Machine Learning 2

CM50265

Genetic Algorithms

- What are they?
- How do they work?
- How are they used in ML?

What are they?

- They are a search method to find optimal solutions to a problem.

- They are based on the biological process of natural selection/evolution.

Natural Selection

- Survival of the fittest.
  - The strongest survive.
  - And breed.
  - And therefore their strength is passed to the next generation.

Genes

- This is done through genes.
- They encode ‘fitness’.
- Offspring receive a mixture of parents genes.
- Some offspring will be an ‘improvement’ others will be less successful.

Moving Target

- The meaning of ‘strongest’ can change over time.

- And natural selection will adapt to the new target.

How do they work?

- Our solution to a problem must be expressible as a ‘genome’.
- And we need a way to measure ‘fitness.’
  - How successful is a given solution.

Process

- Initial Population
- Fitness Function
- Selection
- Crossover
- Mutation.

Initial population

- We start with a population of a reasonable size.
  - ‘Reasonable’ depends on the problem.
  - Typically we keep the population size constant.
- Typically randomly initialised.

Digital Chromosome

- Our candidates all have a number of genes that encode their function.
- This is referred to as a chromosome.
- An array of values.

Fitness Function

- The fitness function evaluates each individual in the population.
  - Each gets a fitness score.
  - The higher the better.
  - Their probability of survival is based on their score.

Selection

- Select the strongest (fittest) to pass their genes on to the next generation.
  - A pair of individuals are selected (the parents) based on their fitness score.
  - The individual with the highest score have more change of being selected.

Crossover

- How do we combine the two chromosomes?
  - In theory we could just randomly select from each pair of genes at random.
  - However, a more constrained approach is used...

Crossover

- Select a random point in the chromosome.
- Switch all the genes between the pair up to the crossover point.
  - This preserves more of the inter-relationship of the genes that a random mixture.

Mutation

- If crossover was all we did at each step we would introduce a problem...
  - A lack of diversity.
- To maintain diversity we introduce mutations into the genes.

Mutation

- With a low probability, change one or more gene values of the new individuals.

- This prevents premature convergence.

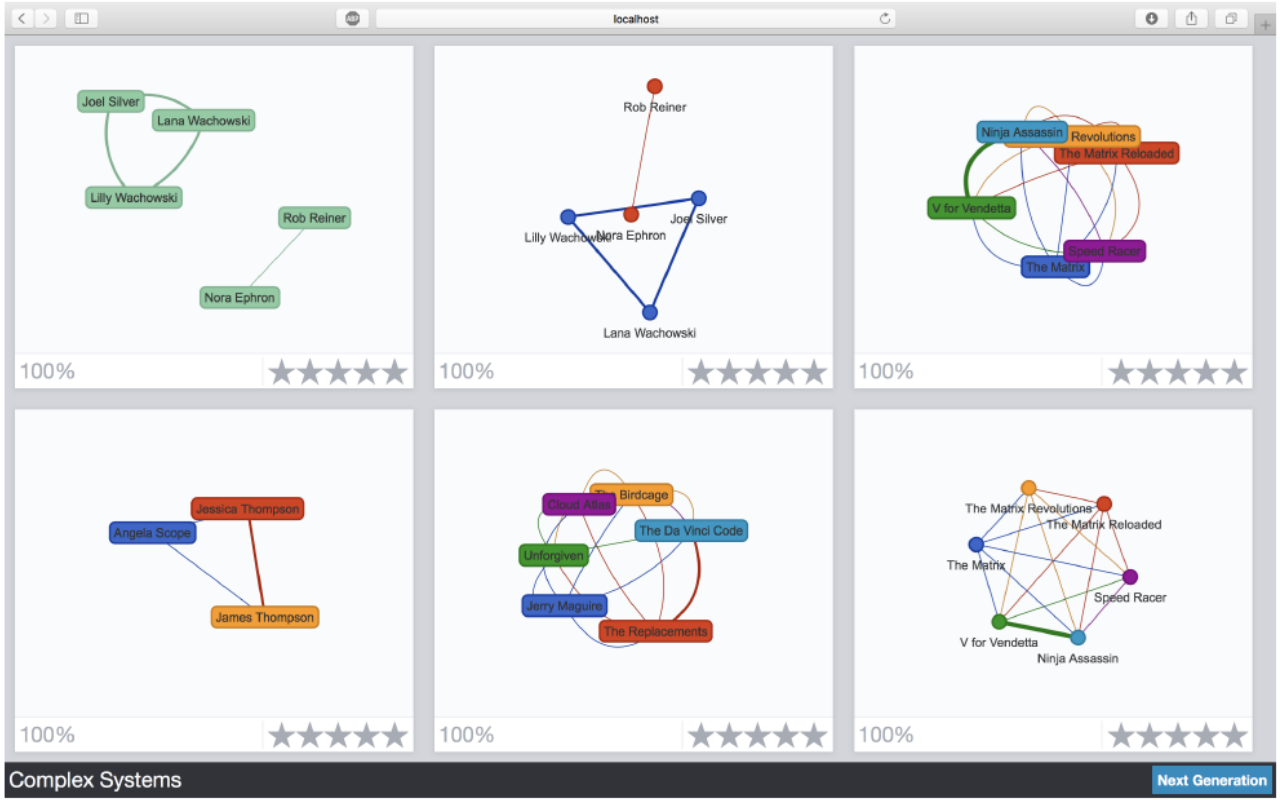
When do we stop?

- Eventually we expect to converge on a solution.
  - New offspring will not be significantly different from their parents.
  - So their offspring will be even less varied.

The ‘human’ fitness function.

- This approach can be used to assist a user to find a solution.

- The user becomes the fitness function and they select the candidates for the next generation.



Kemp T. 2017: Improving Complex System Visualisation with Genetic Algorithms (FYP)

How are they used in ML?

- You should know enough ML to recognise this as an ‘optimisation process.’
- That is stochastic search based.
- And the fitness function is the objective function.

Caution!!!

- If this approach is able to find good solutions why is it not popular?

- Because for almost any problem their is a specific better way to find the solution.

Example

- You want to divide a group of students into teams of roughly equal sizes, randomly.

- You want the teams to be of equal strength.

  - You can assign individual members a strength based on the results of previous tasks.

Chromosome

- Each student will be represented by a gene.
  - The value is the team to which they are assigned.

Fitness function

- Fitness function scores how close each team is to a target strength.

- And how close the teams are to the target size.

Initial Population

- Create a number of initial members of the population.
  - In each member assign each student to a random team.

Crossover and Mutation

- Generating new candidate team mappings will use crossover and mutation.
- Crossover will change the team assignment of several students.
- Mutation will change the team assignment of individual students.

We don’t use fitness to kill.

- It might be tempting to kill off offspring that have teams that are too big or too small.

- Don’t, they may only be a step away from the answer.

Your turn.

- In pairs:
  - Identify an optimisation task you are familiar with.
  - Design a suitable chromosome, fitness function and mutation to find an optimal solution.

Summary

- GAs model the process of natural selection.
- They are a stochastic search based optimisation process.
- The are good at complex searches, but can be inefficient.