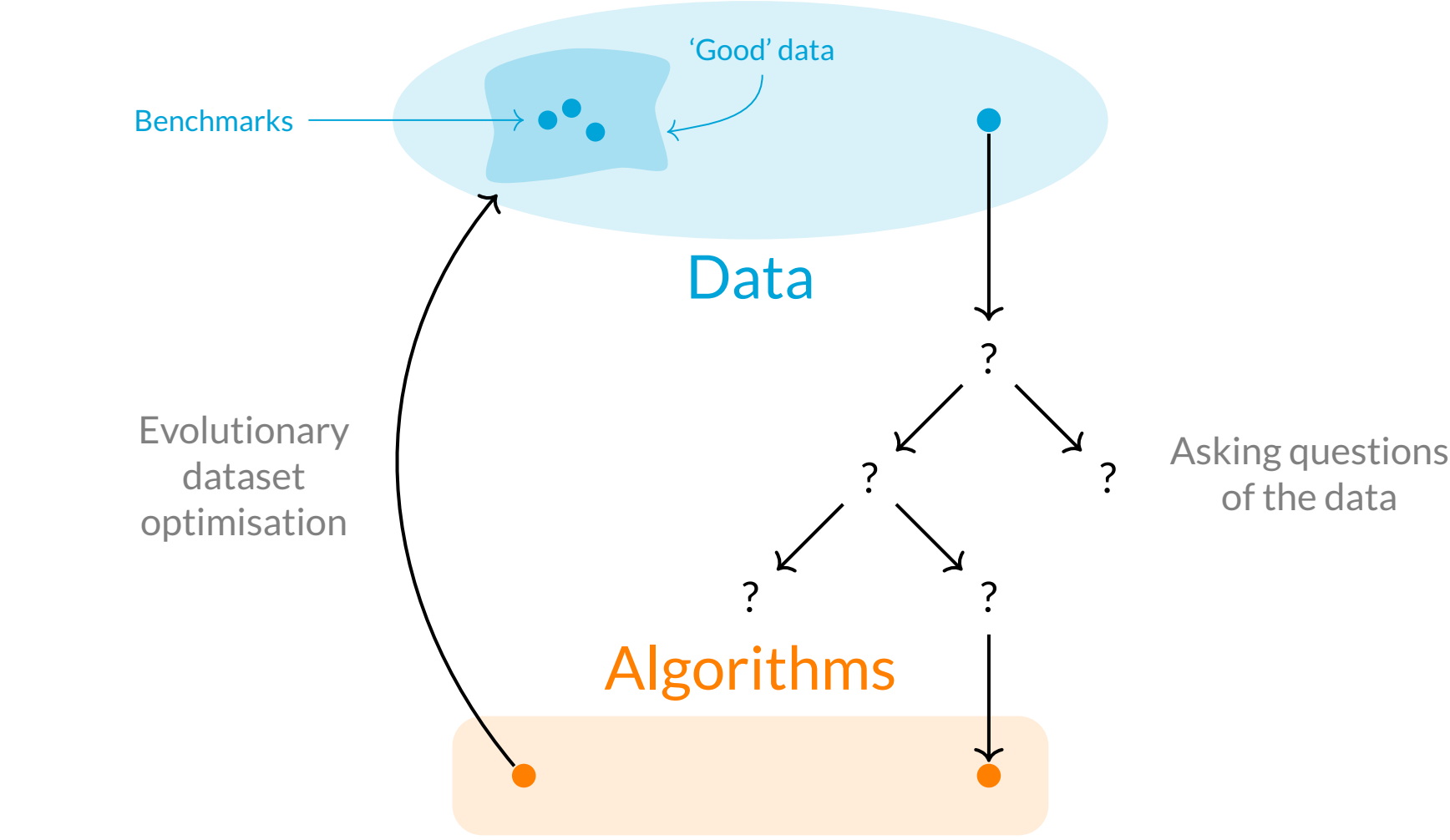


Evolutionary dataset optimisation: learning algorithm quality through evolution

Henry Wilde, Vincent Knight, Jonathan Gillard

Motivating paradigm

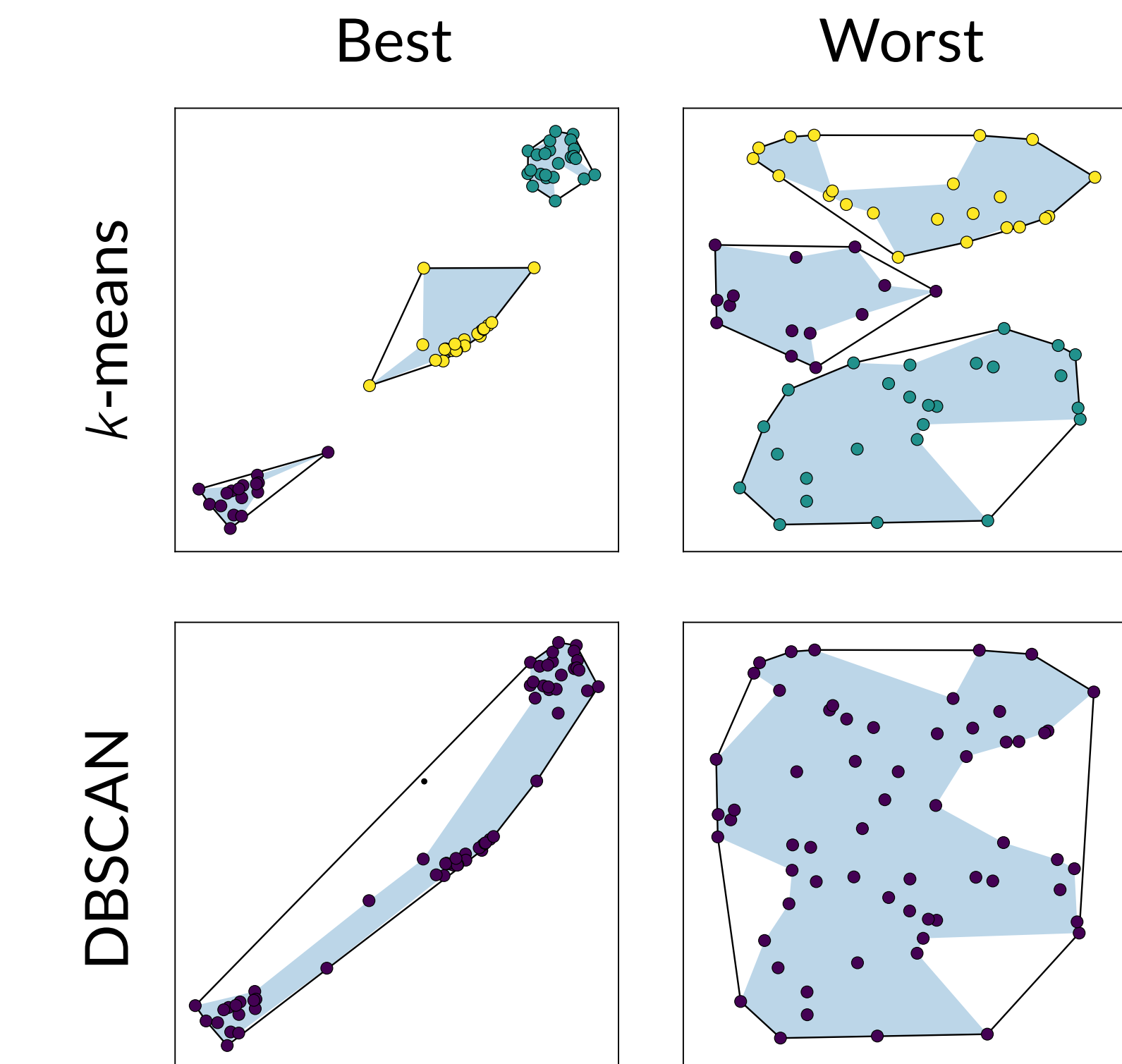


Left: the proposed concept to understand algorithm performance by exploration of ‘good’ (or bad) datasets. Right: the established approach to algorithm selection by metric comparison.

Using an evolutionary algorithm

- Effective in complex domains
- Meaningful and adjustable design
- Transparent and rich solutions
- Potential for suboptimal termination
- Finds the ‘easy’ way out
- Requires careful consideration of fitness

Understanding performance of clustering algorithms



Best and worst datasets from an attempt to maximise the silhouette of *k*-means (top) whilst minimising that of DBSCAN (bottom).

Benchmark datasets are not the only option when understanding algorithm performance



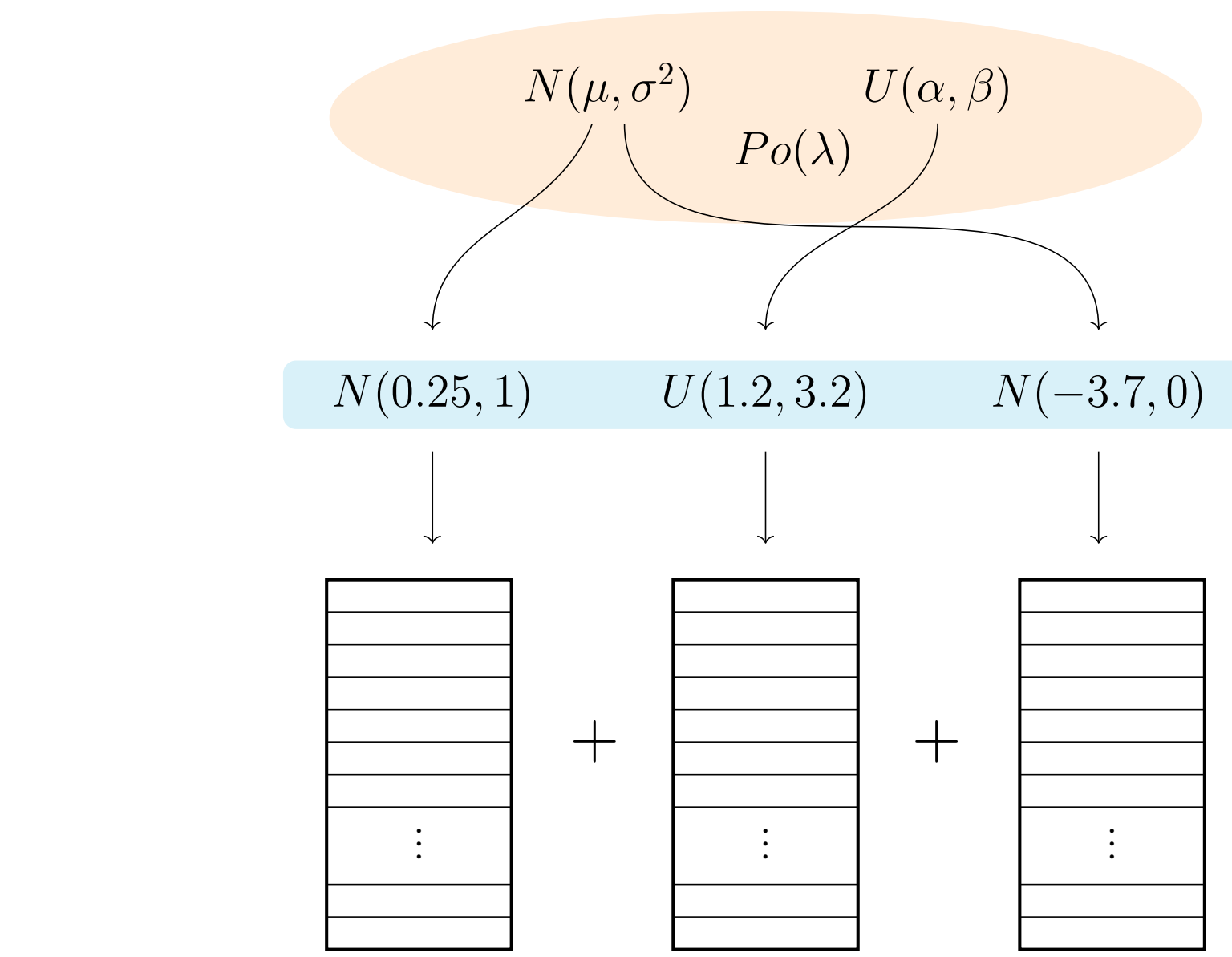
Take a picture to download the full paper

[daffidwilde](#) [edo.readthedocs.io](#) [arxiv.org/abs/1907.13508](#)

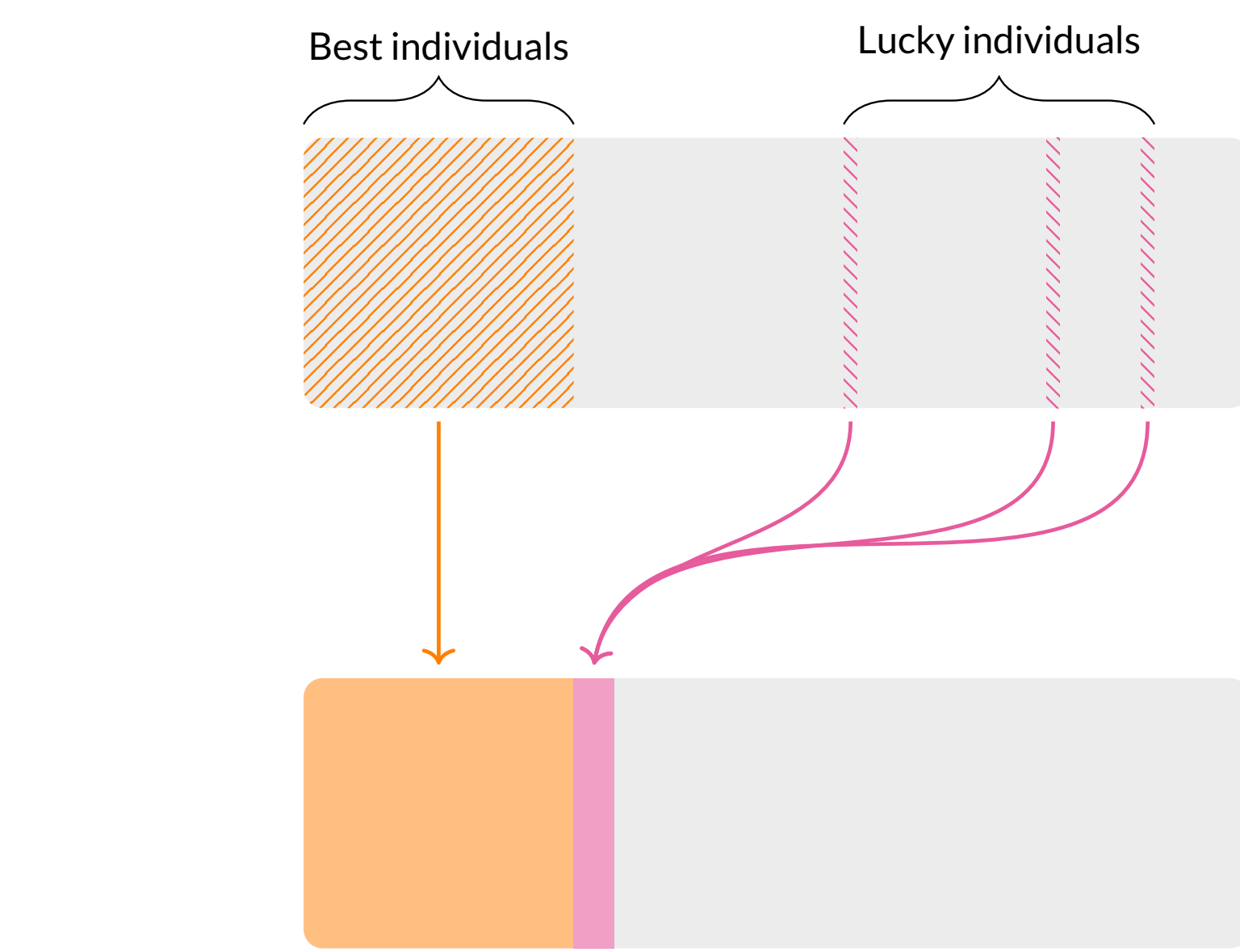


Inner mechanisms

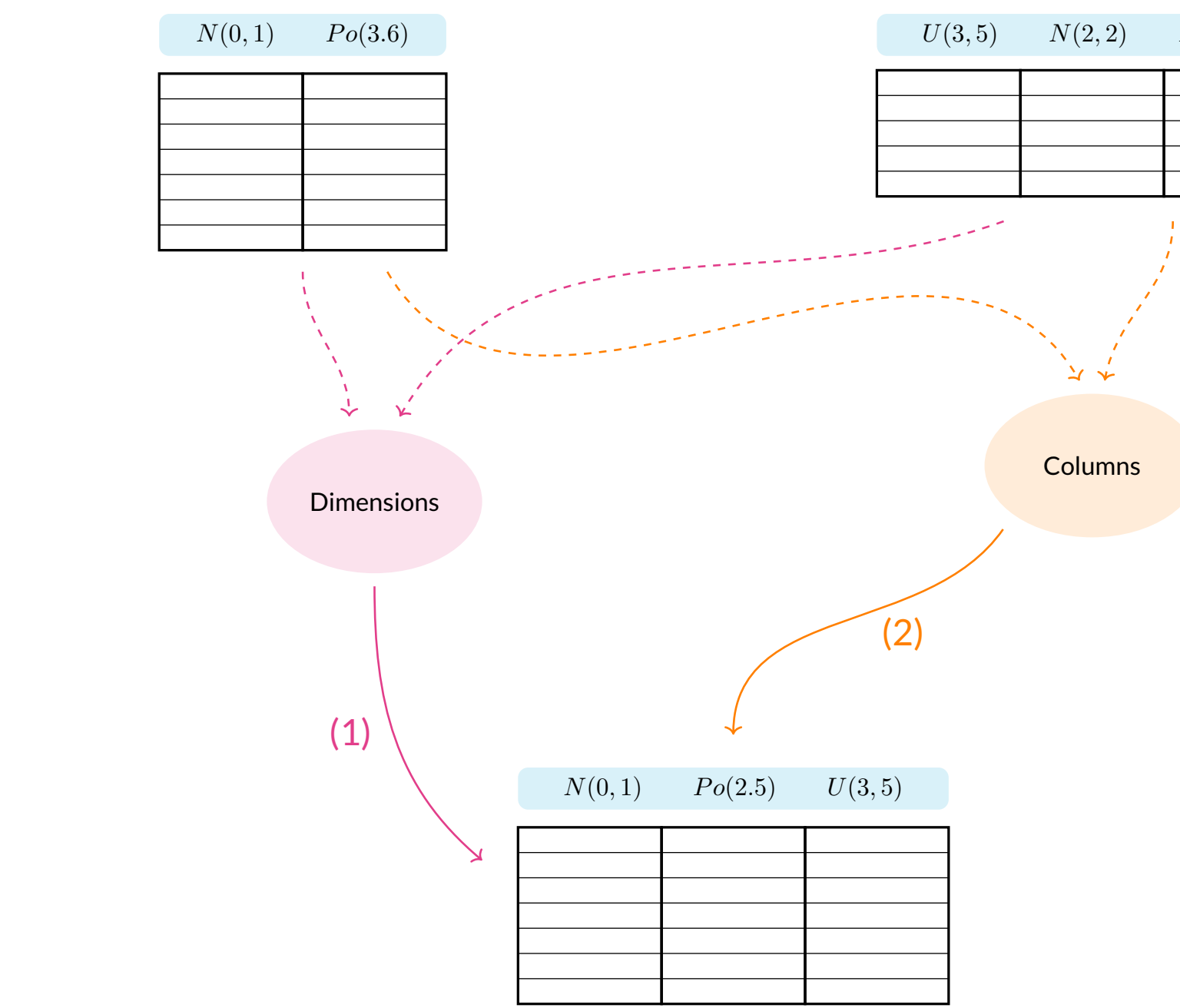
Individual dataset representation



Modified truncation selection



Uniform-sampling crossover



Mutation process adapted for datasets

