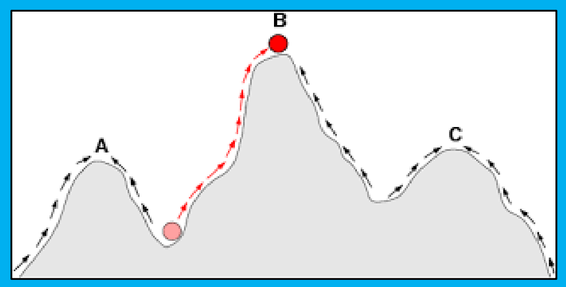
A Survey and Analysis of Benchmark Fitness Functions

Dypsyl Dlamini

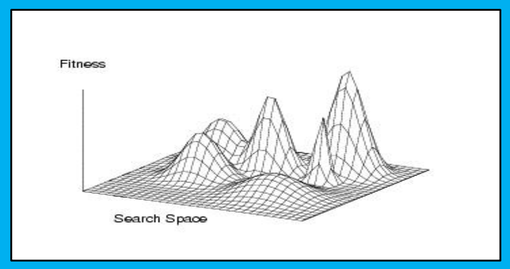
MSc in Big Data

UNIVERSITY of STIRLING

Heuristics are used to solve many real world problems. Binary input vectors are mapped via a fitness function to a single output, and from the set of all possible input vectors the task is to find the one which gives the best output. Where this set of input vectors is small, it is possible to evaluate them all. However this is often not the case. A heuristic attempts to find the best output value by searching the set of all possible input vectors in such a way as to avoid having to evaluate each and every vector in the set.

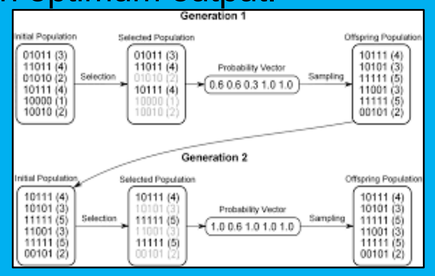


The simplest heuristic is a hill climb. A starting input vector is chosen and each variable in turn is altered. If this change leads to an improvement the change is retained, if not it is discarded. The hill climb is challenged when there are local optima which are lower than the global optimum. Depending on the starting input vector the search may lead away from the global optimum and it may get stuck on a local optimum.

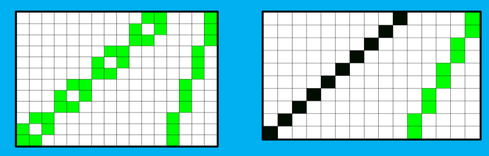


Genetic algorithms search by recombination between two input vectors, then a mutation in the resulting products; and then only retaining the best outcomes. As this process is repeated, the optimum is found in the population of retained input vectors.

Estimation of distribution algorithms guide the search by identifying characteristics in the input vectors and mapping these to a probability distribution, then sampling the distribution to identify further input vectors. The best of these are again mapped to a new probability distribution, this process being repeated to find an optimum output.



Fitness function modelling builds a model of the fitness function and uses this to perform the search. Here a multiple order hyper network with a single output node, several input nodes and no intermediate nodes was used. Weights are assigned to each link between the inputs and the output, as well as to each combination of input nodes with the output node. The structure of the model is discernible



Given that fitness function models have discernible features, is it possible to use this to classify fitness functions in such a way that a suitable algorithm can be identified?

This Project will look at a number of benchmark fitness functions and generate the corresponding fitness function model. The optimum solution for each problem will then be identified using each of the search heuristics described. The objective of these experiments is to investigate whether the fitness function model can be used to classify which heuristic could be most economically applied.