Evolutionary Methods on Synthetic Data

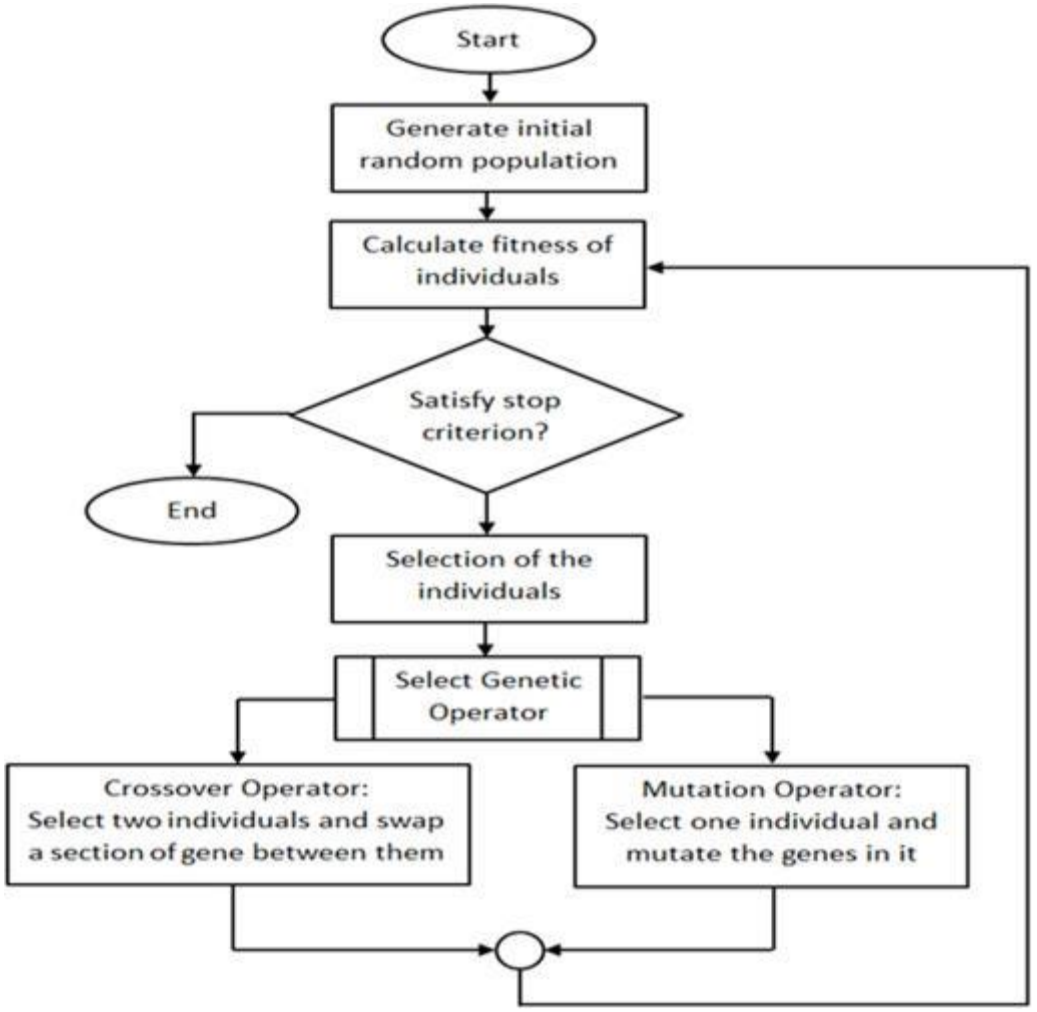
Yingrui Chen, Mark Elliot, Joe Sakshaug

University of Manchester

Evolutionary Methods

- Artificial Intelligence
- Machine learning
- Key Features:
  - A set of criteria (or objectives) to optimise
  - A set of passible actions
    - e.g. Mutation, crossover, selection.
  - Selection mechanism
  - Iteration
  - Able to improve on the basis of an evaluation against the citeria

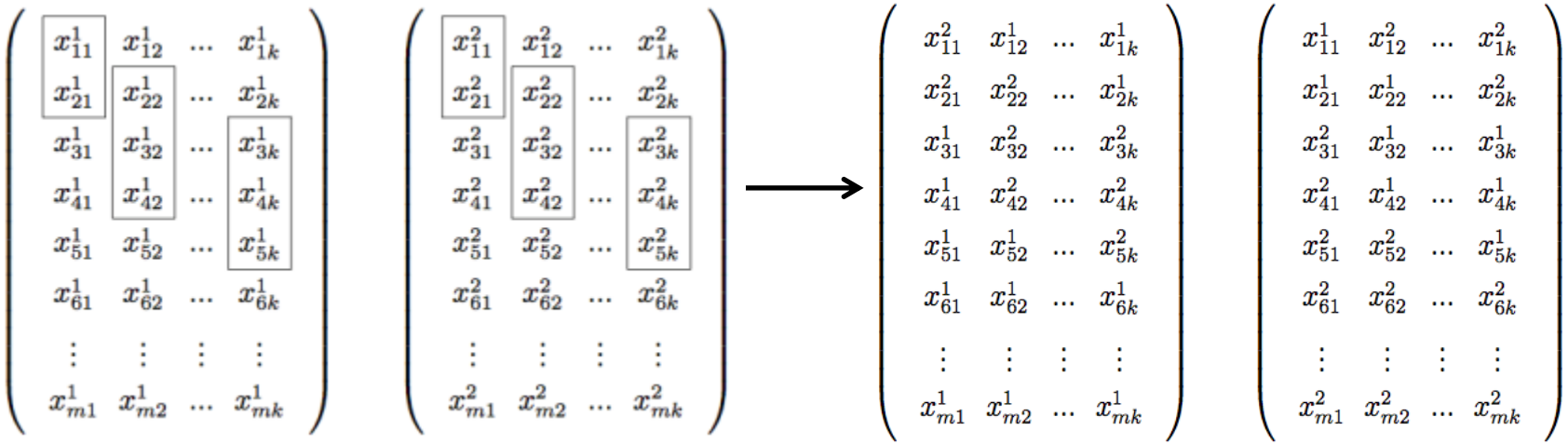
Genetic Algorithms



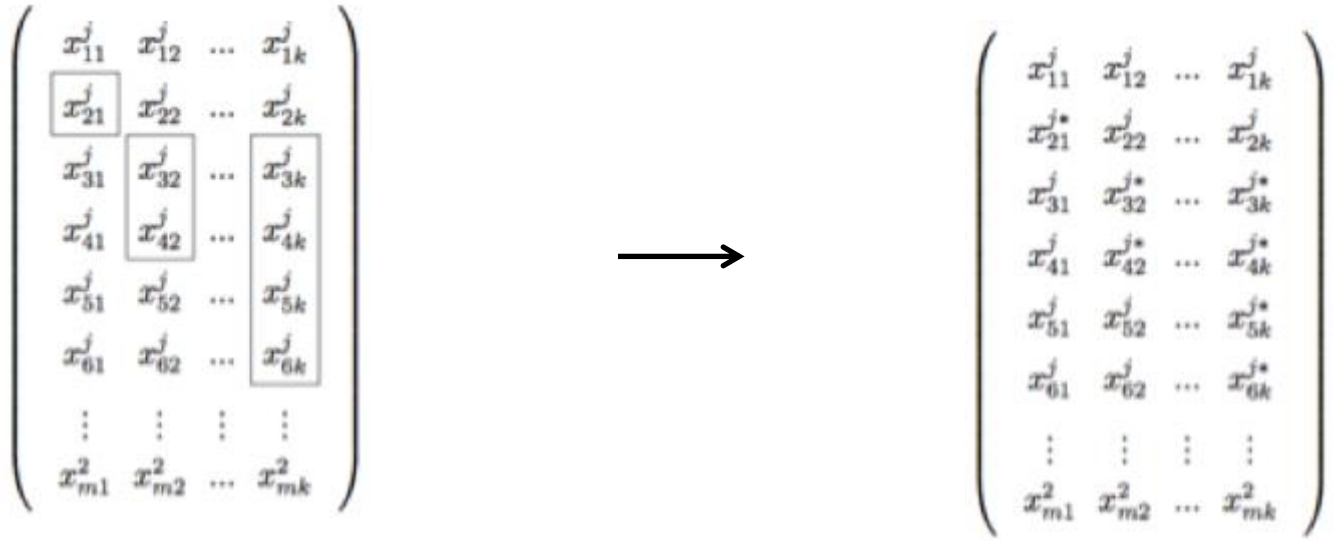
Wahab,.M, Nefti-Meziani, S., and, Atyabi, A. (2015), A comprehensive review of Swarm Optimisation Algorithms, *PLoS one*, 10(5)

- Objectives/Criteria:
  - selected statistical properties
- Initial population:
  - A group of synthetic datasets that are generated using the univariate distributions of the original dataset.
- Selection mechanism:
  - ranking based selection scheme.

- Crossover operator:



- Mutation operator



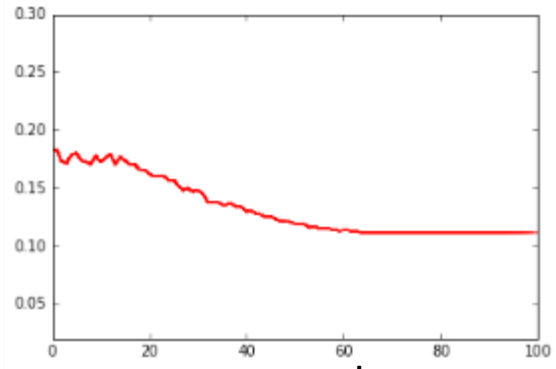
Pilot Experiments

- Dataset: 1000 cases, 4 variables, categorical, single-level microdata
- Single objective: Mean of differences between synthetic and original datasets in values of Cramer's V for all bivariate cross-classifications:

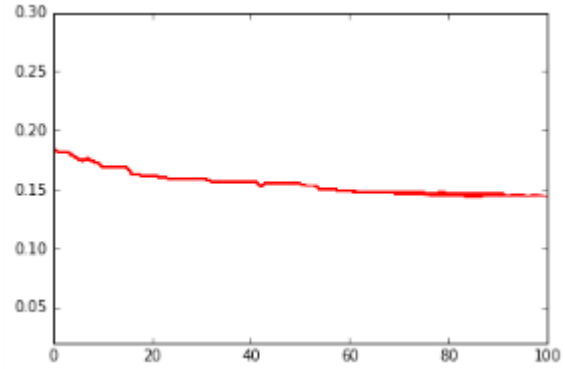
$$\mathrm{F} = \binom{4}{2}^{-1} \sum_{i=1}^{3} \sum_{j=i+1}^{4} \left| \phi_c\left(x_i', x_j'\right) - \phi_c\left(x_i, x_j\right) \right|$$

Results

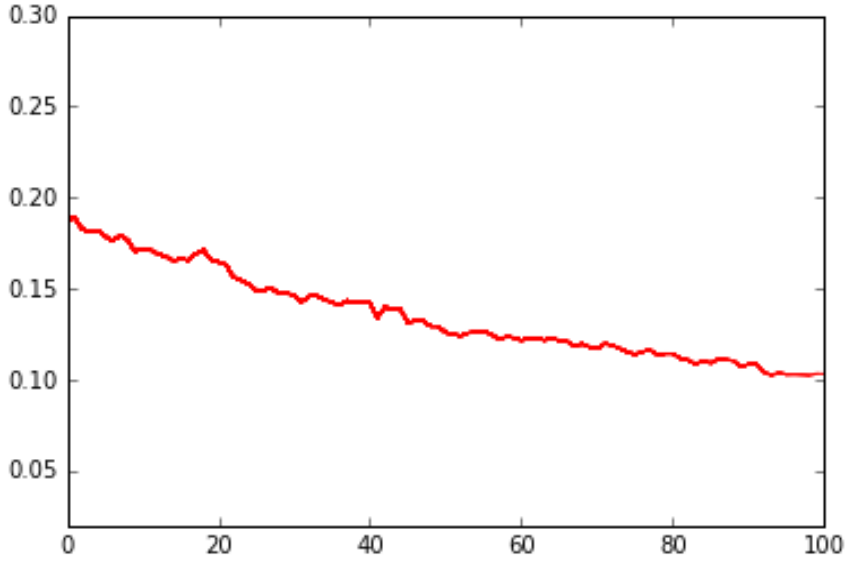
Minimum divergence (best fitness) values
in current iteration vs. number of iterations



crossover only



mutation (mutation rate=0.1) only



General GA model

Next steps

- More crossover/mutation mechanisms
  - Randomly select attributes to crossover and mutate.
  - Adaptive crossover/mutation.
  - Targets: Not forcing all candidates into evolution, trying to keep more features from good parents.
- Multi-objectives
  - Other analytical properties (e.g. the full hierarchical equivalence structure).
  - Disclosure risk measures.