2 Genetic Algorithms

Genetic algorithms borrow optimisation strategies from those found in nature. One of the main benefits of genetic algorithms is the flexibility in terms of categorical, discrete and continuous variables. GAs are naturally suited to the creation of novel topologies[4] [25], and can often produce more targeted, specialised architectures that deviate from the generalised templates that are commonly used due to their effectiveness across a range of domains. GAs have also been used on a more granular level, creating novel variants of network components suited to a specific task [25]. Evolutionary optimisation is composed of two main processes, Mutation and Selection. Mutation in this application broadly describes an operation which alters the traits, or creates new members of a population. Selection conversely, is a process which contracts the population based on some criterion. This is commonly some fitness evaluation on the objective function.

2.1 Selection

The selection process in a GA defines much of the character of the overall system. Due to tendency of naive evolutionary systems to become trapped in local optima, care must taken to maintain diversity in the population. The selection process can take place prior to the mutation of population or subsequently, reducing the expanded population back to its original size. This can change the expression of the overall system significantly as the former can result in a greedier search, however, the latter generally significantly larger computational costs, warm-starting or surrogate evaluation methods sub-sample the expanded population [23], [34], [37].

Truncation Selection, also known as Elitist Selection, is a simple approach to selection where the population is ranked by a criterion and all members below a threshold P are eliminated. This process is shown in figure 1. Tournament selection is a less greedy approach to selection where the population is randomly sampled for a batch of members S. An evaluation is then done within this subset with an approach similar to elitist selection where the highest performing member of the sample is used to create an offspring which replaces the lowest performing. As the sample size grows so does the overall selection pressure of the process as samples are more likely to contain higher performing individuals. This system is described in Algorithm 1.

A variant of tournament selection was used by [3], in which the population is selected against based upon the "age" of the network. This approach was used to set a new SOTA, performing at scale on ImageNET. This method maintains a population of networks along with a list which retains the age of each network ranking each from oldest to youngest based on when they were added to the population, as shown in Figure 2. At each iteration a sample is draw from the population and the highest performing model is used to generate a new network. This new network becomes the youngest in the population and replaces the oldest network which is removed.