Master of Technology in Knowledge Engineering

Unit 7:

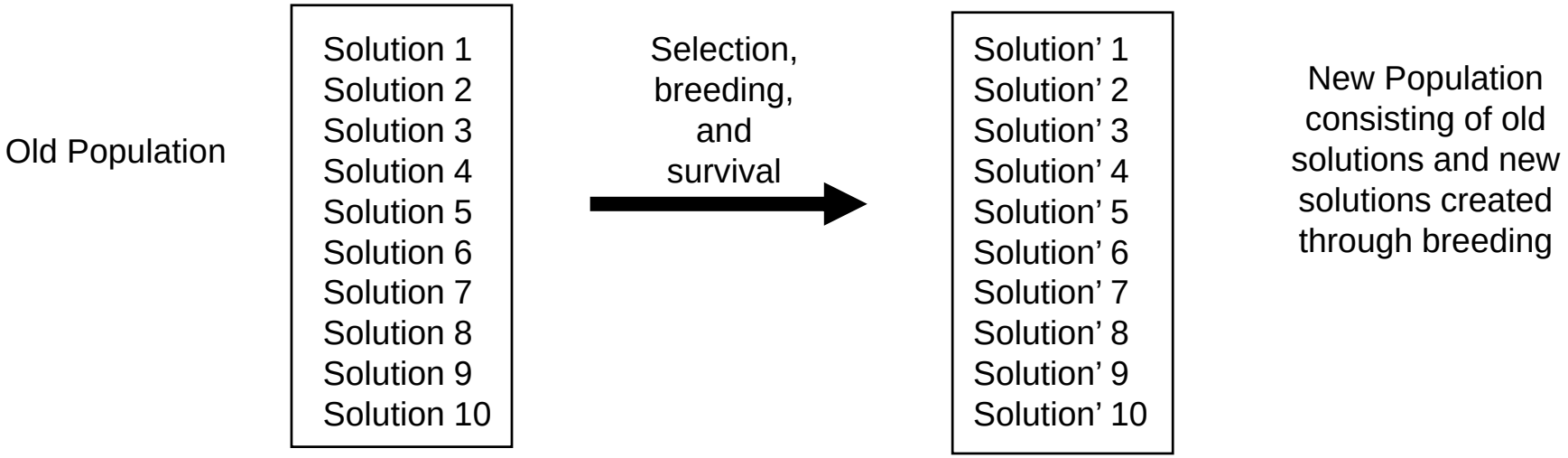
Developing Intelligent Systems for Performing Business Analytics

GA Modeling

Dr. Zhu Fangming
Institute of Systems Science,
National University of Singapore.
E-mail: isszfm@nus.edu.sg

GA – An Overview (1)

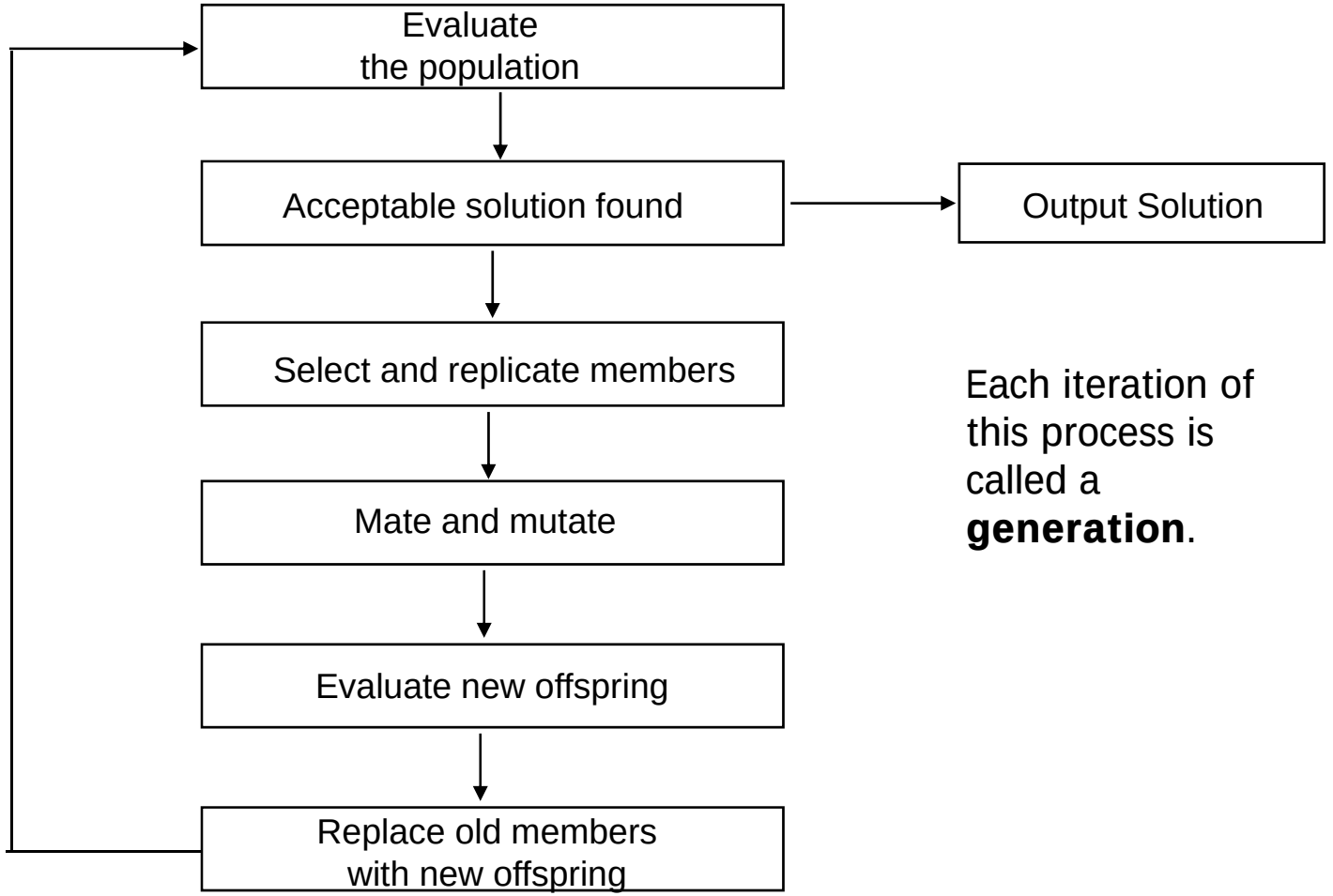
- General-purpose search algorithm that uses principles inspired by natural population genetics to evolve solutions to problems
- Seeks to produce better (fitter) individuals (solutions) by combining the best of the existing ones (via selection & breeding) and eliminating weaker members (survival of fittest)



GA – An Overview (2)

- **Initial population**
  - A set of candidate solutions (organisms)
  - Can be initialized using whatever knowledge is available about the possible solutions or created randomly
  - Should represent a random sample of the valid search space
- **New population**
  - Candidate solutions in the current population are selected for replication (breeding) based on their relative fitness
  - Very fit candidates might be chosen several times for breeding, while poor-performing ones might not be chosen at all
  - After breeding, the new population will hopefully be fitter than the old one

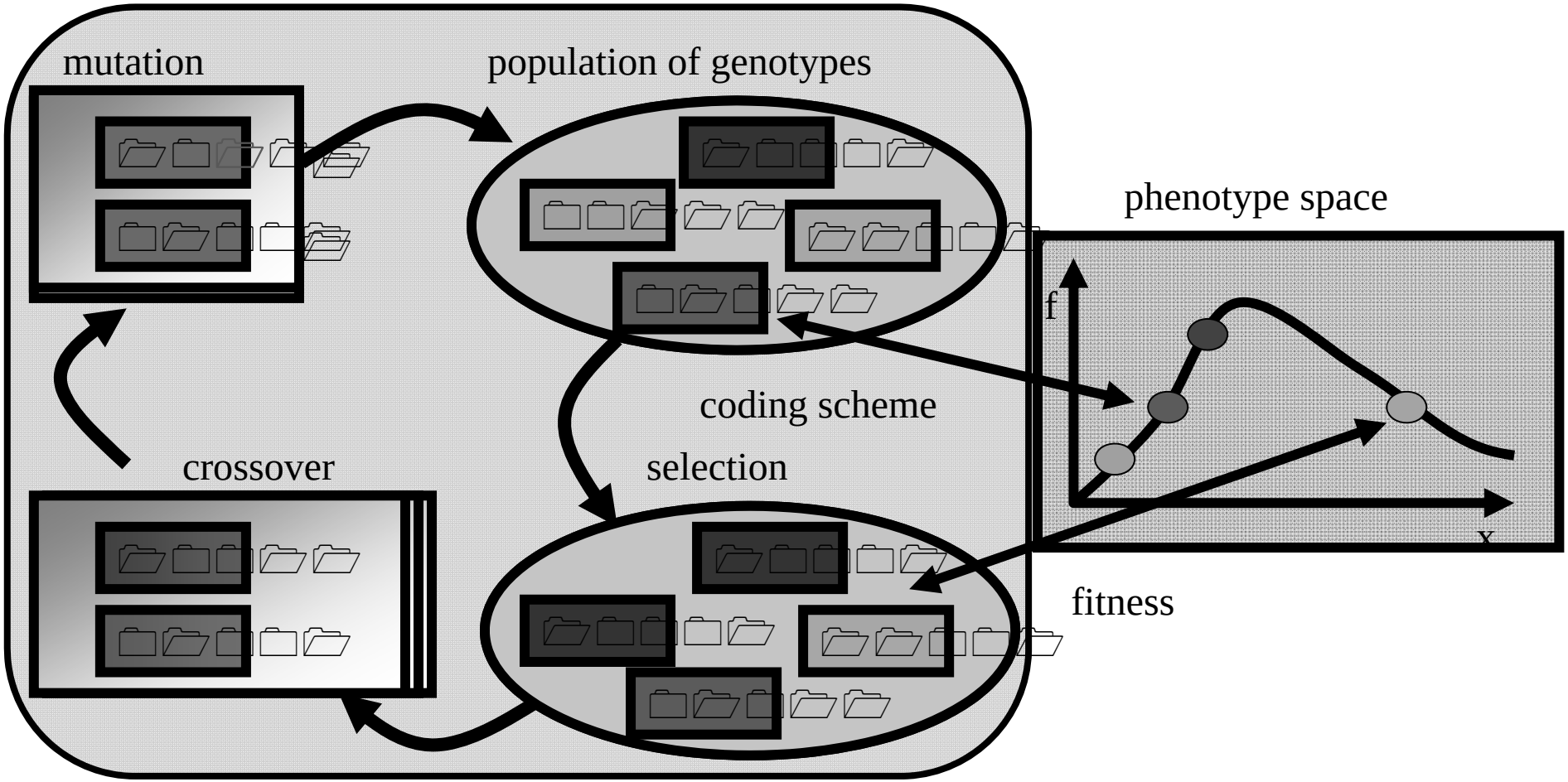
GA Search Process



Each iteration of this process is called a **generation**.

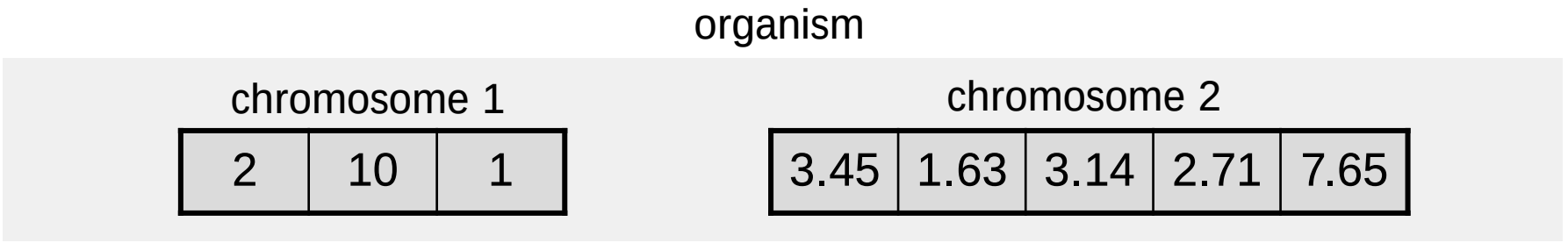
Evolution as Optimisation

GA search process



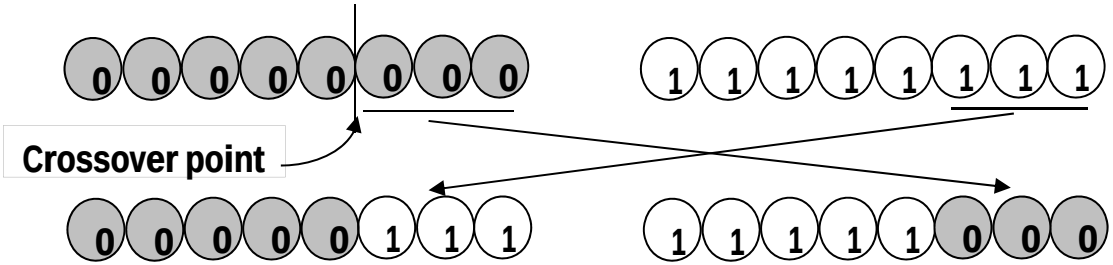
GA Terminology

- **Population** ~ a set of candidate solutions (organisms)
- **Organism** ~ represents a candidate solution to the optimisation problem. Consists of one or more chromosomes. Many problems typically use only one chromosome
- **Chromosome** ~ a collection of genes that are related by functionality
- **Gene** ~ a particular feature of the organism (e.g. black hair v.s. blonde hair) ~ a variable in the optimisation problem

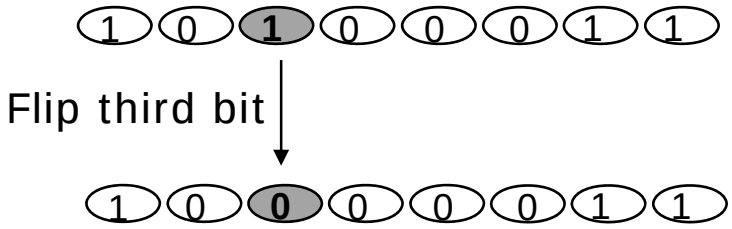


GA Mechanisms

- **Crossover**
  - Combine the genes from two parents to form two similar offspring
  - Tends to conserve the genetic information that was present in the parent chromosomes – i.e. tries to keep the "good" features of the parents
  -



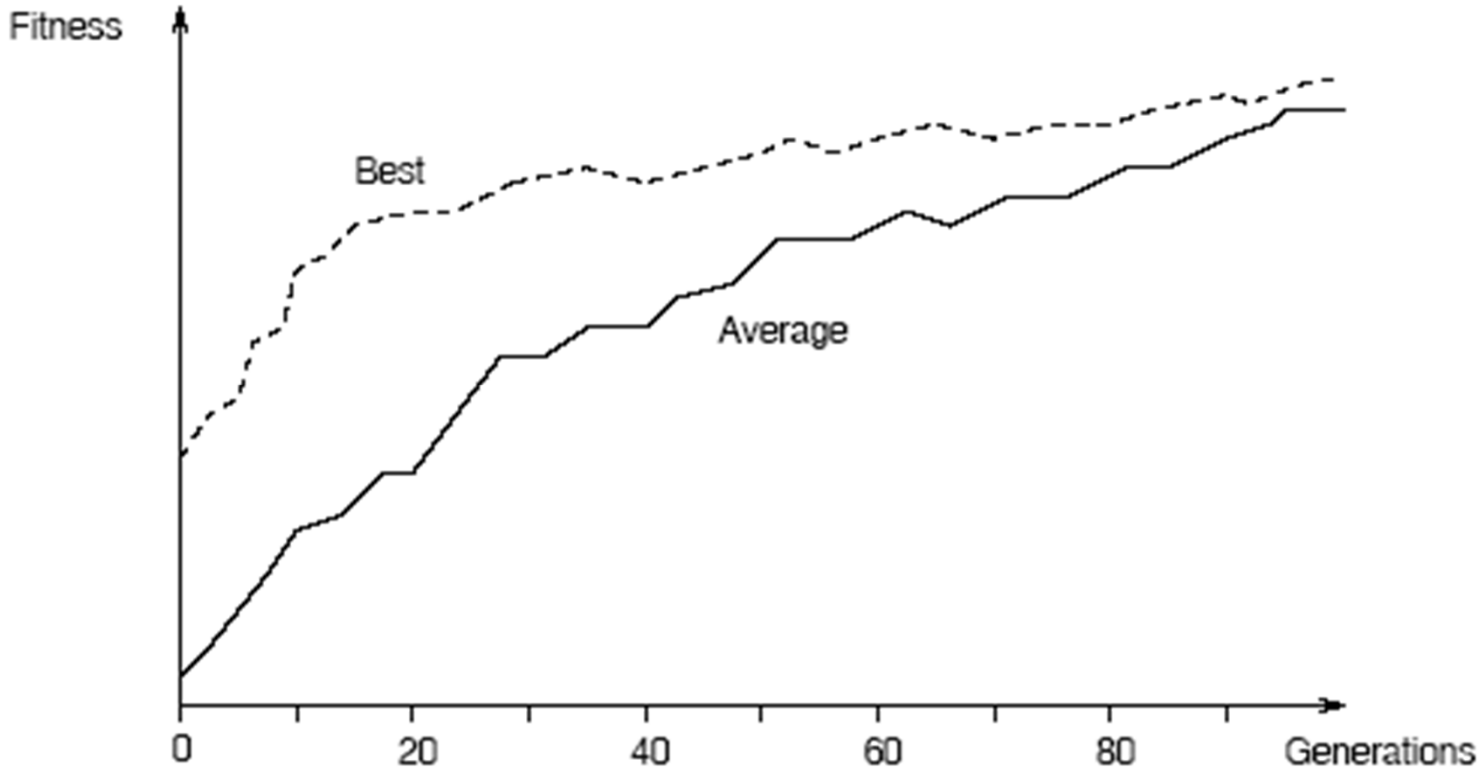
- **Mutation**
  - Randomly alters one or more genes (a direct analogy from nature, plays the role of regenerating lost genetic material).
  - Helps ensure population diversity



GA Mechanisms

- Selection and reproduction in GA are used in the given sequence
  1. **Select** fitter solutions (or individuals) to contribute genetic material to the next generation
     → selection method
  2. **Crossover** or recombine genetic materials from two parents to create new offsprings
     → crossover operator and crossover rate
  3. **Mutate** randomly selected components of new offsprings
     → mutation rate (and implicitly mutation operator)
  4. **Replace** some existing solutions in the current generation with the new offsprings to form the next generation
     → replacement strategy

A Sample GA Run



Taken from: Beasley, D. R. Bull, R. R. Martin. An Overview of Genetic Algorithms: Part 1, Fundamentals. University Computing, 1993.

When & How to Use GAs?

- When would you want to apply GA on a problem?
  - No idea how to reasonably solve the problem
  - You cannot enumerate all possible solutions
  - You know how to evaluate how good or bad a solution is
- What do we need to use Genetic Algorithms?
  - A **goodness measure**
    $\rightarrow$ this is eventually translated into the fitness function
  - A **representation** of the solution
    $\rightarrow$ the more natural, the better
  - A **problem model** which defines the behaviour

Applications of GA

- GAs can be applied to a wide range of optimization and learning problems
  - Routing and scheduling, machine vision, engineering design optimization, gas pipeline control systems, machine learning
  - Hundreds of applications have now been discovered and a variety of commercial software tools have been introduced.

- Numerical function optimisation
  - GA has been shown to outperform conventional optimisation techniques on difficult, discontinuous, multimodal, noisy functions.

Applications of GA

- Combinatorial optimisation
- Financial forecasting
- Design optimisation
- Pattern recognition
- Machine learning
- etc...

GA Tools

- Matthew's GAlib, A C++ Library of Genetic Algorithm Components developed at MIT Technology Center
  - The library includes tools for using genetic algorithms to do optimization in any C++ program using any representation and genetic operators.
  - http://lancet.mit.edu/ga/
- JGAP, Java Genetic Algorithms Package
  - a Genetic Algorithms and Genetic Programming component provided as a Java framework.
  - Easy to use, highly modular
  - Can plug in custom genetic operators
  - .Net version available
  - http://jgap.sourceforge.net/

GA Tools

- Excel-based GA tools
  - Excel Solver
    - http://www.solver.com/
  - Evolver
    - http://www.palisade.com/evolver/
  - SolveXL
    - http://www.solvexl.com/
  - GeneHunter
    - http://www.wardsystems.com/genehunter.asp

How to Load the Excel Solver Add-in

- **Open Excel**
- Click the **File** tab, and then click **Options**.
- Click **Add-Ins**, and then in the **Manage** box, select **Excel Add-ins**.
- Click **Go**.
- In the **Add-Ins available** box, select the **Solver Add-in** check box, and then click **OK**.
- If **Solver Add-in** is not listed in the **Add-Ins available** box, click **Browse** to locate the add-in.
- If you get prompted that the Solver add-in is not currently installed on your computer, click **Yes** to install it.
- After you load the Solver add-in, the **Solver** command is available in the **Analysis** group on the **Data** tab.

Excel Solver

- Excel Solver Help
  - http://www.solver.com/excel-solver-help

#Click HELP to get the help file.

GA Modelling – Excel Solver

Step 1:

- Decide what you are optimising for (e.g. maximise profit). This 'goodness measure' or objective will eventually translate into a fitness function

Step 2:

- Represent each solution chromosome as an array of real numbers or integers
- E.g. to find the best split of funds that will allow a company to maximise profits

advertising	marketing	production	salary		profit
8%	12%	50%	30%		$2780

One solution **Fitness function**

GA Modelling – Excel Solver

Step 3:

- Write a **fitness function** that evaluates the goodness of a solution chromosome.
- This function takes as input a candidate solution and returns a number that indicates how good the solution is (e.g. the amount of expected profit)

Step 4:

- Define any constraints on the values of your solution chromosome.

Exercise 1

Your company is launching a new product soon. You are to consider the different media for its best publicity. The four types of media for advertising are TV, Newspaper, Radio and WWW. The maximum budget for each type of media is 20, 100, 50, 10 respectively. The total amount that you are allowed to spend on the publicity is 100.

Assuming the equation for measuring the publicity result is 0.4*Radio+0.8*Newspaper/(0.3*TV+0.5*WWW), how much will you spend on each type of media? Set the Excel spreadsheet as shown.

	A	B
1	TV	20
2	Newspaper	100
3	Radio	50
4	WWW	10
5	Result	=0.4*B3+0.8*B2/(0.3*B1+0.5*B4)

Exercise 2

The following jobs can be processed on any of the 5 machines. How can these jobs be assigned to the machine so that the total processing time for the jobs is minimum. The time taken to process each job on each machine is known.

Set the Excel spreadsheet as shown. In the Formulas menu, select Define Name. Type MCData into the Name field and B10:F15 in Refer to field.

	A	B	C	D	E	F
1		M/C ID	Process time			
2	Job1	1	=HLOOKUP(B2,MCData,2)			
3	Job2	2	=HLOOKUP(B3,MCData,3)			
4	Job3	3	=HLOOKUP(B4,MCData,4)			
5	Job4	4	=HLOOKUP(B5,MCData,5)			
6	Job5	5	=HLOOKUP(B6,MCData,6)			
7		Total	=SUM(C2:C6)			
8						
9						
10	M/C ID	1	2	3	4	5
11	Job1	12	45	23	33	12
12	Job2	34	13	8	14	25
13	Job3	22	13	33	15	24
14	Job4	14	56	23	12	26
15	Job5	4	13	23	34	27

MCData

Exercise 3

You have a group of students whose average overall performance are known. You are required to divide them into 3 groups so that the members of each group can interact well with each other. To ensure that they can interact well with each other the deviations of their performance should be minimum.

Set the Excel spreadsheet as shown

	A	B	C	D	E	F
1	Student	Performance	Group	Group1	Group2	Group3
2	Tom	50	1	IF(C2=1, B2,"0")	IF(C2=2, B2, "0")	IF(C2=3, B2, "0")
3	Jerry	67	2	IF(C3=1, B3, "0")	IF(C3=2, B3, "0")	IF(C3=3, B3, "0")
4	Ann	34	3	IF(C4=1, B4, "0")	IF(C4=2, B4, "0")	IF(C4=3, B4, "0")
5	Bob	55	1	IF(C5=1, B5, "0")	IF(C5=2, B5, "0")	IF(C5=3, B5, "0")
6	June	80	2	IF(C6=1, B6, "0")	IF(C6=2, B6, "0")	IF(C6=3, B6, "0")
7	Nancy	90	3	IF(C7=1, B7, "0")	IF(C7=2, B7, "0")	IF(C7=3, B7, "0")
8	Mike	77	1	IF(C8=1, B8, "0")	IF(C8=2, B8, "0")	IF(C8=3, B8, "0")
9	Dolly	55	2	IF(C9=1, B9, "0")	IF(C9=2, B9, "0")	IF(C9=3, B9, "0")
10	Paul	66	3	IF(C10=1, B10, "0")	IF(C10=2, B10, "0")	IF(C10=3, B10, "0")
11	Lucy	83	1	IF(C11=1, B11, "0")	IF(C11=2, B11, "0")	IF(C11=3, B11, "0")
12	John	44	2	IF(C12=1, B12, "0")	IF(C12=2, B12, "0")	IF(C12=3, B12, "0")
13	Mary	73	3	IF(C13=1, B13, "0")	IF(C13=2, B13, "0")	IF(C13=3, B13, "0")
14	Number			COUNTIF (D2:D13,">0")	COUNTIF (E2:E13,">0")	COUNTIF (F2:F13,">0")
15	Average			SUM(D2: D13)/D14	SUM(E2: E13)/D14	SUM(F2: F13)/D14
16	Std Dev			STDEV(D 2:D13)	STDEV(E 2:E13)	STDEV(F 2:F13)
17						
18				SUM(D16 :F16)		