Montreal Optimization Days May 2005

Misclassification Minimization by Variable Neighborhood Search

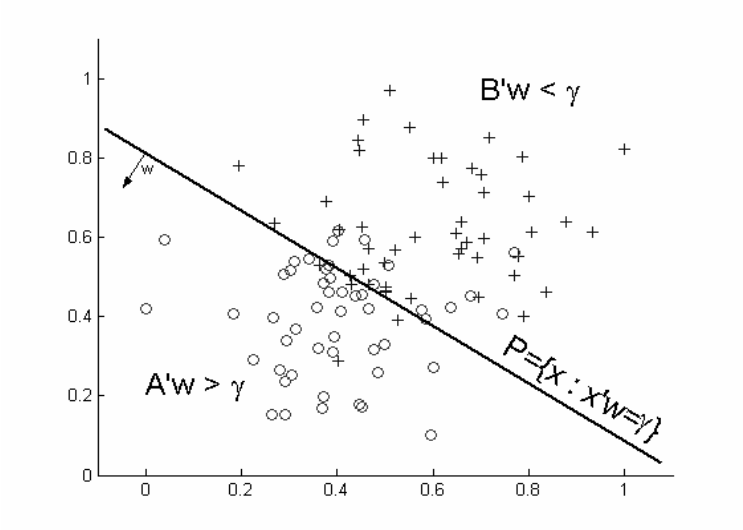
Alejandro Karam (alejandro.karam@hec.ca)
Pierre Hansen (pierre.hansen@gerad.ca)

Today's Talk

- The problem:
  - Linear discrimination criteria
  - Misclassification minimization
- Our approach:
  - Decomposition
  - VNS
- Numerical Results
  - Benchmarks
  - Acceleration of exact solution

Linear Discrimination

Two sets of points, A (o) and B (+), in Euclidean space R^n



Criteria to minimize on training set

- (minus) Likelihoods
  - Logit
  - Probit
- Ratio of variances:
  - Fisher's LDA
- Deviations from central score:
  - Quadratic
  - Absolute

Criteria to minimize on training set

- Distance to plane of misclassified points:
  - L_1 Manhattan
  - L_2 Euclidean
  - L_∞ Max
  - L_p other arbitrary p

$$\|x\|_p = \left(\sum_i |x_i|^p \right)^{\frac{1}{p}}$$

Criteria to minimize on training set

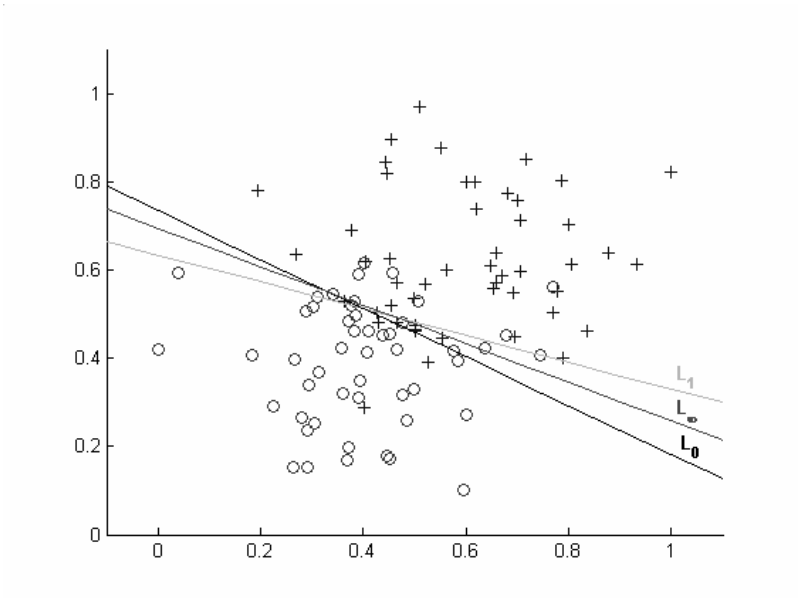
- Distance to plane of misclassified points:
  - L_1 Manhattan — 2n linear programs
  - L_2 Euclidean — non convex quadratic
  - L_∞ Max — MIP with n binary variables
  - L_p other arbitrary p — heuristic

Criteria to minimize on training set

Number of misclassified points

- Sometimes referred to as the L_0-norm
- Several exact and heuristic approaches in the literature
  - Many exact solutions proposed are problematic
- Liittschwager & Wang (1978)
  - Perhaps the first correct formulation

Linear Discrimination



Misclassification minimization

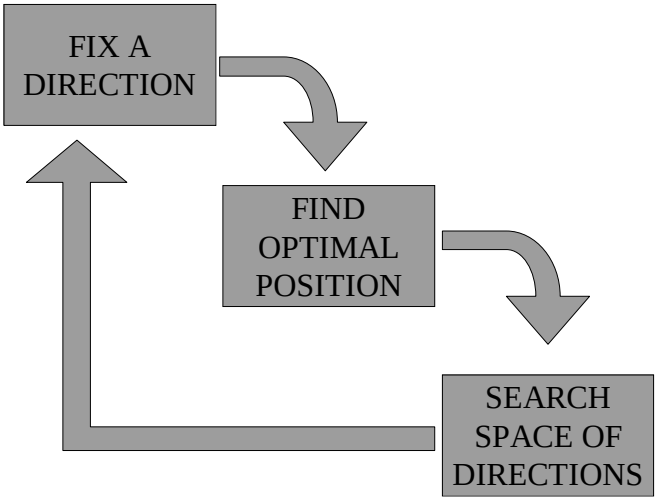
Exact solution is difficult

- MIP with at least k binary variables
  - k is the number of points!
  - (Decomposition and other tricks aside)
- NP-Complete
- Interesting heuristics in the literature
  - Not all scale up gracefully

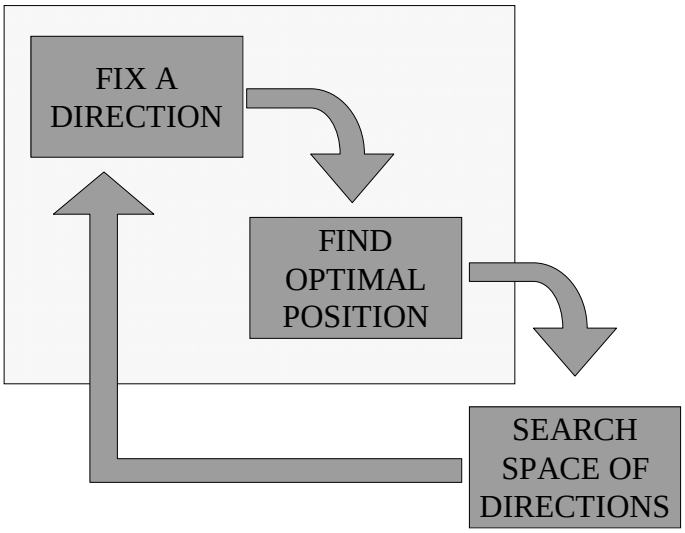
Today's Talk

- The problem:
  - Linear discrimination criteria
  - Misclassification minimization
- Our approach:
  - Decomposition
  - VNS
- Numerical Results
  - Benchmarks
  - Acceleration of exact solution

Solution approach

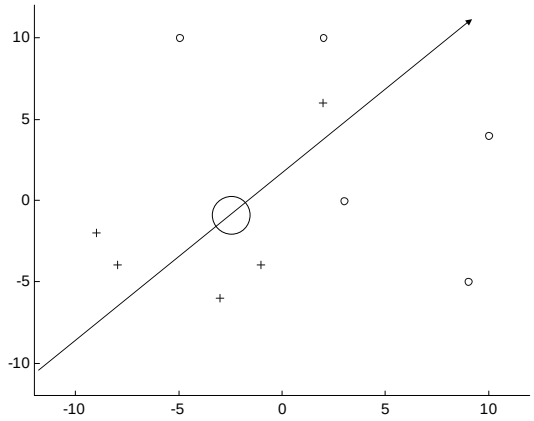


Solution approach



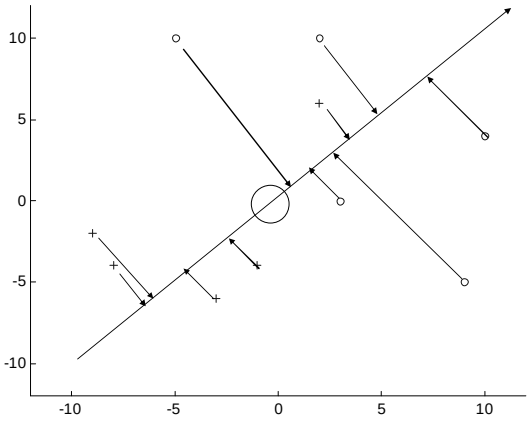
Solution Approach

The decomposition:
- fix a direction
- project points into ray
- find best position for plane



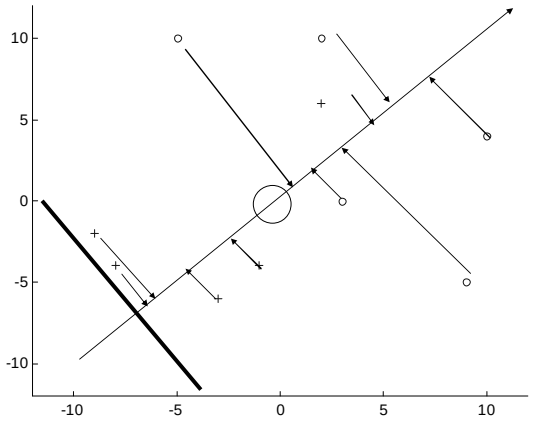
Solution Approach

The decomposition:
- fix a direction
- project points into ray
- find best position for plane



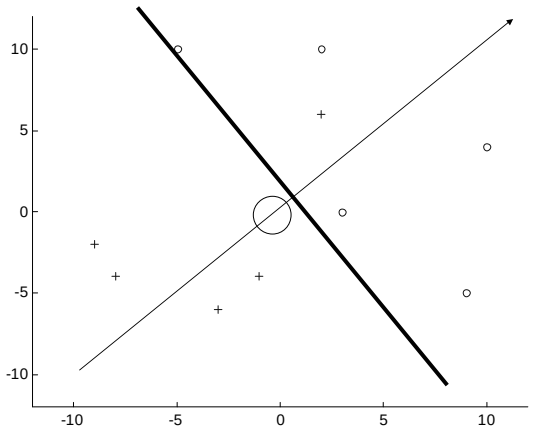
Solution Approach

The decomposition:
- fix a direction
- project points into ray
- find best position for plane

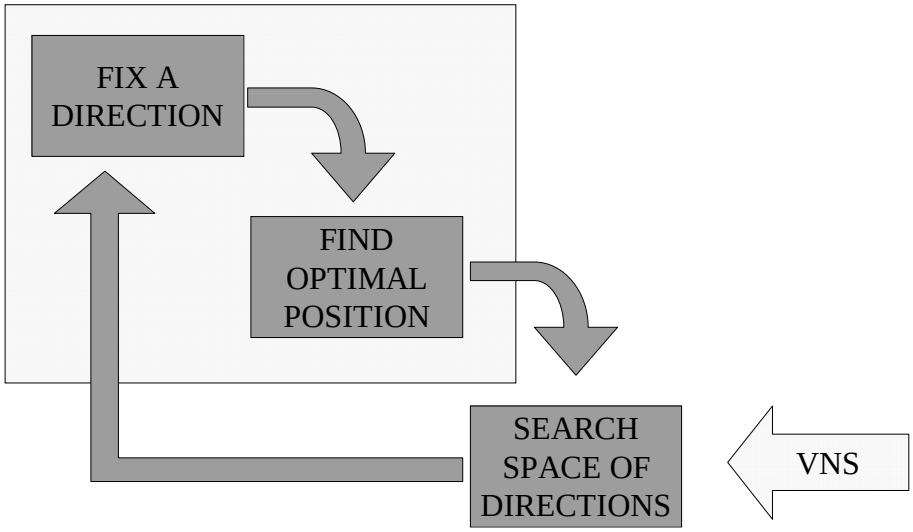


Solution Approach

The decomposition:
- fix a direction
- project points into ray
- find best position for plane



Solution approach



Variable Neighborhood Search

Key ideas:

- Exploit valuable information in local minima
- Random perturbations increasingly "far" from incumbent solution

Ingredients:

- Local descent method
- Distance and neighborhoods on solution space

} problem dependent

Variable Neighborhood Search

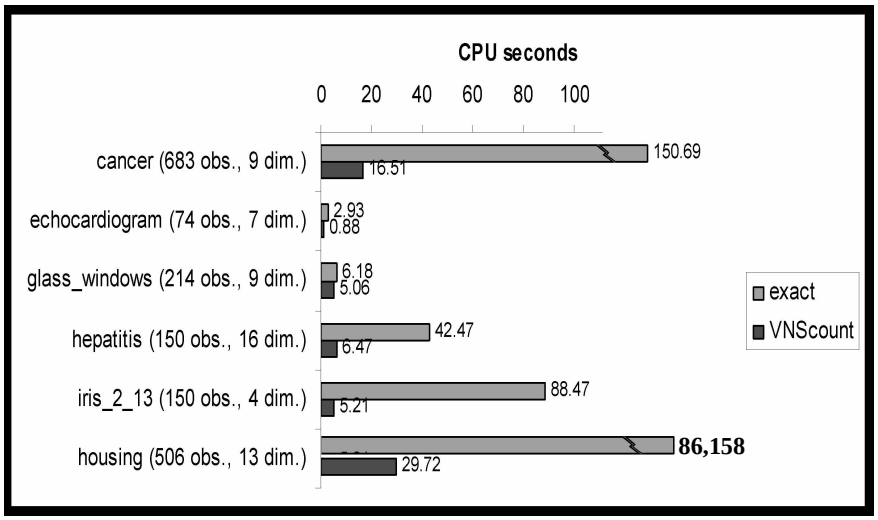
Implementation:

- Local descent method — Downhill simplex method of (Nelder and Mead 1965)
- Distance and neighborhoods on solution space — Difference of embedded in space of angles

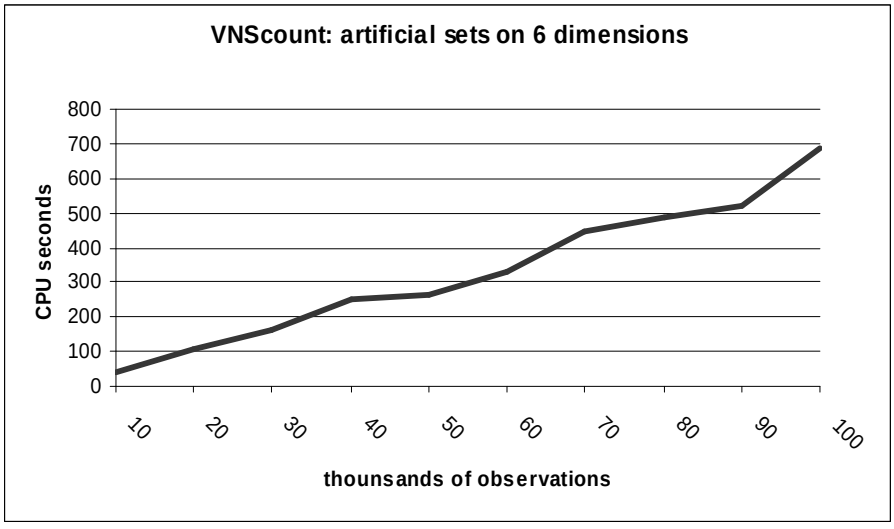
Today's Talk

- The problem:
  - Linear discrimination criteria
  - Misclassification minimization
- Our approach:
  - Decomposition
  - VNS
- Numerical Results
  - Benchmarks
  - Acceleration of exact solution

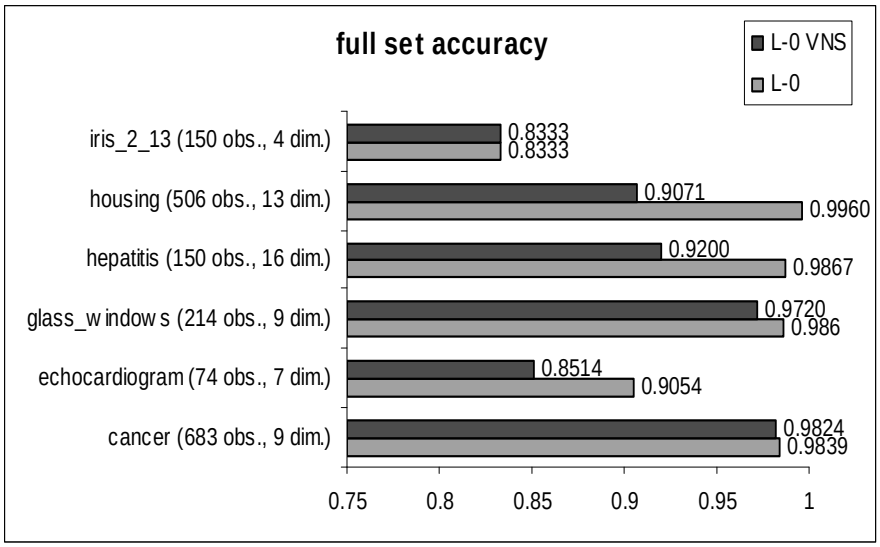
Numerical results: it's fast …



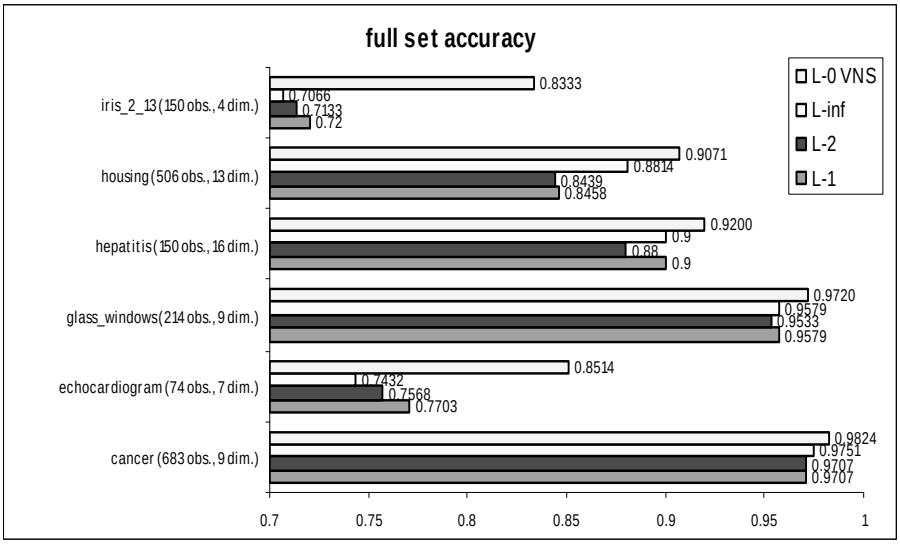
Numerical results: scales up gracefully …



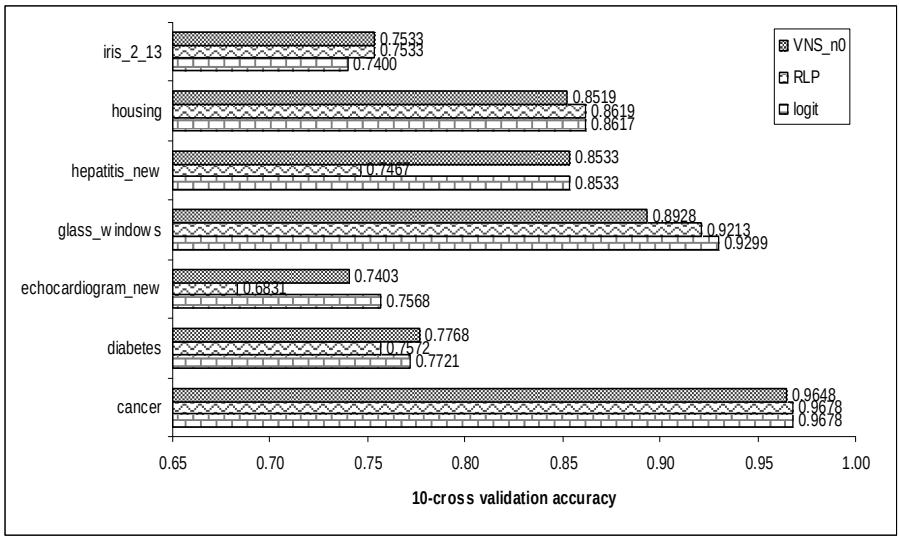
Numerical results: it's not always accurate!



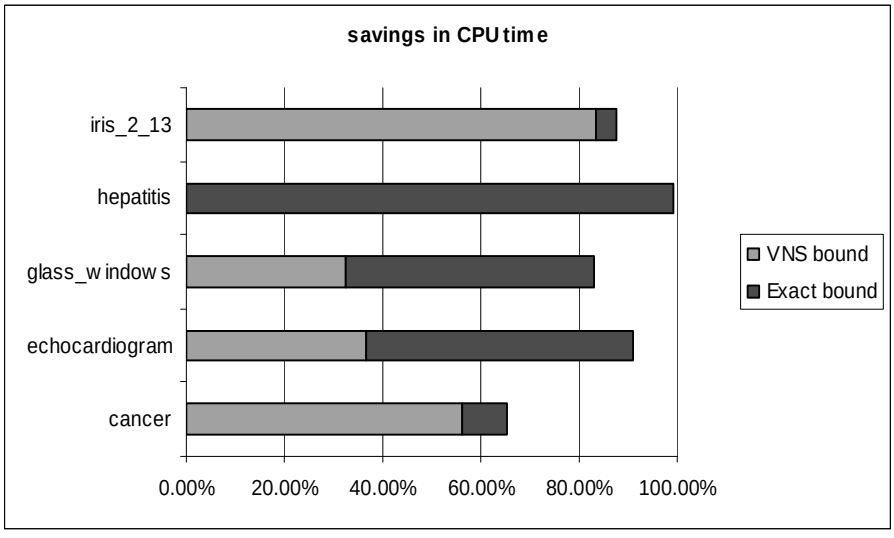
Numerical results: but really not that bad …



Numerical results: … and it generalizes well



Numerical results: acceleration of exact solution



Conclusions

- Fast alternative to exact L_0 solution
- Good generalization properties
- Seems competitive with alternative linear discriminants
- Scales well to large number of observations
- It might be worth a try!