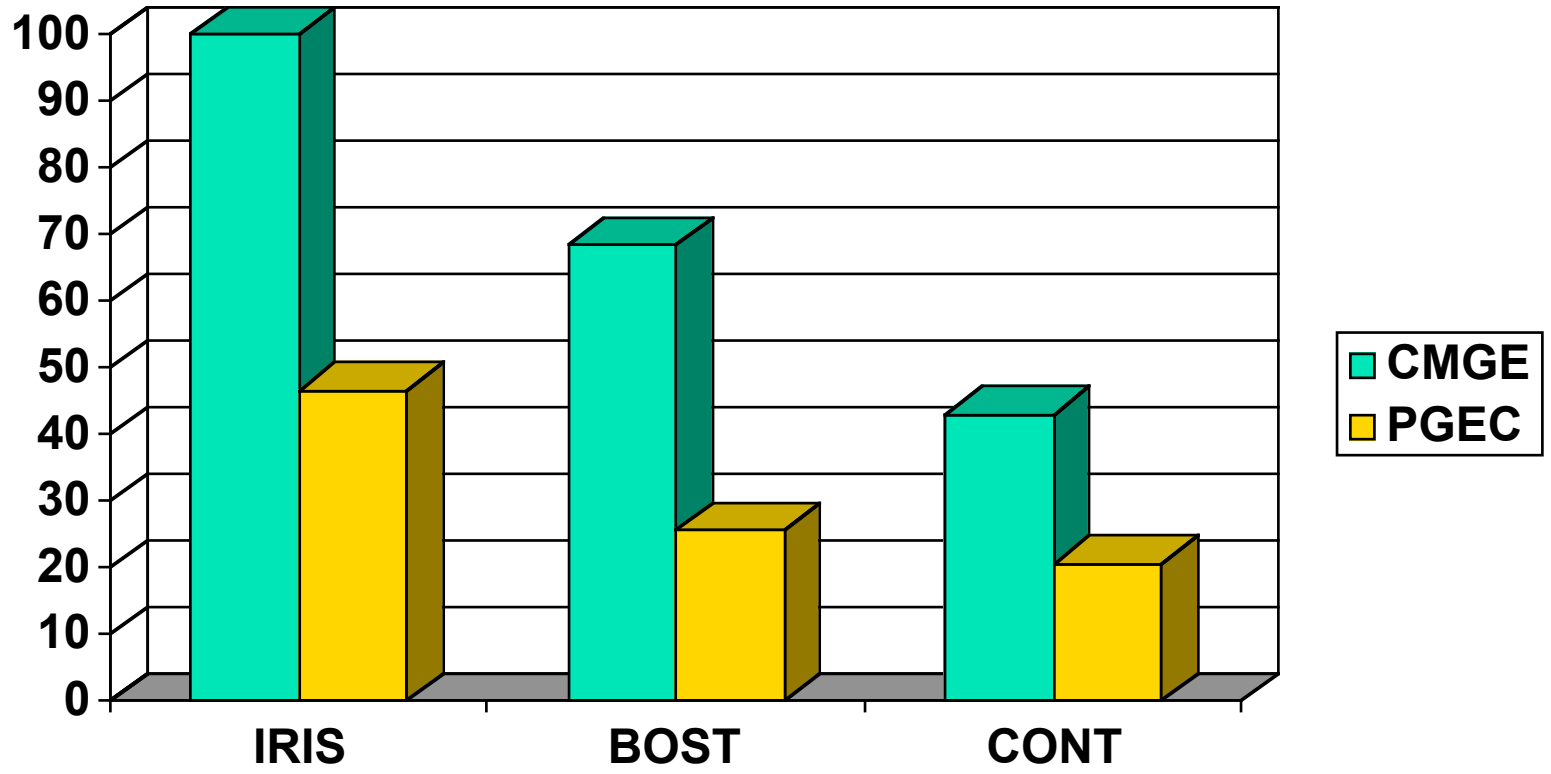
Behavioural Assessment

- PGEC
  - Not cooperative
  - Discrimination
- CMGE
  - Cooperative
  - Novelty Detection
- Classifier ‘Coverage’
  - Compare “Winning” classifier membership to median in-class membership
  - High → unique classifier
  - Low → overlapping classifier behaviours
  - Medium → no clear distinction

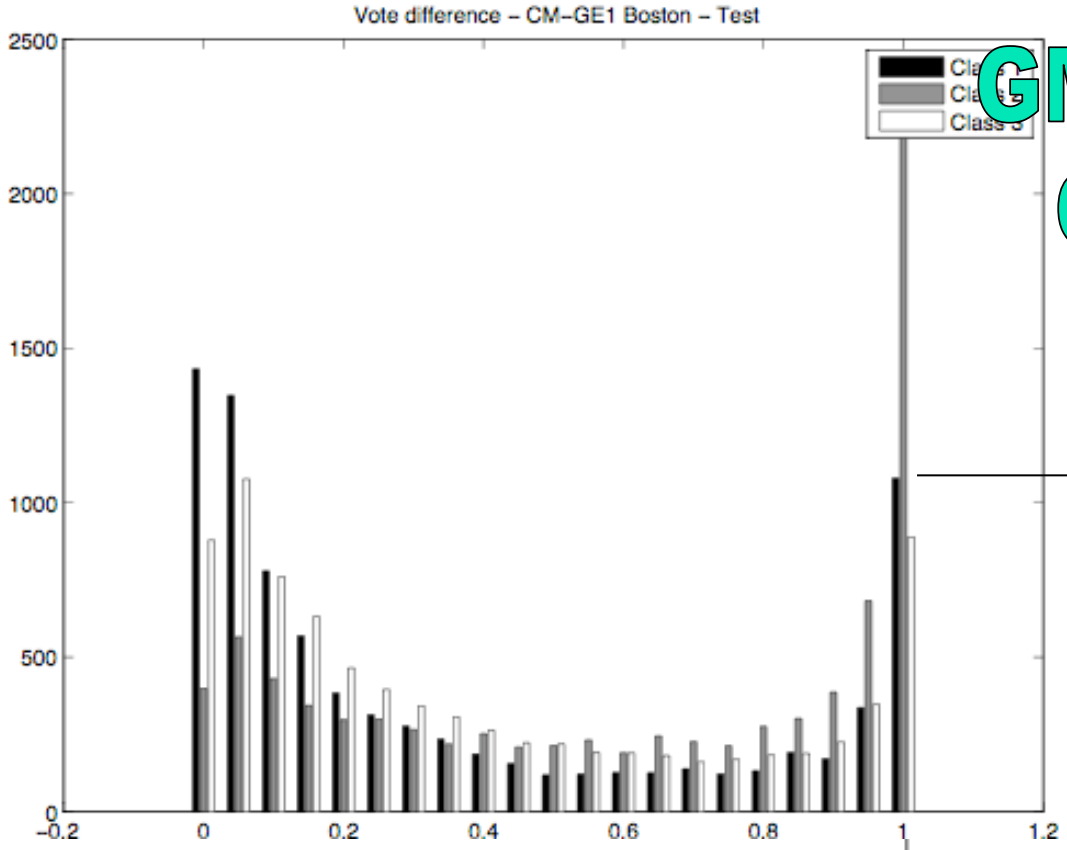
Classifier Count



Median test ‘Score’:
CMGE1 versus PGEC

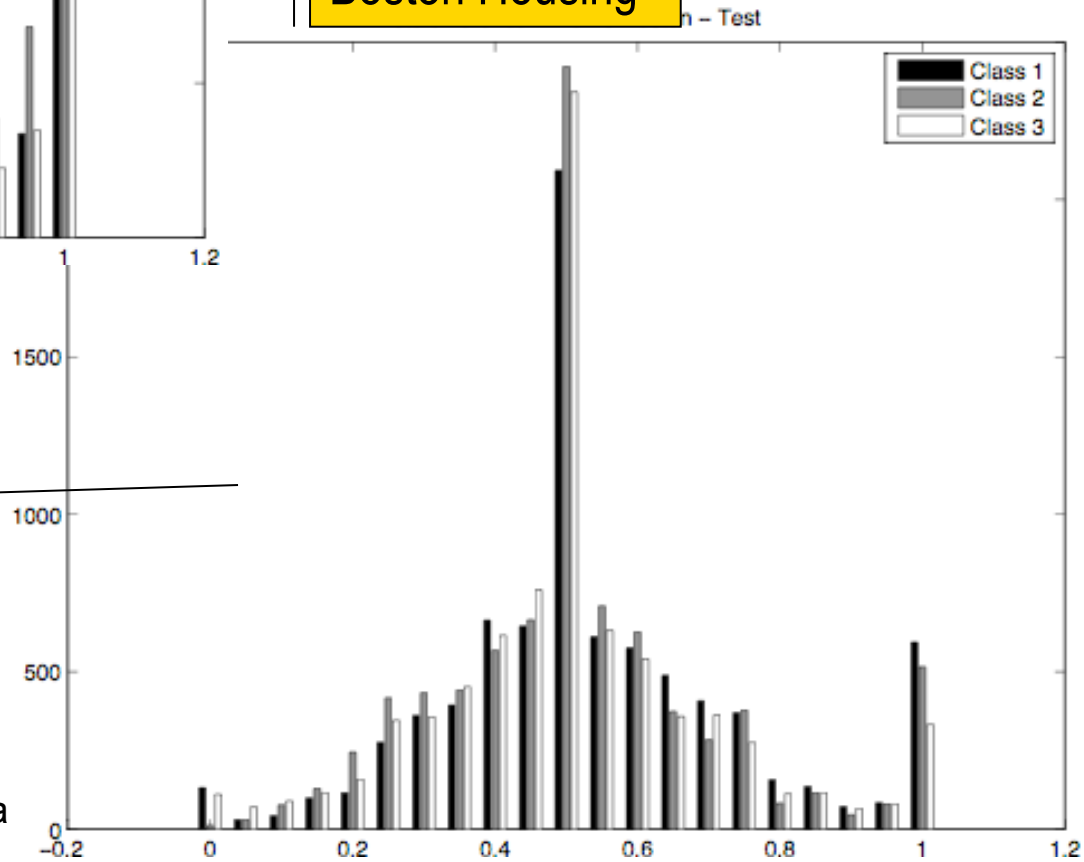


GMGE versus PGEC: Class Membership Difference



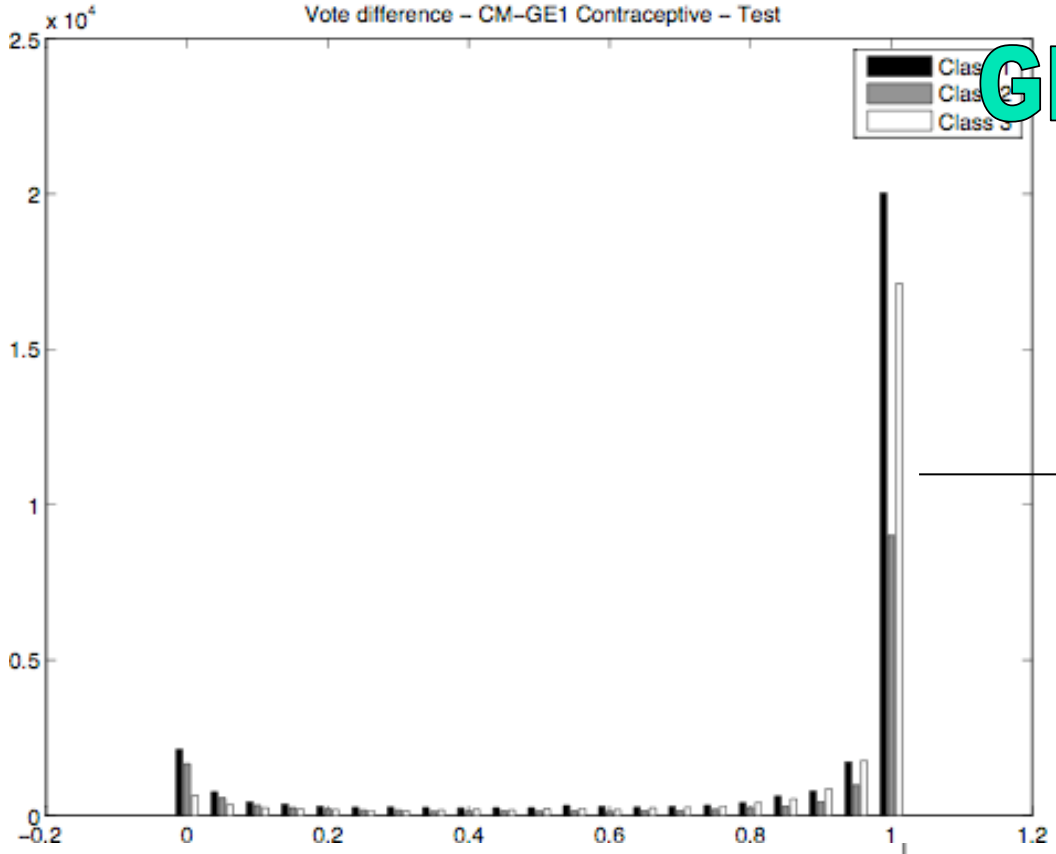
CMGE1:
Boston Housing

PGEC:
Boston Housing



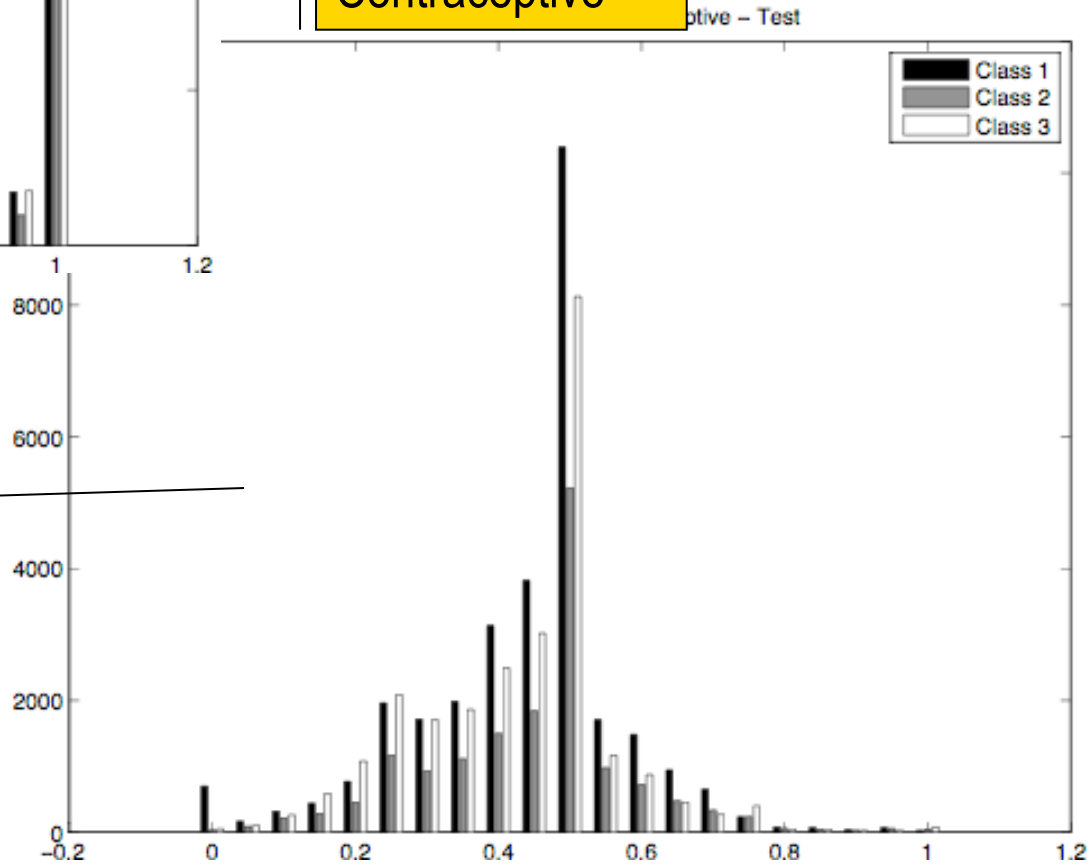
Da

GMGE versus PGEC: Class Membership Difference

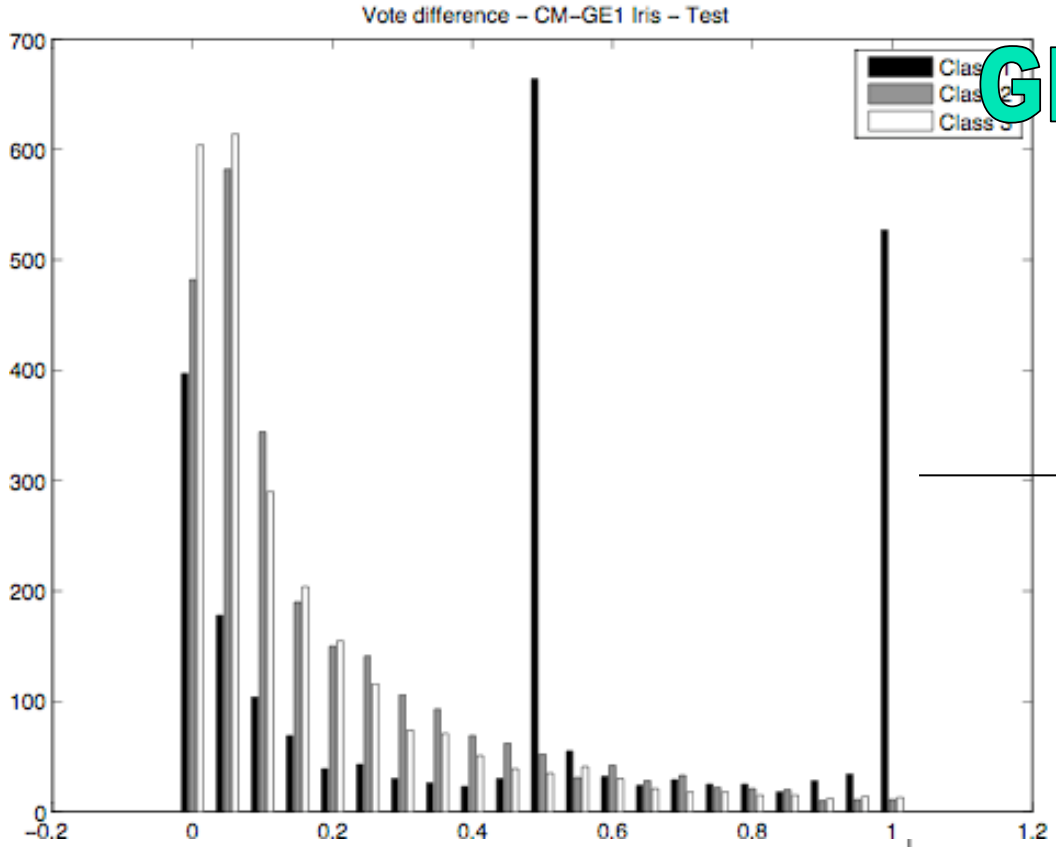


CMGE1: Contraceptive

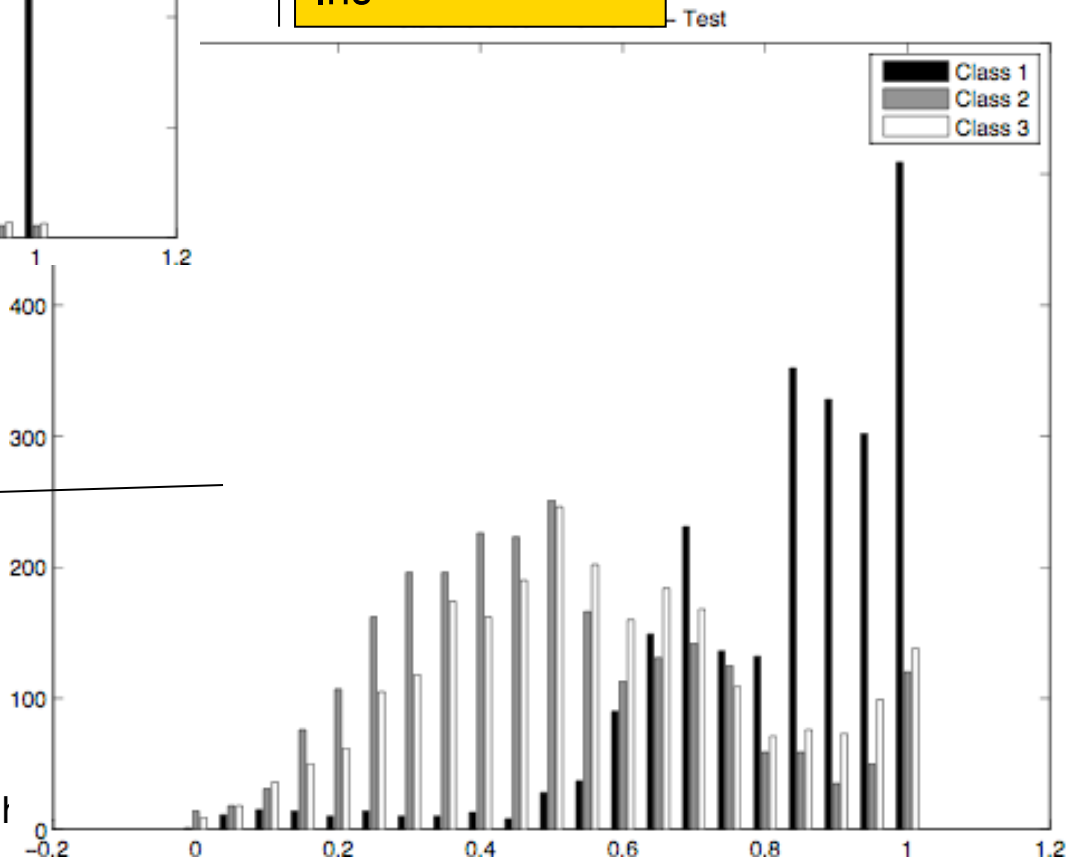
PGEC: Contraceptive



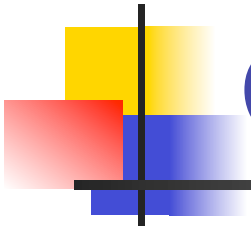
GMGE versus PGEC: Class Membership Difference



CMGE1:
Iris

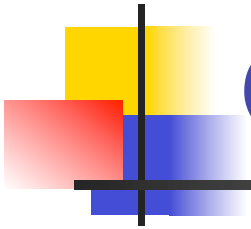


PGEC:
Iris



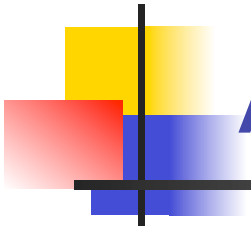
Conclusions

- Cooperative Competitive Model of GP
  - No a priori specification of cooperating individuals
  - Problem decomposition as a learning artifact
  - Competitive coevolution → Scalable
  - Cooperative coevolution → Encourages non-overlapping behaviour



Outstanding Issues

- Locating Test points more effectively
  - Bias exemplar selection
  - Pareto rank
- Bid based cooperative coevolution
  - No clustering step
- One Class learning for GP Classification
  - Novelty based classification paradigm



Acknowledgements

- Andrew McIntyre
  - CMGE cooperative-competitive model
- Michal Lemczyk
  - PGPC competitive model
- Peter Lichodzijewski
  - Bid based cooperative model
- Robert Curry
  - One Class GP Classifier
- NSERC, MITACS, SwissCom, TARA