Inspire...Educate...Transform.

Inspire…Educate…Transform.

Simulations, Evolutionary searches

Dr. K. V. Dakshinamurthy

President, International School of Engineering

Why do we need samples

SAMPLING STRATEGY

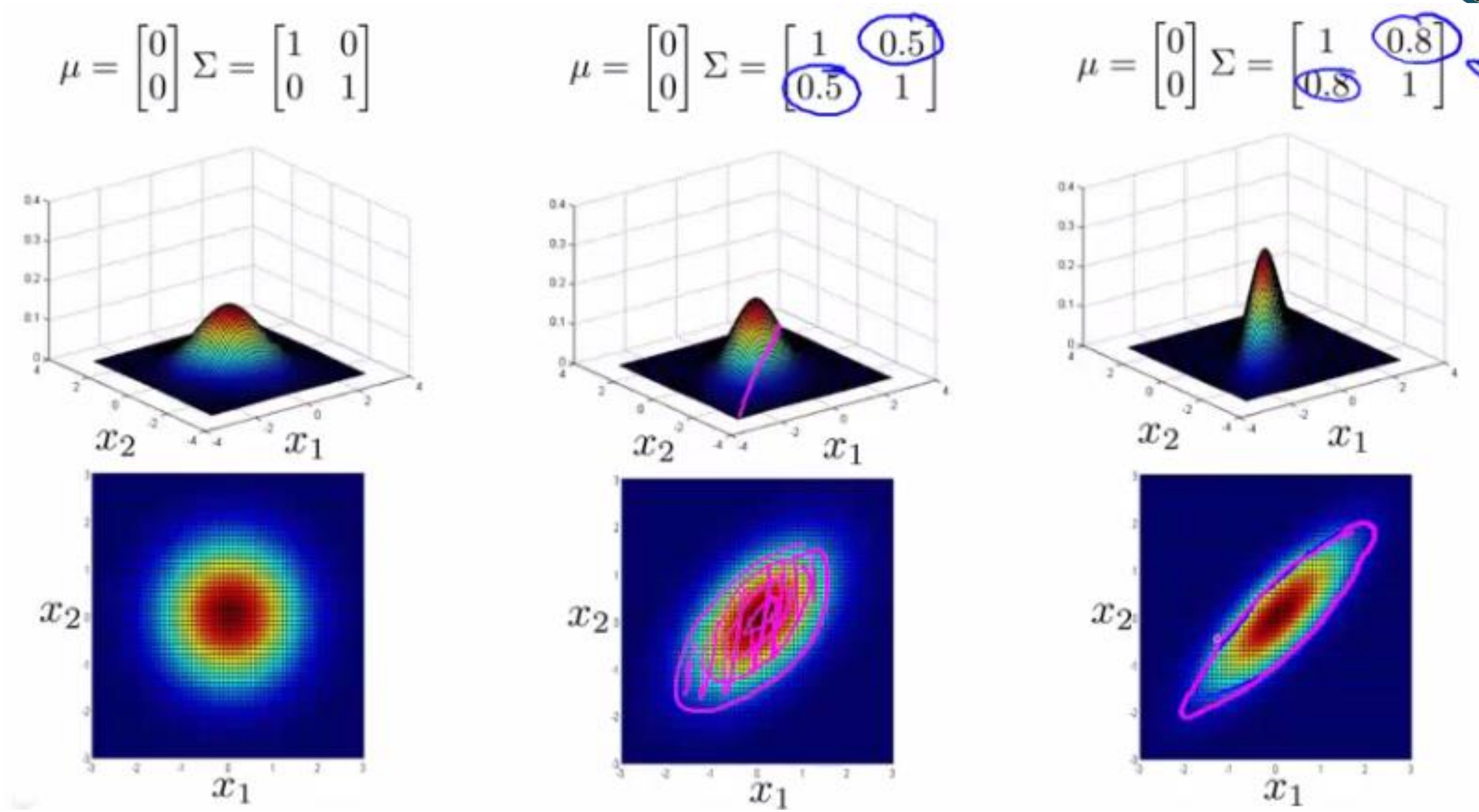
Learn to sample

- Given a distribution, let us learn to sample from it
- First, let us understand why and then learn the strategy
- Then check how good this process is

Measuring central tendencies

- Computing mean median and mode is a lot of calculus for a continuous distribution
  - $\sum$ *must be replaced by* $\int$ *sign*.
  - $Mean = \int xp(x)dx$
  - $Median = \int_0^M p(x)dx = \int_M^k p(x)dx = \frac{1}{2}$
  - $Mode\ is\ where\ \frac{d(p(x)}{dx} = 0$

Even normal distribution becomes complex



Multivariate Gaussian

$$Bell\ curve\ in\ multivariate\ case =$$

$$2\pi^{\frac{-N}{2}}|\Sigma|^{\frac{-1}{2}}e^{(\frac{-1}{2}(X-\mu)^T\Sigma^{-1}(X-\mu))}$$

$$\mu = \frac{1}{m}\sum_{i=1}^{m} x^{(i)}$$

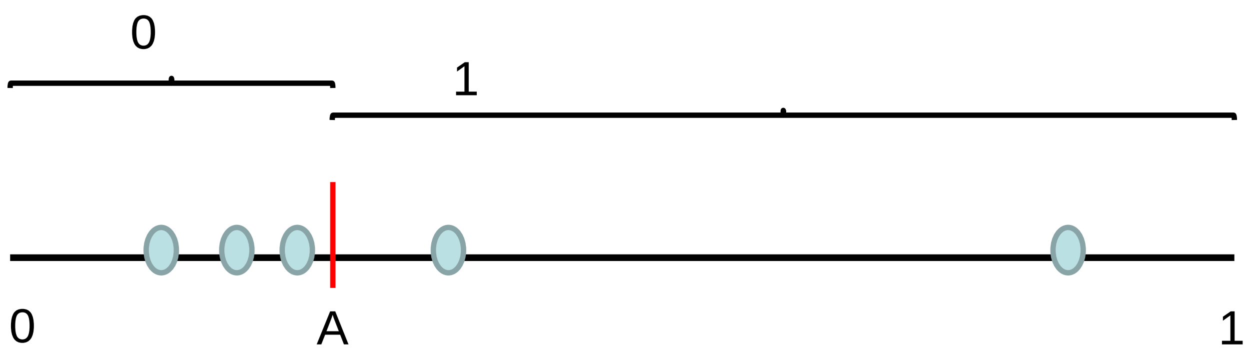
$$\Sigma = \frac{1}{m}\sum_{i=1}^{m}(x^{(i)} - \mu)(x^{(i)} - \mu)^T$$

Estimating from samples

- Let us have IID (identically, independently distributed samples) of a distribution
- Then, mean is

$$\frac{1}{M}\sum_{m=1}^{M} x[m]$$

Creating samples from a discrete distribution



P(0)= A and P(1)=1-A

Randomly generate a number between 0 and 1; If it is less than A, sample is 0 else it is 1

Sample =(0,1,0,0,1)

Discretizing a continuous variable

- Equal bin, equal frequency
- With class variable, sort and decide where to cut (cut wherever the class variable changes its values)
- Many such algorithms exist

Age	25	32	34	35	35	37	37	38
Loan	0	1	1	0	0	0	1	1

Sampling based estimations

$$T_{\mathcal{D}} = \frac{1}{M} \sum_{m=1}^{M} X[m]$$

Hoeffding Bound:

$$P_{\mathcal{D}}(T_{\mathcal{D}} \notin \underbrace{[p-\epsilon, p+\epsilon]}) \leq 2e^{-2M\epsilon^2}$$

The probability of getting the sample estimate outside the range of ε is given by right hand side. So, this is the probability of a bad estimate.

It goes down exponentially with sample size and tolerance.

For additive bound ε on error with probability > 1-δ:

$$M \geq \frac{\ln(2/\delta)}{2\epsilon^2}$$

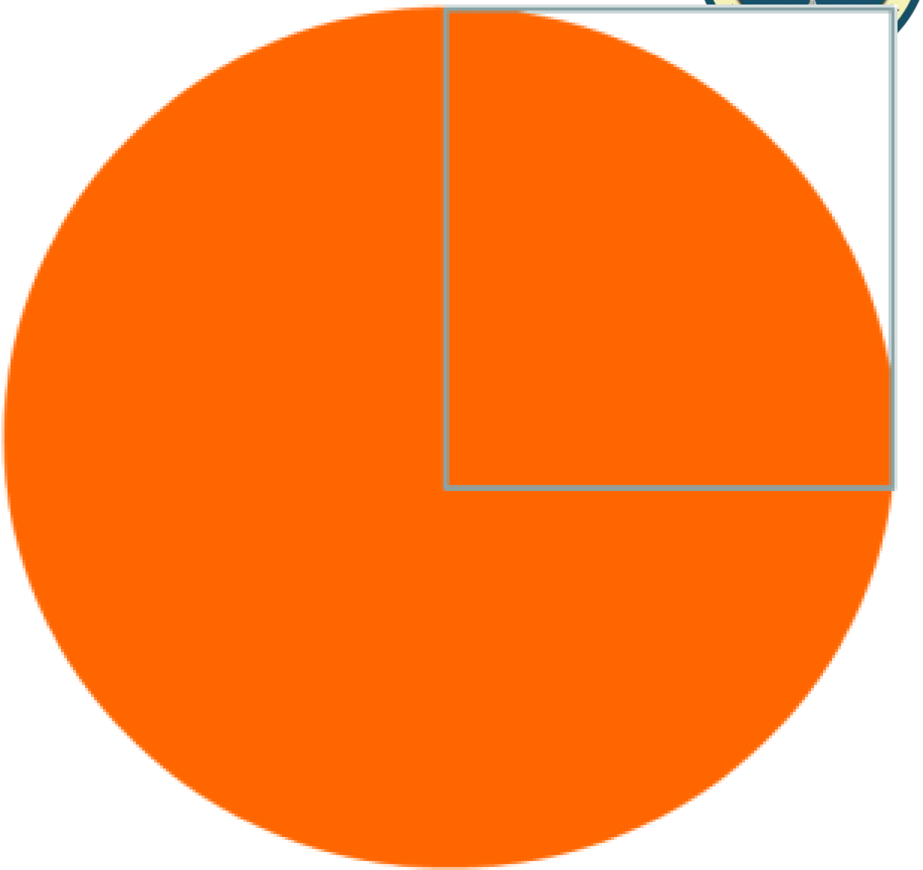
It turns out that we do need a large number of samples in most cases.

Computing central tendencies

APPLICATIONS OF SAMPLING:

Monte Carlo Simulations

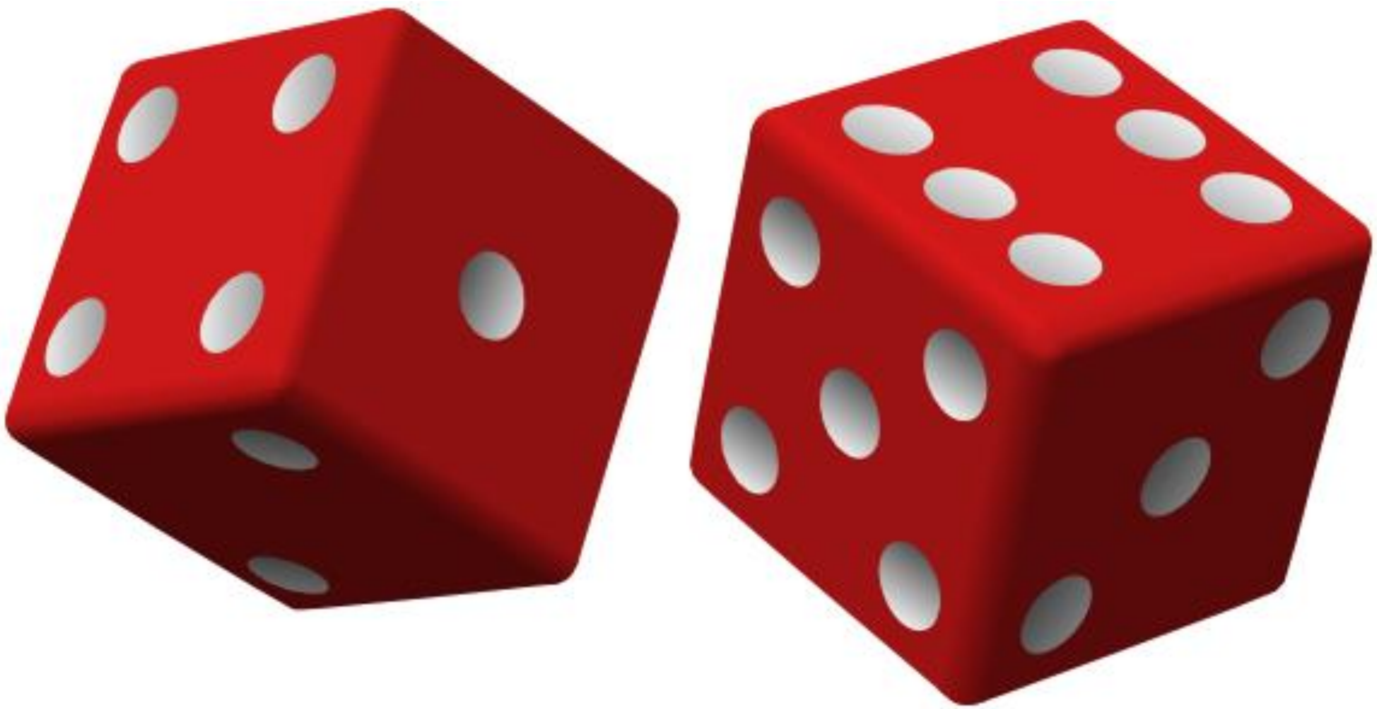
Value of Pi



When to stop?

- Define a small tolerance δ delta
- If the difference between nth iteration and n+1th iteration is less than delta, stop. Else, continue

Probability



Capital markets

Outcome	Green	Red	White
1	0.8	0.06	0.9
2	0.9	0.2	1
3	1.05	1	1
4	1.1	3	1
5	1.2	3	1
6	1.4	3	1.1

GENERATING AND UNDERSTANDING BEHAVIOR OF TEXT, IMAGES, AUDIO

Transition probabilities

- Process
  - Learn the distribution from a large corpus
  - Collate counted probabilities as a matrix (columns, current state and rows future state)
  - Text, images, music etc.

Learning distributions

	a	b	c	d	e	f	g	h	I
A	0.08	0.09							
B									

Larger corpuses yield better descriptions

Character generation

- Random
  - No use
- First order (character-Character transitions)
  - t I amy, vin. id wht omanly heay atuss n macon aresethe hired boutwhe t, tl, ad torurest t plur I wit hengamind tarer-plarody thishand.
- Second order (Previous 2 characters, di-gram)
  - Ther I the heingoind of-pleat, blur it dwere wing waske hat trooss. Yout lar on wassing, an sit." "Yould," "I that vide was nots ther.

Applications of Markov chains

- 3rd order
  - I has them the saw the secorrow. And wintails on my my ent, thinks, fore voyager lanated the been elsed helder was of him a very free bottlemarkable,

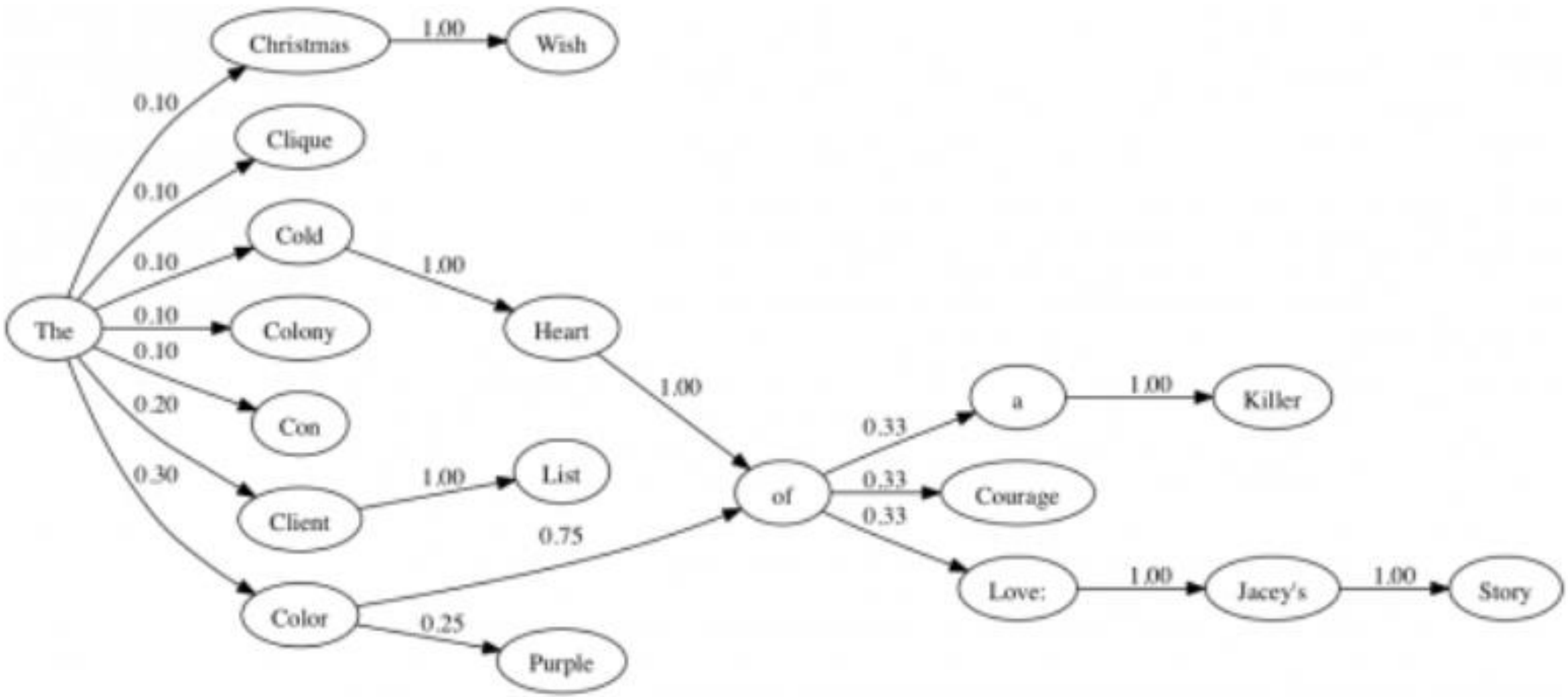
- 4 gram
  - His heard." "Exactly he very glad trouble, and by Hopkins! That it on of the who difficentralia. He rushed likely?" "Blood night that.

Text generation

- First order <u>word</u> markov chains
  - Construct a matrix with transition probabilities of all word pairs and the probability of succeeding
  - Start random and select the words obeying the distribution

Let us generate new movie titles

- http://www.soliantconsulting.com/blog/2013/02/title-generator-using-markov-chains
  - The Christmas Wish, The Client, The Client List, The Clique, The Cold Heart of a Killer, The Colony, The Color of Courage, The Color of Love: Jacey's Story, The Color Purple, The Con
  - Constructing a list of unique words that appear in these titles: "The", "Christmas", "Wish", "Client", "List", "Clique", "Cold", "Heart", "of", "a", "Killer", "Colony", "Color", "Courage", "Love:", "Jacey's", "Story", "Purple", "Con".



Probability diagram

Process

- We can start with the word “The” and for each new word desired, we’ll choose from the distribution which next node to move to accordingly.
- Some possible generated strings include new titles “The Color of a Killer” or “The Cold Heart of Love: Jacey’s Story”.
- With more data, a more complicated graph can be constructed, and more interesting and varied strings can be generated as a result. You should use as much data in your corpus as you can find!
- Small tweaks like setting a minimum length, choosing to start with start words and end with end words might help

Markov Chain in Text

- http://en.wikipedia.org/wiki/Mark_V_Shaney

- Finding who the author of a work is
  - Compare the transitional probabilities of unknown work and known works

Music: Monte Carlo

rand_music.wav

First order transitions

1st-order matrix

Note	A	C♯	E♭
A	0.1	0.6	0.3
C♯	0.25	0.05	0.7
E♭	0.7	0.3	0

Markov second order

2nd-order matrix

Note	A	D	G
AA	0.18	0.6	0.22
AD	0.5	0.5	0
AG	0.15	0.75	0.1
DD	0	0	1
DA	0.25	0	0.75
DG	0.9	0.1	0
GG	0.4	0.4	0.2
GA	0.5	0.25	0.25
GD	1	0	0

Music_Markov Chain

first_score.wav

https://ssodelta.wordpress.com/2014/03/20/generating-mozart-from-markov-chains/

https://medium.com/@omgimanerd/generating-music-using-markov-chains-40c3f3f46405#.2ydm6na5

http://www.cs.uml.edu/ecg/uploads/AIfall11/SimoneHill.FinalPaper.MarkovMelodyGenerator.pdf

NLP

- HMM models
  - Learn transition probabilities of parts of speech using 2nd order
  - Gene sequences

AN INTERESTING PROPERTY OF TRANSITION PROBABILITY MATRICES

Understanding Markov Chain

- Let us say, my class has the following properties
  - 90% of the interested students will remain interested; 10% doze off
  - 50% of the dozing students get interested; remaining stay as it is

What is the steady state?

- Let us start with 100 interested and 0 boring
- Work out what happens after 5 periods

Using matrices to solve the problem

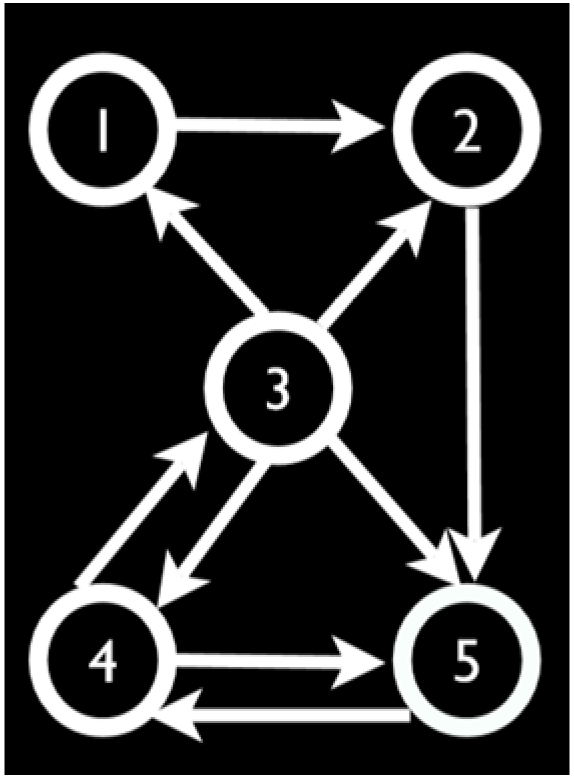
- Transition matrix
  - Each columns: one state transition
- Transition matrix * current distribution = Next distribution

This is a Markov chain

- Finite states
- Fixed transition probabilities
- From one state, I should be able to go to any state
- Column stochastic matrices

Other applications

- Google Page rank



$$\begin{bmatrix} 0 & 0 & 1/4 & 0 & 0 \\ 1 & 0 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 1/4 & 0 & 1 \\ 0 & 1 & 1/4 & 1/2 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1/20 \\ 5/20 \\ 1/10 \\ 5/20 \\ 7/20 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1/4 & 0 & 0 \\ 1 & 0 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 1/4 & 0 & 1 \\ 0 & 1 & 1/4 & 1/2 & 0 \end{bmatrix} \begin{bmatrix} 1/5 \\ 1/5 \\ 1/5 \\ 1/5 \\ 1/5 \end{bmatrix}$$

$$r^{(t+1)} = H \quad r^{(t)}$$

	Iteration 0	Iteration 1	Iteration 2	Page Rank
P_1	1/5	1/20	1/40	5
P_2	1/5	5/20	3/40	4
P_3	1/5	1/10	5/40	3
P_4	1/5	5/20	15/40	2
P_5	1/5	7/20	16/40	1

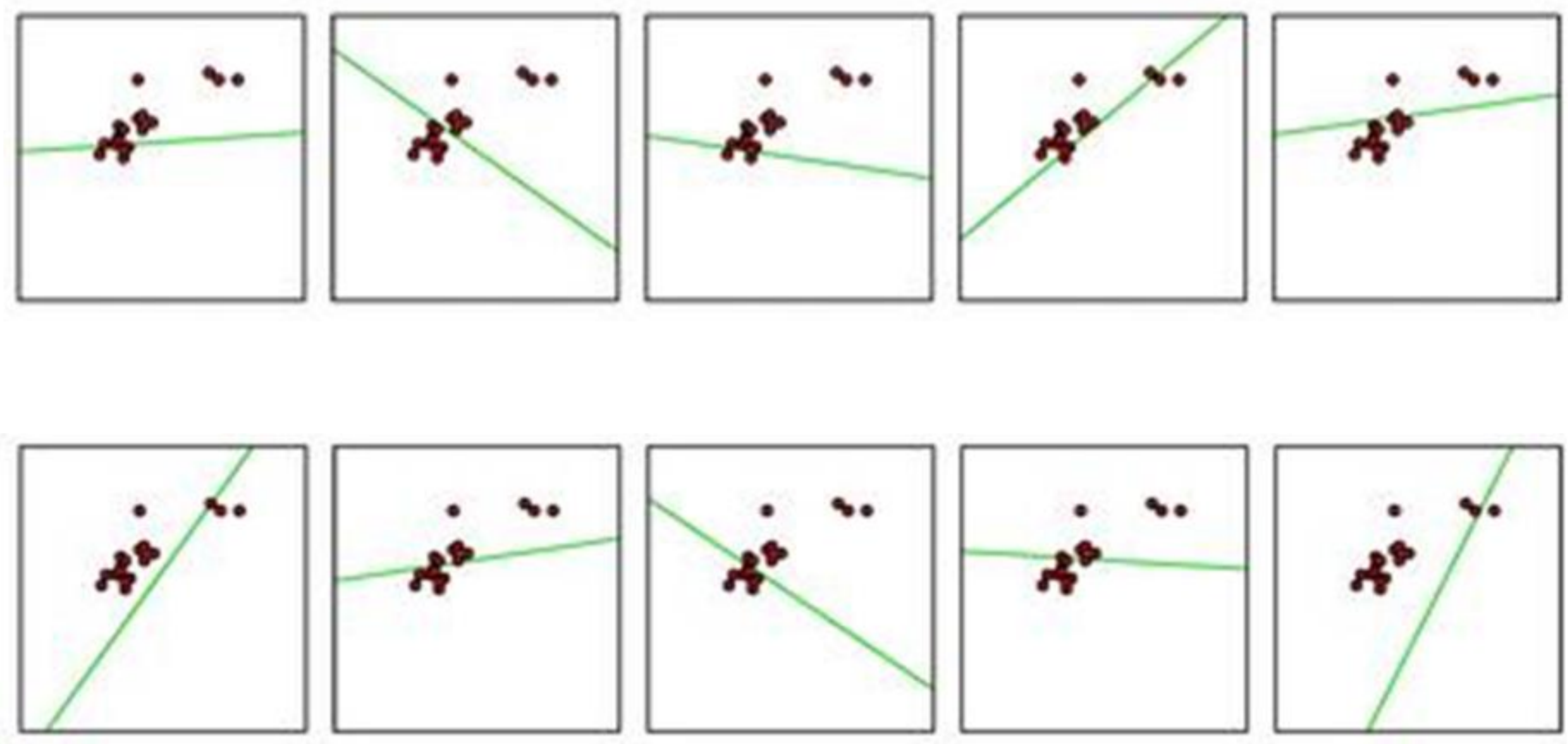
Guided random searches

CAN SAMPLING BE USED TO SEARCH FOR MAXIMA AND MINIMA?

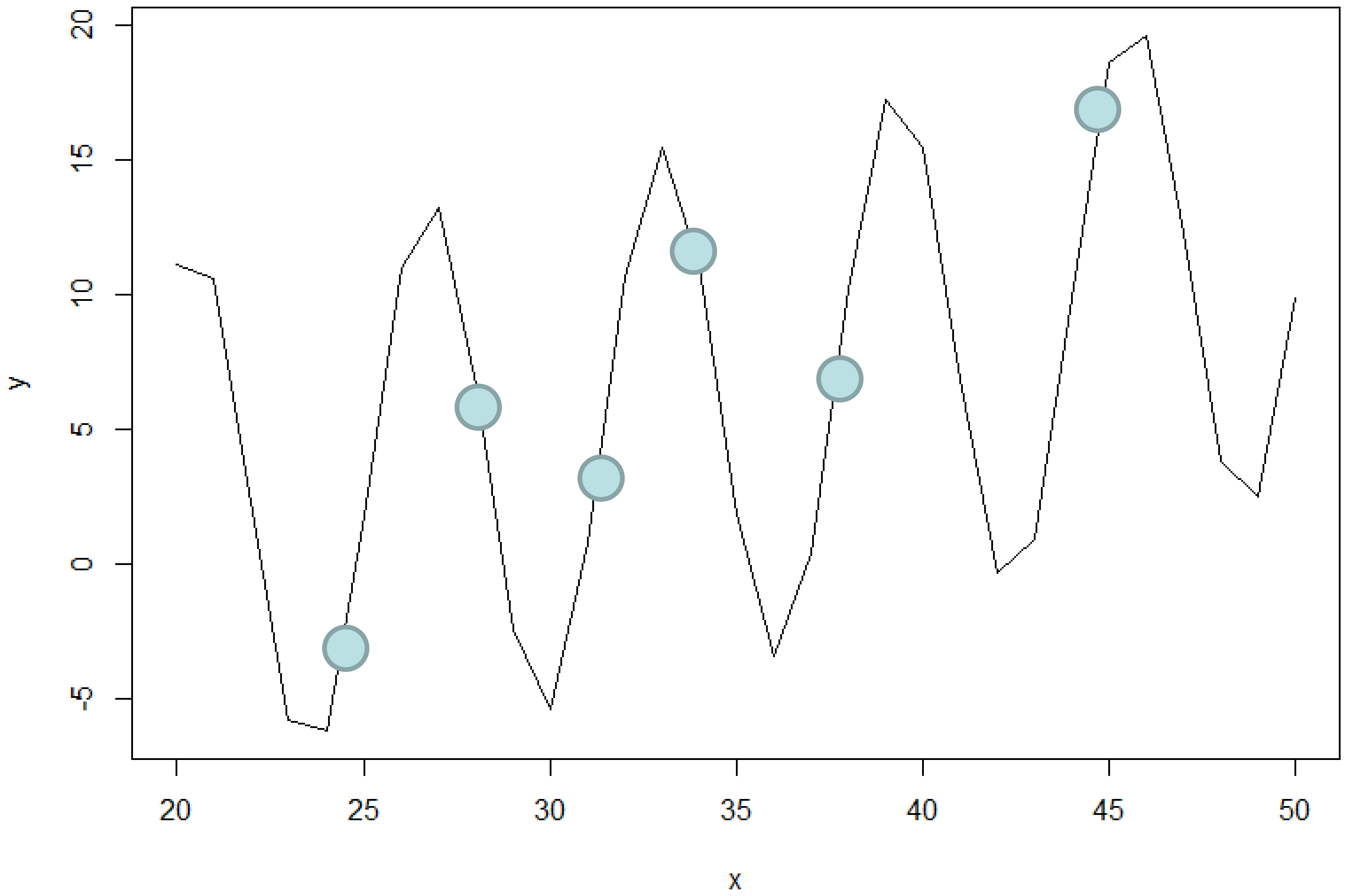
What Sampling and MonteCarlo are not

- They are random searches
- But, not optimization methods
- Can you do linear regression with Montecarlo?

Linear regression using Monte Carlo



A complex function and we need to find minima



A possible approach

- Do a million simulations and pick the most optimal solution
- Not guaranteed that we get anywhere close to the solution

- What to do?
  - Have memory

Genetic algorithms

GUIDED RANDOM SEARCHES

What does survival of the fittest mean

Summary of biology

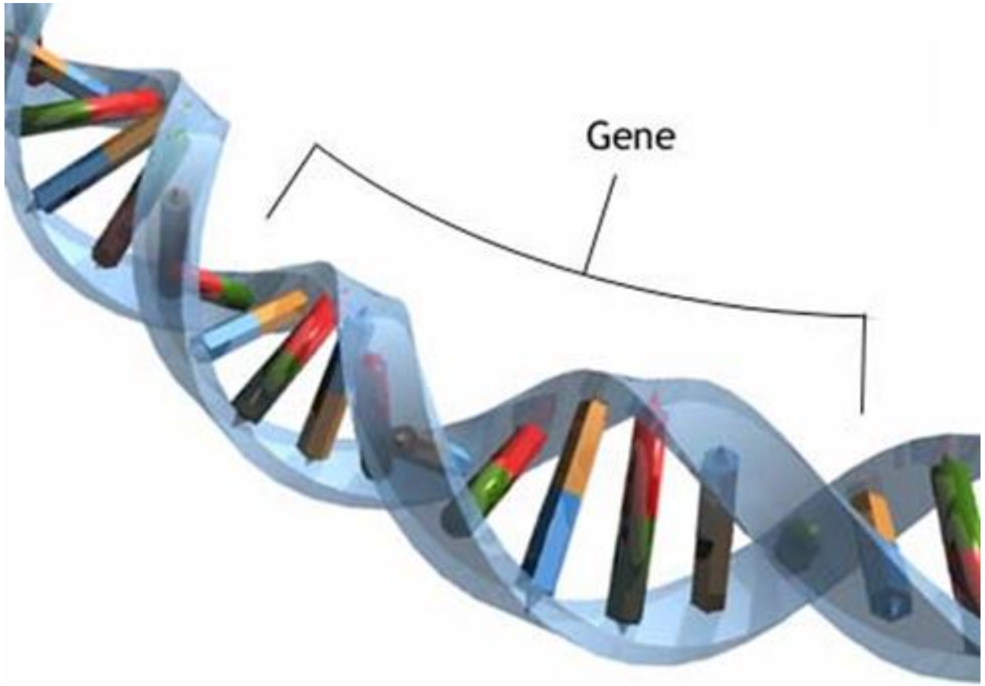
- Not all members survive until reproduction. Only a few learn to earn lunch. Others are lunch!
- The strong ones then need to attract and mate other members of the species that survived.
- The offspring generation hence is normally better than the parents generation
- Over generations, species develop

Can a similar theory be applied to optimization

- Start with random solutions
- Kill the weak ones (memory!)
- Reproduce from the good ones
- Will the newer set be more optimal than its parents?

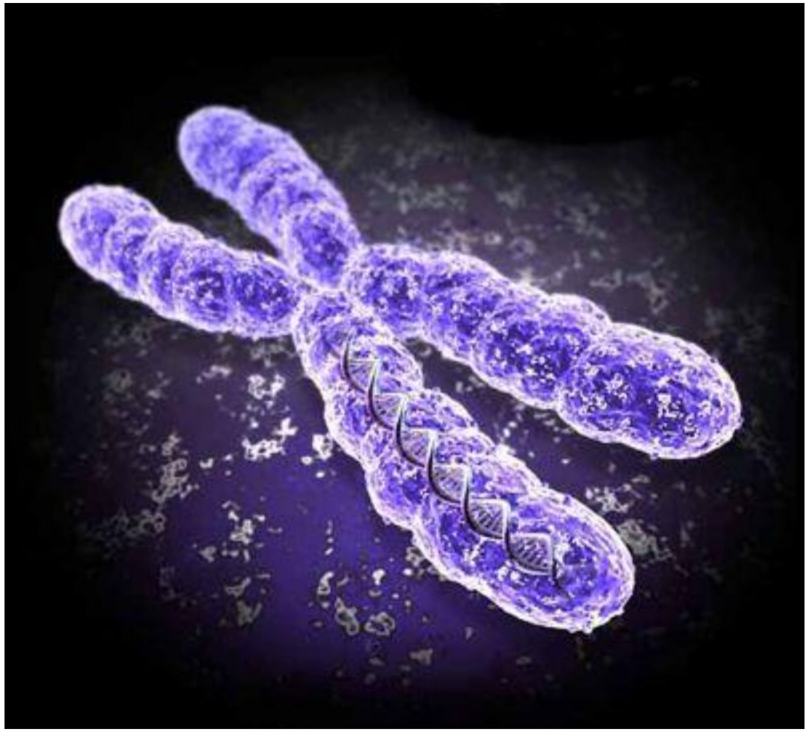
Gene

Gene: The reproducible building block of chromosome



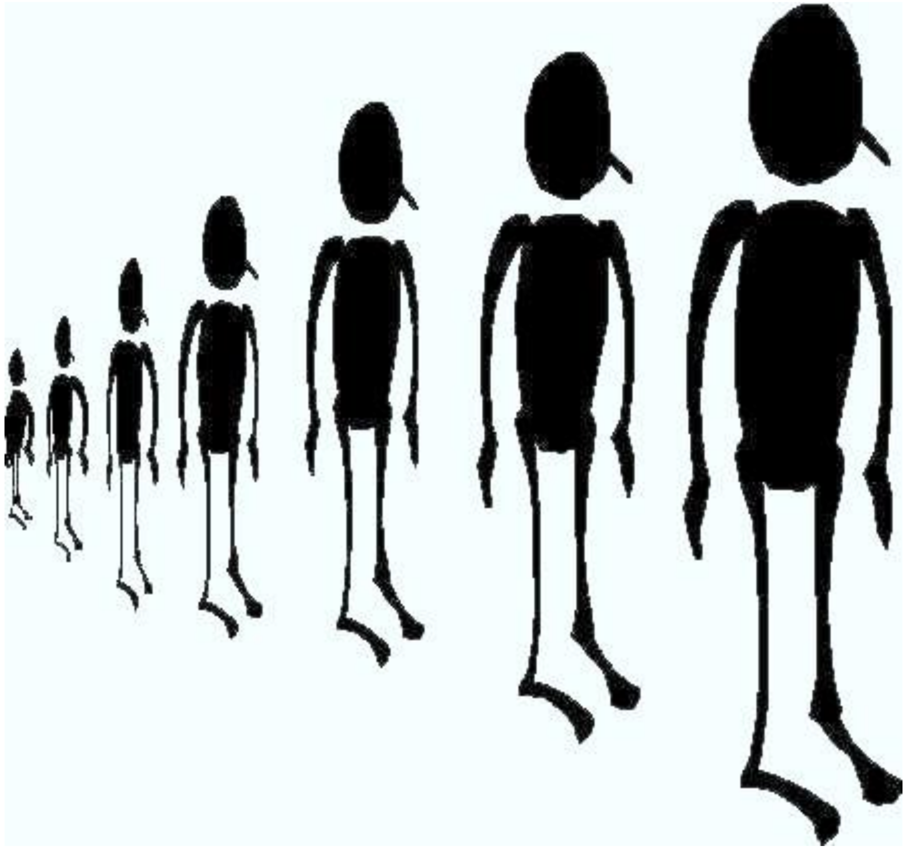
Beg, Borrow or Steel

- Chromosome: A possible solution



Population

- A set of chromosomes



Generation

- A population derived from the fittest chromosomes of the previous population

Reproduction, Mating and Mutation

- Create a new solution from the old fitter solutions

```
{
    initialize population;
    evaluate population;
    while TerminationCriteriaNotSatisfied
    {
        select parents for reproduction;
        perform recombination and mutation;
        evaluate population;
    }
}
```

So, GA at a snapshot

- Create a schema to create possible solutions in a bit/byte format
- Start with a number of random solutions
- Identify the best ones
- Create new ones from them using some exchange mechanism
- Continue with the last two steps until a solution is found

Some unique differentiators

- GAs work with a coding of the parameters and not the parameters themselves
- GAs search from a population of points and not a single point
- GAs work with the objective function and not the derivatives
- GAs transition based on random rules and not on deterministic rules

OPERATIONAL ASPECTS

Innovations: Creating a coding scheme

- Knap Sack
- Polynomial fit
- TSP
- Portfolio allocation

- Define objective function and solution

Population

Chromosomes could be:

– Bit strings	(0101 ... 1100)
– Real numbers	(43.2 -33.1 ... 0.0 89.2)
– Permutations of element	(E11 E3 E7 ... E1 E15)
– Lists of rules	(R1 R2 R3 ... R22 R23)
– Program elements	(genetic programming)
– ... any data structure ...	

How do we decide mating pool

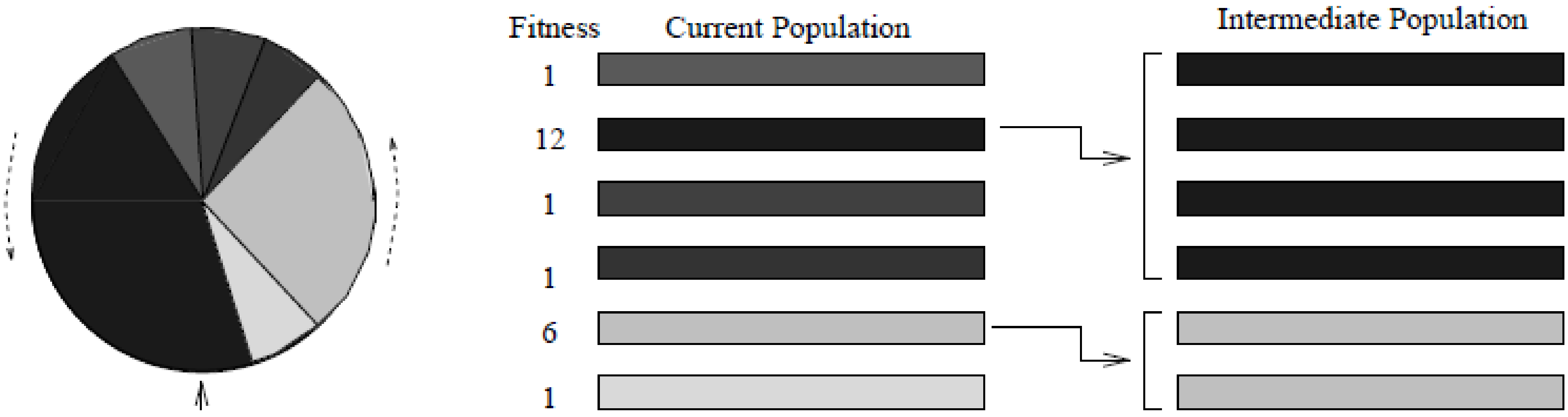


Figure 2. Selection application

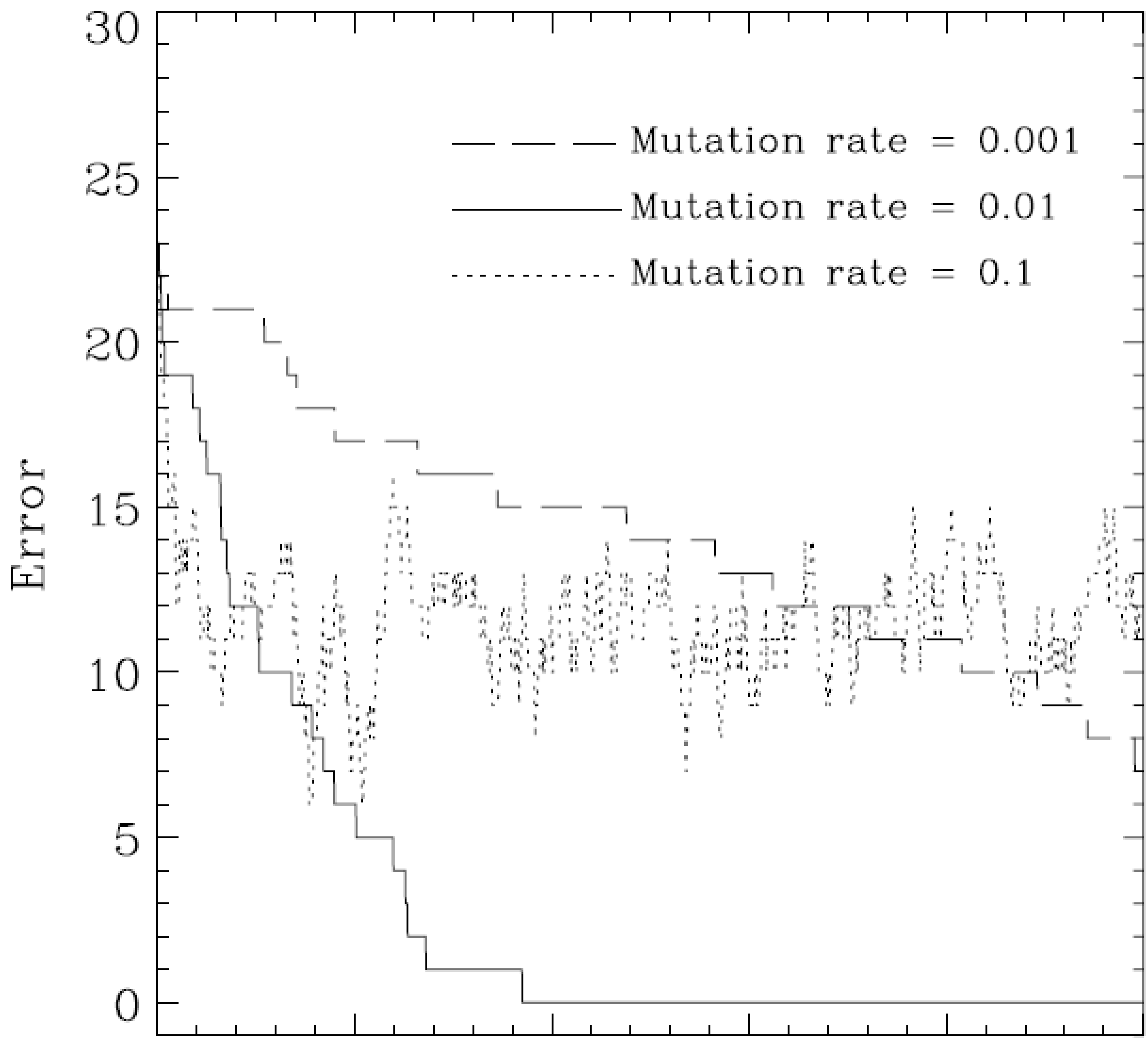
Mutation or mating?

- A low probability of mutation
- Come down with generation

Non uniform mutation

- Initially mutate heavily
- As the number of generations increase mutate less

What about mutation?



How do we determine reproduction

- Most common: Cross over
  - npoint crossover: n crossover points are randomly selected and the segments of the parents, defined by them, are exchanged for generating the offspring.
  - uniform crossover: the values of each gene in the offspring are determined by the uniform random choice of the values of this gene in the parents.

For real coding

Arithmetical crossover (Michalewicz, 1992)

Two offspring, $H_k = (h_1^k, ..., h_i^k, ..., h_n^k)$ $k = 1, 2$, are generated, where $h_i^1 = \lambda c_i^1 + (1 - \lambda)c_i^2$ and $h_i^2 = \lambda c_i^2 + (1 - \lambda)c_i^1$. λ is a constant (uniform arithmetical crossover) or varies with regard to the number of generations made (non-uniform arithmetical crossover).

- Flat crossover: An offspring, H is generated, where hi is a randomly (uniformly) chosen value of the interval [c1i; c2i]

Cross over

- Hello World
- Knap Sack
- Polynomial fit
- TSP
- Portfolio allocation

Inversion or Mutation

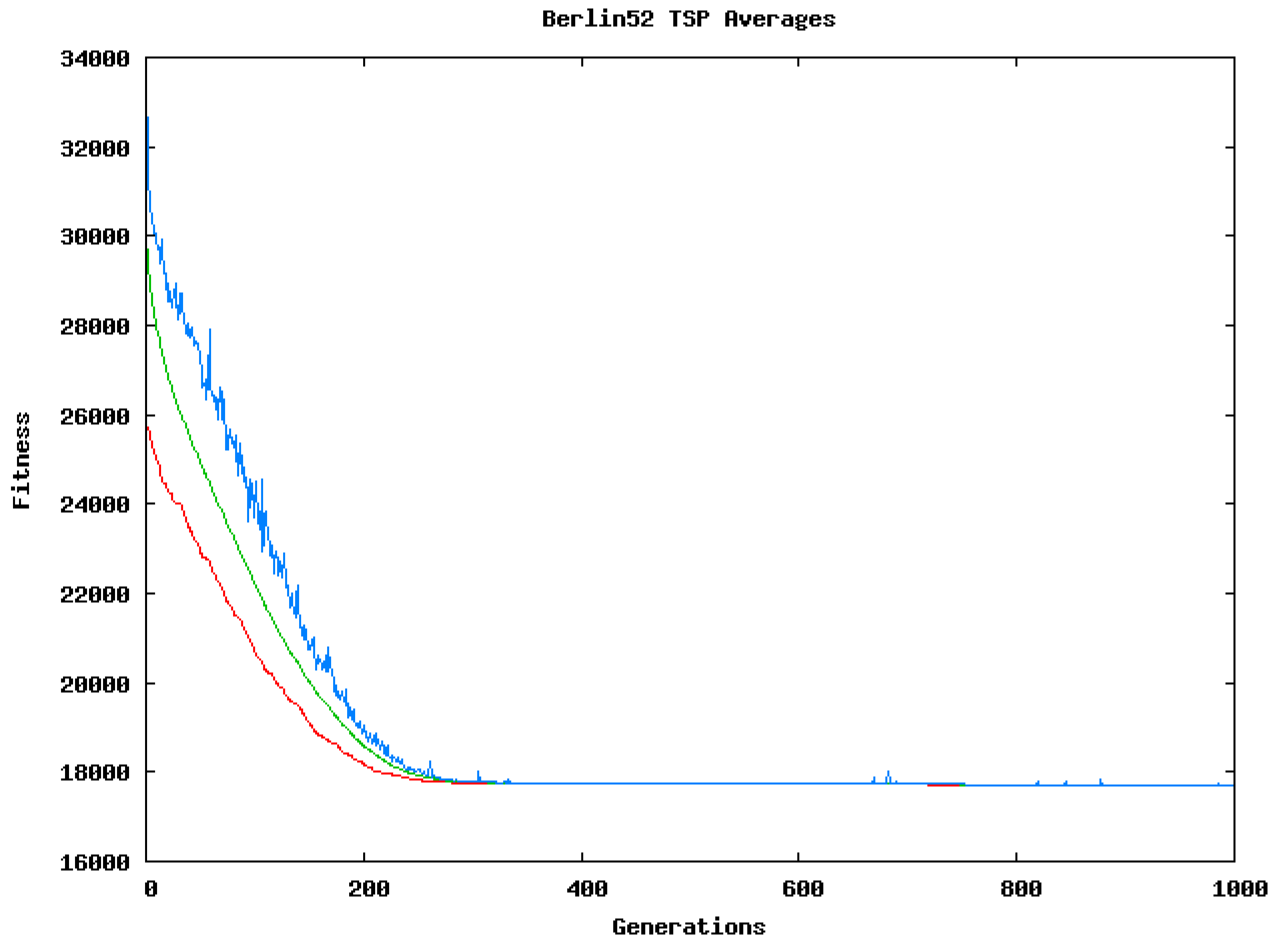
- Flip
- Random generation
- An unexpected operation

Mutation scheme

- Hello World
- Knap Sack
- Polynomial fit
- TSP
- Portfolio allocation

Convergence of a GA

When to stop?



EXAMPLES

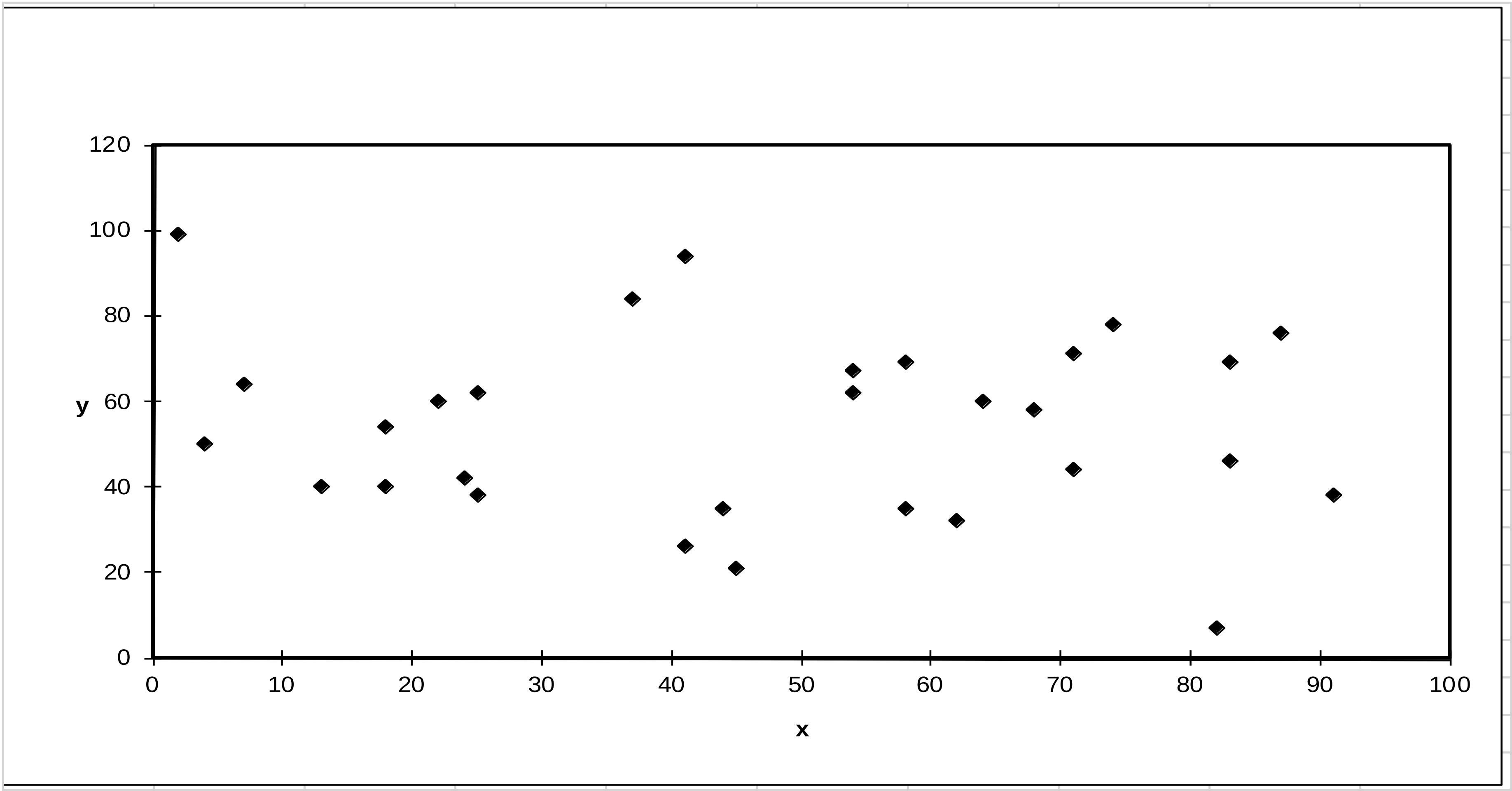
A Simple Example

The Traveling Salesman Problem:

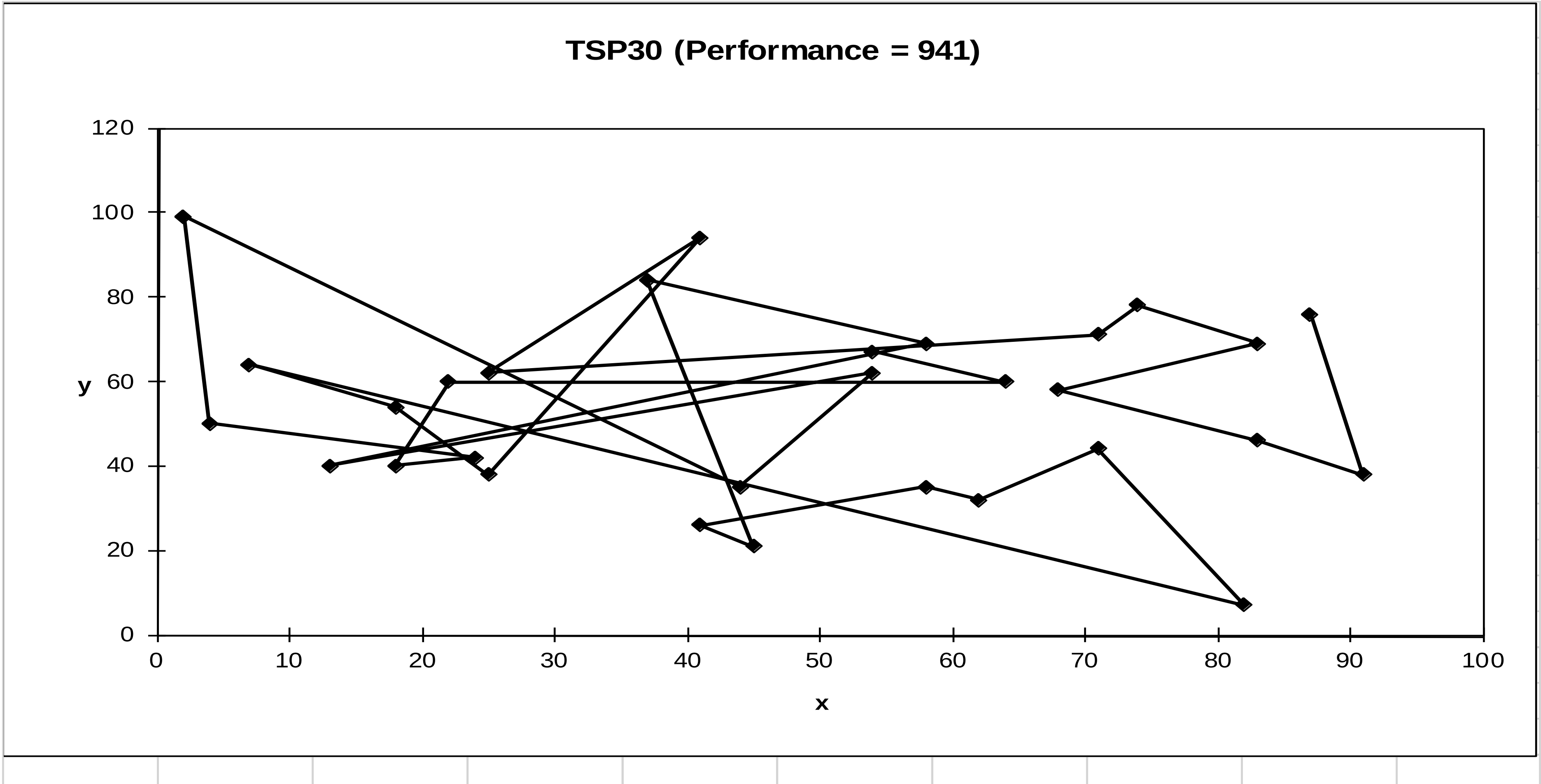
Find a tour of a given set of cities so that

- each city is visited only once
- the total distance traveled is minimized

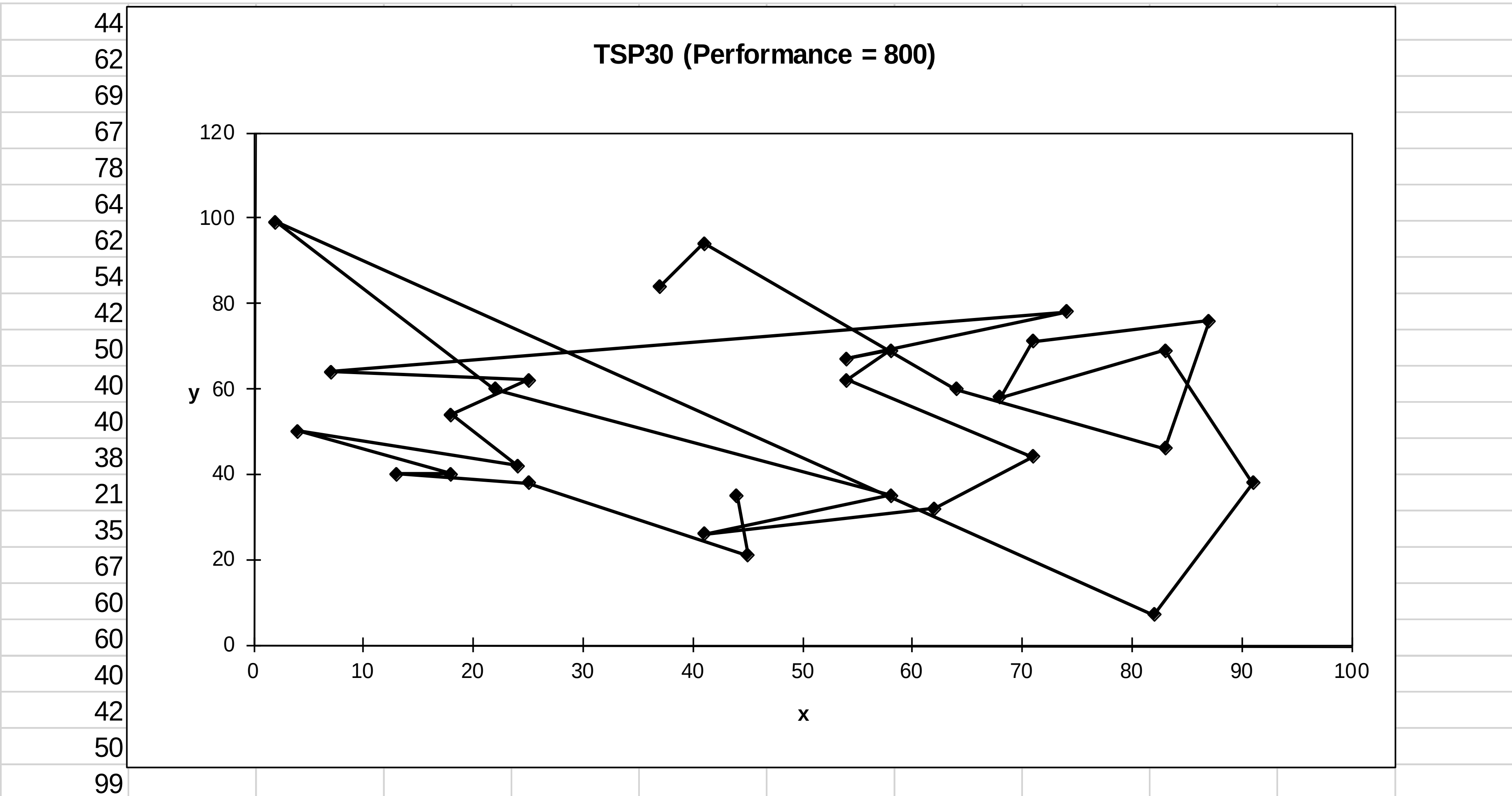
TSP Example: 30 Cities



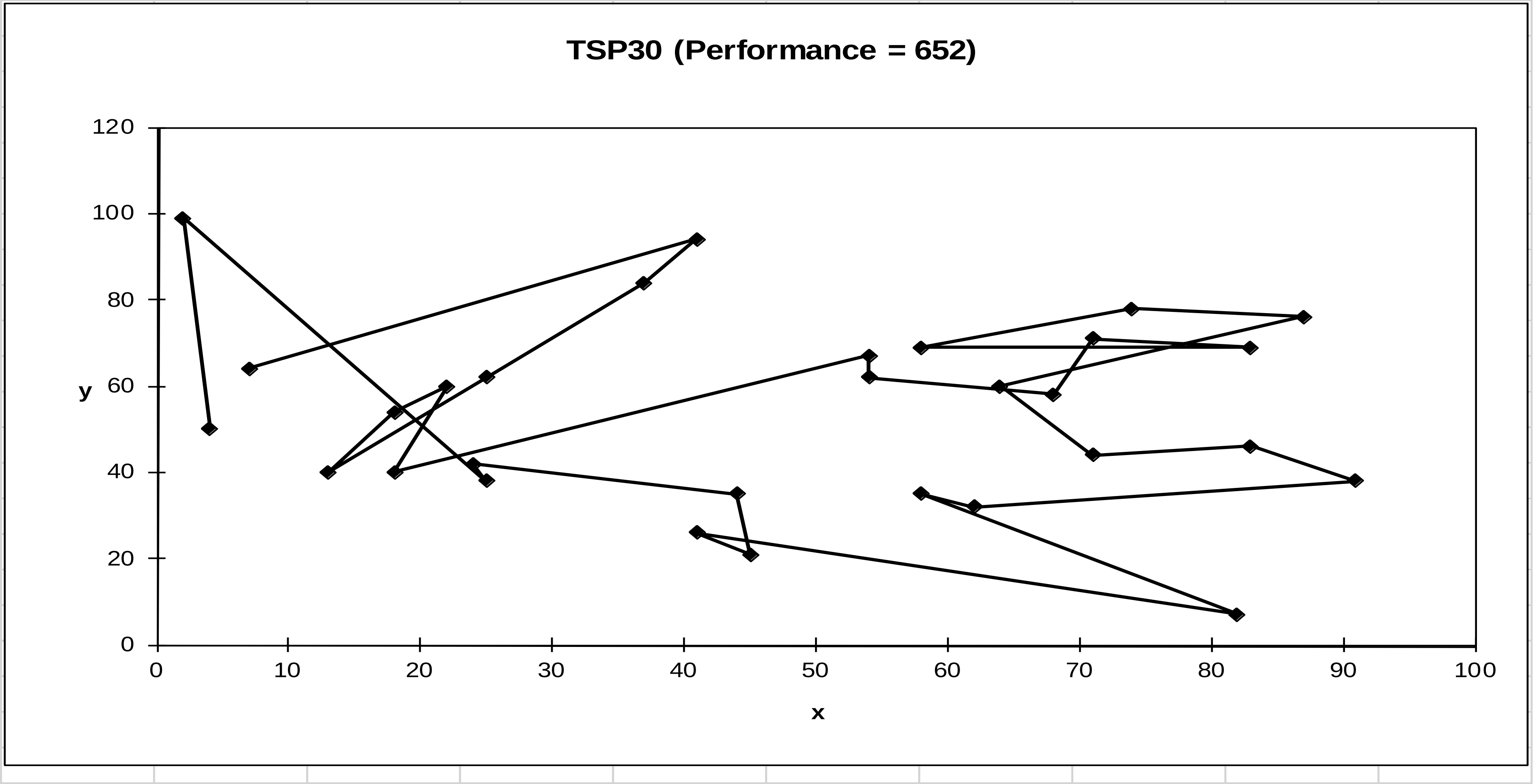
Solution ¡ (Distance = 941)



Solution j(Distance = 800)

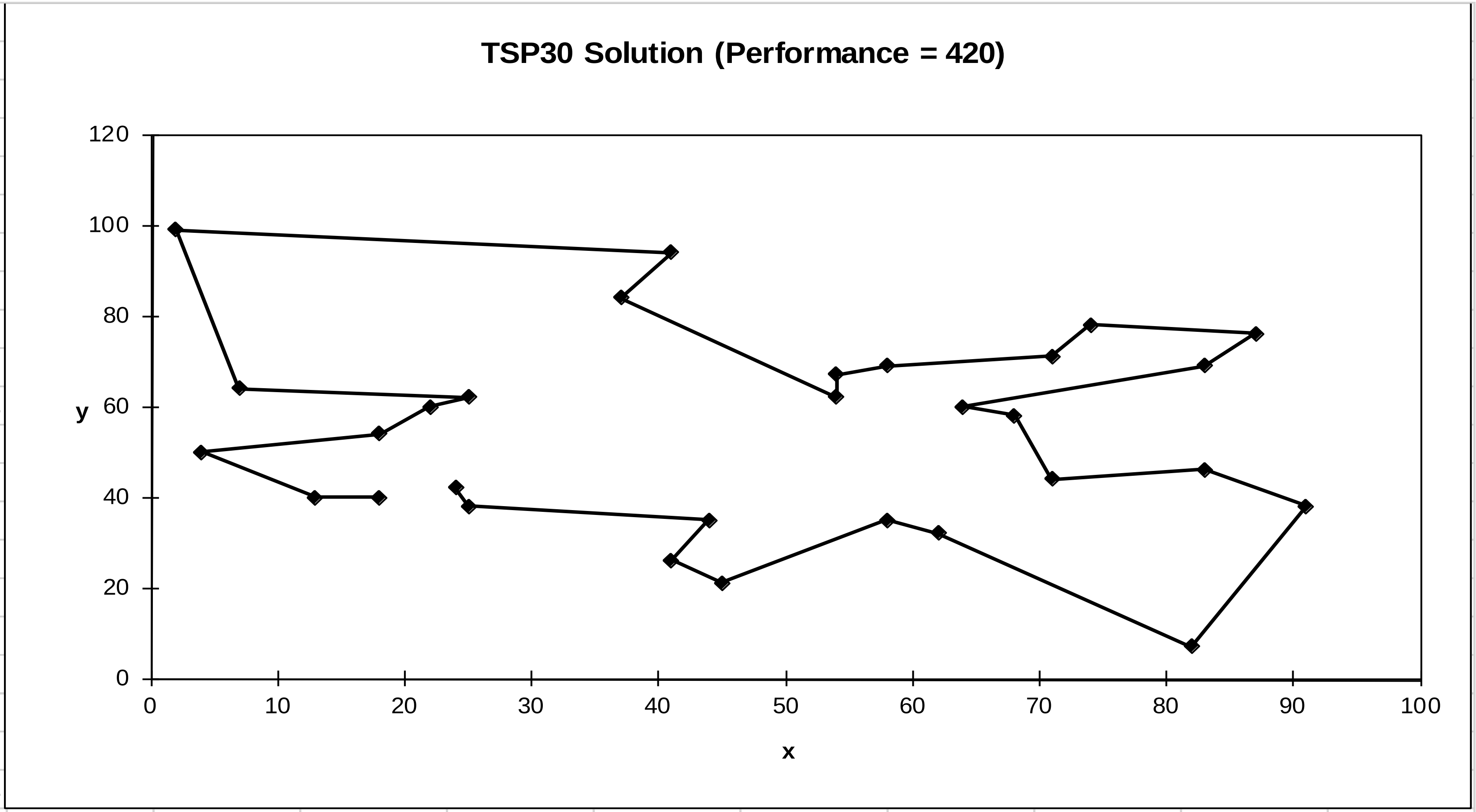


Solution $_k$(Distance = 652)

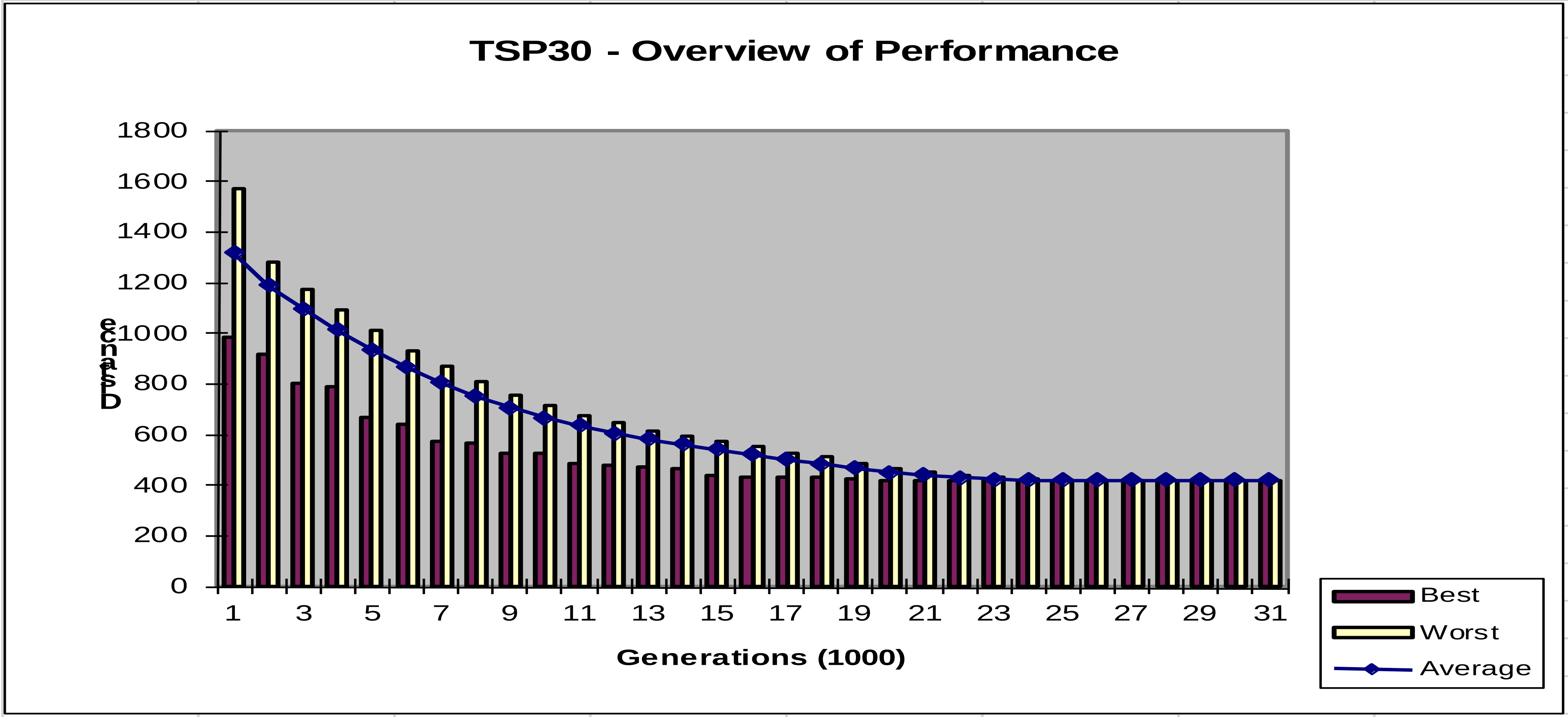


Best Solution (Distance = 420)

42
38
35
26
21
35
32
7
38
46
44
58
60
69
76
78
71
69
67
62
84
94



Overview of Performance



GA and Decision trees

- Create many trees
- Define mating (interchanging branches) and mutation (replacing a node with another)
- Define objective function (accuracy)
- Pick parents and evolve!!

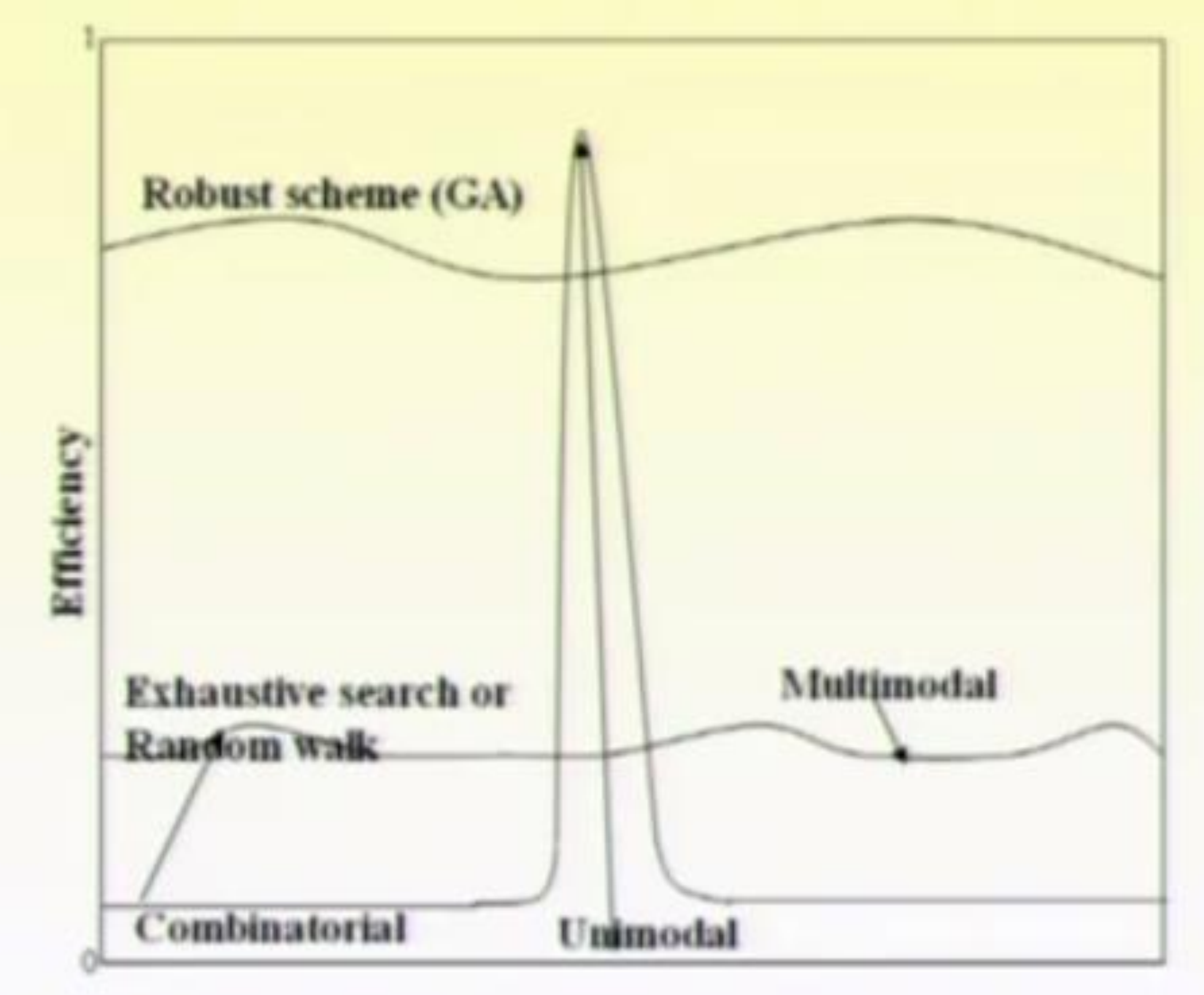
Comparison

Gradient descent and Traditional methods

- Require derivatives
- Gets stuck at local minima
- Quick where they work

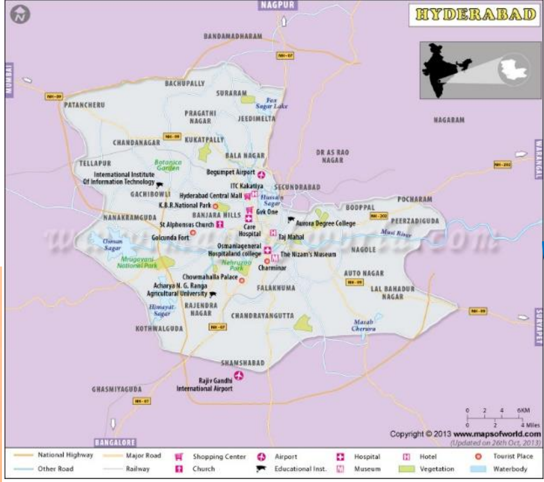
Genetic algorithms

- Substitute in the function
- Can yield global minima
- Time consuming


1
Robust scheme (GA)
Efficiency
Exhaustive search or
Random walk
Multimodal
Combinatorial
Unimodal
0

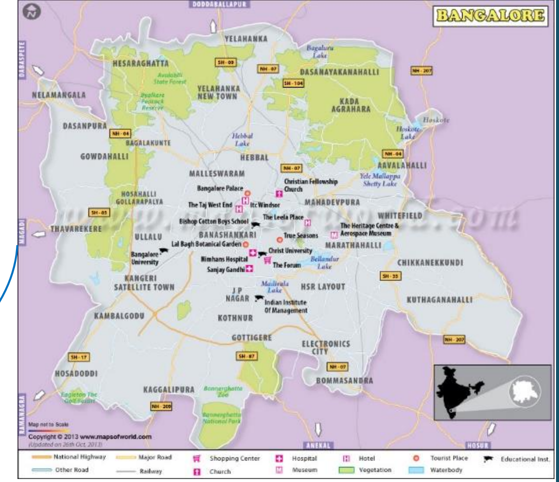
Searching for optimal solution

- Enumerative search (search all possibilities): Works for very small data sets
- Random search: Works in cases where fluctuations are not much
- Calculus/Graphical driven search: Works well, fast but gets stuck at local minima
- Evolutionary search



INSOFE

Inspire...Educate...Transform.



HYDERABAD

Office and Classrooms

Plot 63/A, Floors 1&2, Road # 13, Film Nagar,
Jubilee Hills, Hyderabad - 500 033
+91-9701685511 (Individuals)
+91-9618483483 (Corporates)

BENGALURU

Office

Incubex, #728, Grace Platina, 4th Floor, CMH Road,
Indira Nagar, 1st Stage, Bengaluru – 560038
+91-9502334561 (Individuals)
+91-9502799088 (Corporates)

Classroom

KnowledgeHut Solutions Pvt. Ltd., Reliable Plaza,
Jakkasandra Main Road, Teacher's Colony, 14th Main
Road, Sector – 5, HSR Layout, Bengaluru - 560102

Social Media

Web:	http://www.insofe.edu.in
Facebook:	https://www.facebook.com/insofe
Twitter:	https://twitter.com/Insofeedu
YouTube:	http://www.youtube.com/InsofeVideos
SlideShare:	http://www.slideshare.net/INSOFE
LinkedIn:	http://www.linkedin.com/company/international-school-of-engineering

This presentation may contain references to findings of various reports available in the public domain. INSOFE makes no representation as to their accuracy or that the organization subscribes to those findings.