

Validation of MVGA by a toy example

Bowen Deng

Dept. of Prob. and Stat.

Review

Toy Example

Parameter Selection on Original Problem

MVGA

We could generalize the GA algorithm to multiple value case.
Most steps are similar, but the crossproduction is much harder.

Let $k_{\min} = k_{\max} = 2$,

Father:

11223300

Mother:

22331100

Their child should inherit two 1s, two 2s and two 3s. But the child should be as distinct from its parents as possible.

Crossproduct

Denote father: $F = (a_1, \dots, a_n)$, mother: $M = (b_1, \dots, b_n)$,
 $a_i, b_i \in \{0, 1, \dots, t\}$.

$a_i = \rho$ means i -th gene is in the ρ -th set, if $\rho = 0$, i -th gene is not selected.

The motivation is to find a feasible solution corresponding to child
 c_1, \dots, c_n under constraint:

$$\sum_{i=1}^n [(1 - x_i)I(a_i = \rho) + x_i I(b_i = \rho)] \in [k_{\min}, k_{\max}]$$

where $x_i = 0$ represents $c_i = a_i$, and $x_i = 1$ represents $c_i = b_i$.

Moreover,

$$\sum_{a_i \text{ or } b_i > 0} x_i \in \left[\frac{s - c}{2}, \frac{s + c}{2} \right].$$

where $s = \#\{a_i \text{ or } b_i > 0\}$.

We should minimize c (ILP).

Discussion

At first glance, we plug an ILP inside of a genetic algorithm, the computation is expensive.

However, we only need to involve $|F \cup M| \leq 2tk_{\max} \ll n$ binary variables and one integer variable.

We consider the hidden set detection problem:

Suppose there are $S_1, \dots, S_t \subset \{1, \dots, n\}$, we would like to detect those sets without prior knowledge.

We could use a test set M and get its matching score with those hidden sets. For convenience, let $S(M) = n - \min_{1 \leq \rho \leq t} |M \Delta S_\rho|$.

An iterative detection is to find maximizer $\hat{M} = \arg \max S(M)$, deleting \hat{M} , repeat.

The simultaneous version would be $\max(M_1, \dots, M_t) = \sum_{i=1}^t S(M_i)$.

This problem is similar to our task, and we use this toy problem to test the methods.

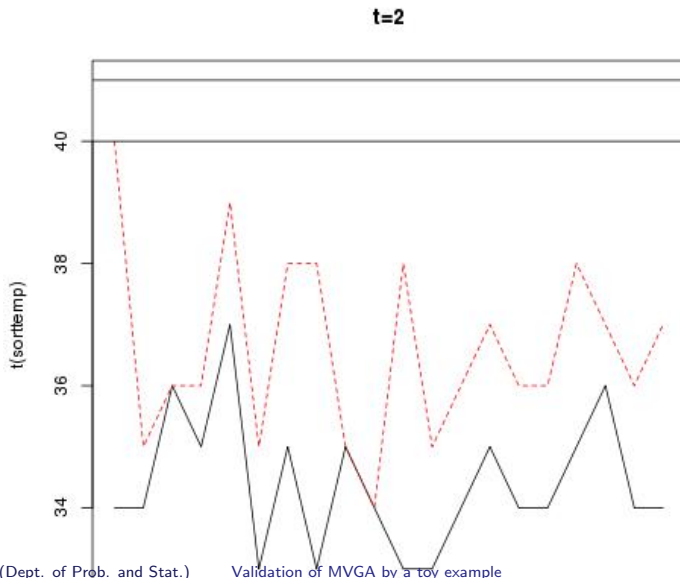
We use MVGA to test whether it converges in this problem.

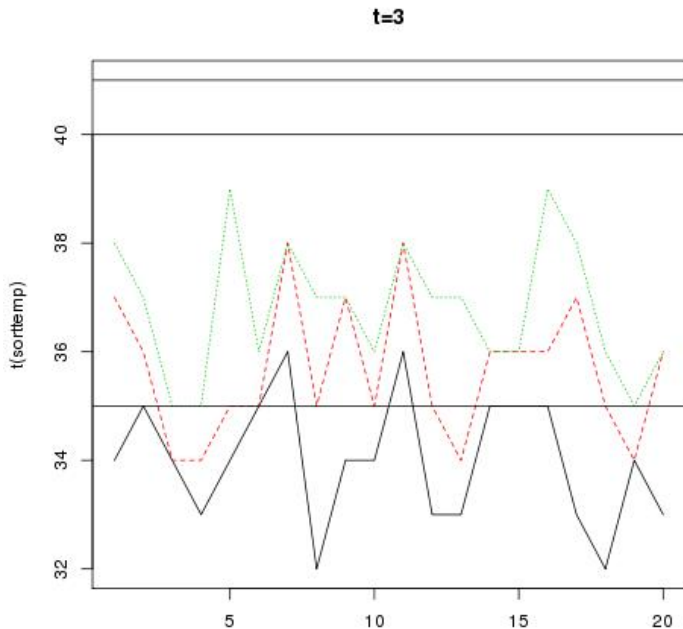
For the sake of computation cost, we let

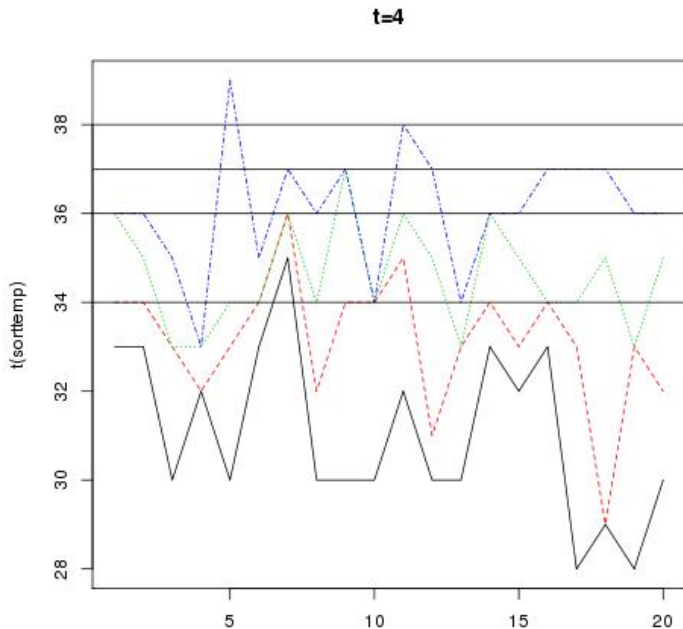
$S_1 = \{1, \dots, 5\}$, $S_2 = \{6, \dots, 10\}$, $n = 20$. And we find M_1 , M_2 simultaneously using MVGA.

Population	Iteration	Score	Time(s)
100	1	34	52
	10	37	135.41
	20	39	206.07
10	1	32	12
	10	34	10.19
	20	33	44.97
50	20	38	115.43

We use permutation test to identify the best t .
For simulation data with $t = 2$.







loose to $0 \leq x \leq 1$.

$$u = f(u)$$

lars