Test-1 (NNFL): Model Solution

(Partial marks awarded wherever applicable for partially correct answers)

1. Compared to the Standard PSO algorithm, the Accelerated PSO algorithm favors exploration or exploitation? [1M]

Ans.: Exploitation

2. Cite two main limitations of gradient descent method. [2M]

Ans.: i) Local minima problem; ii) objective function has to be differentiable everywhere in the search space

3. What are the (i) order and (ii) defining length of the schema **1*00*01* ? [2M]

Ans.: i) 5; ii) 6

4. If crossover probability is 0.9 then what is the probability that the chromosome 11100000 will retain the schema **10*0** after uniform crossover? [4M]

Ans.: Schema: **10*0**

Probability that any of the 3 fixed bits is changed = 0.25 (for uniform crossover)

Probability that any of the 3 fixed bits remains unchanged = 0.75

Probability that all 3 fixed bits remain unchanged = 0.75^3 = 0.422

Probability that the schema will be lost = (1-0.422)*0.9 (including the effect of crossover probability)

= 0.52

Hence, probability that the schema is retained by the string = 1-0.52 =0.48

5. What is the difference between AdaGrad and RMSProp optimizations? (Answer in one short sentence without any equation) [1M]

Ans.: RMSProp gives more importance to the recent past values of gradient while AdaGrad gives equal importance to all past values.

6. In the context of GA, elitism favors exploration or exploitation? [1M]

Ans.: Exploitation

7. Two decision variables x1 and x2 are each varying between -10 and +10. If they are encoded as 5 bit binary strings each, then obtain the chromosome corresponding to the point (-4, 6). [4M]

Ans.: $\in = \frac{10-(-10)}{2^5-1} = \frac{20}{31}$

$x1 = x_{min} + \in \cdot x_{1d}$

-4 = -10 + (20/31)* x1d

x1d = 9.3 = 9 i.e. 01001

Similarly, x2d = 24.8 = 25 i.e 11001

8. In the Local Best variant of the PSO, usually the neighborhood is expanded or shrunk or left unchanged with increase in number of iterations? [1M]

Ans.: Expanded

9. What type of constraints are considered in Gradient descent with Momentum method? Equality(E) / Inequality(I) / Both E and I / Neither E nor I. [1M]

Ans.: Neither E nor I

10. In Nesterov Accelerated Gradient method, gradient is computed at a future/ past/ present value? [1M]

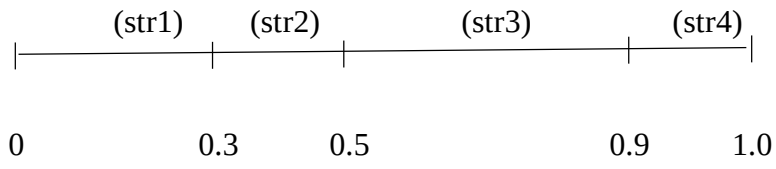
Ans.: Future

11. Consider the given GA population at the 10th Generation. Obtain the mating pool using ranking selection with the random numbers 0.48, 0.27, 0.81, 0.62. Write your answer as a), b), c), d). [4M]

GEN #10	Fitness Value	Mating Pool
string1	25	a) _____
string2	20	b) _____
string3	45	c) _____
string4	10	d) _____

Ans.:

GEN #10	Fitness Value	Rank	Selection Probability	Mating Pool
string1	25	3	0.3	string2
string2	20	2	0.2	string1
string3	45	4	0.4	string3
string4	10	1	0.1	string3
		10		



12. Compute the optimal values of x1 and x2 for the problem given below. [4M]

minimize $x_1^2 + x_2^2$

sub. to $2x_1 + 3x_2 = 4$

Ans.:

Lagrangian,

$$\mathcal{L} = x_1^2 + x_2^2 + \lambda(2x_1 + 3x_2 - 4)$$

$$\frac{\partial \mathcal{L}}{\partial x_1} = 2x_1 + 2\lambda = 0$$

$$\frac{\partial \mathcal{L}}{\partial x_2} = 2x_2 + 3\lambda = 0$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = 2x_1 + 3x_2 - 4 = 0.$$

Solving the above equations for x_1, x_2 & λ, we get-

$$x_1 = \frac{8}{13}, \quad x_2 = \frac{12}{13} \quad \text{Ans.}$$

$(f_{min} = 1.23.)$

N.B.: Those who have answered 1.23 are awarded +2.

13. The Hessian matrix as obtained at some optimal point is shown below. The point is a maximizer/ minimizer / saddle point? [1M]

$$\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -2 \\ -1 & -1 & 1 \end{bmatrix}$$

Ans.: minimizer

$$det\,(1) = 1\ ;\ det\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1\ ;\ det\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & -2 \\ -1 & -1 & 1 \end{pmatrix} = 1$$

14. Assuming static penalty with unity penalty coefficient, compute the fitness of the given GA population in the given optimization problem. Write answer as: a) b) [4M]

maximize $x_1^3 + x_2^3 + x_1^2 x_2^2$

subject to $x_1 + x_2 \leq 5$

and $x_1^2 + x_2^2 \leq 20$

Population	Fitness
1) [2, 5]	a) _____
2) [3, 3]	b) _____

Ans.:

Fitness

a) $2^3 + 5^3 + 2^2 \times 5^2 - 1\cdot(2+5-5)^2 - 1\cdot(2^2+5^2-20)^2 = 148$

b) $3^3 + 3^3 + 3^2 \times 3^2 - 1(3+3-5)^2 - 0 = 134.$

15. For the given maximization problem, assuming the parent chromosomes in RCGA to be (4,1) and (-3,3) obtain the two children chromosomes by linear crossover. [4M]

$$\text{maximize} \quad x_1^2 + 2x_2^2 + 2x_1x_2 ; \quad -5 \le x_1, x_2 \le 5.0$$

Ans.:

$P_1 = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$, $P_2 = \begin{bmatrix} -3 \\ 3 \end{bmatrix}$

$\therefore C_1 = 0.5 P_1 + 0.5 P_2 = \begin{bmatrix} 0.5 \\ 2.0 \end{bmatrix} \Rightarrow f_1 = (x_1^2 + 2x_2^2 + 2x_1x_2)\Big|_{\substack{x_1=0.5 \\ x_2=2.0}}$

$= 10.25$

$C_2 = 1.5 P_1 - 0.5 P_2 = \begin{bmatrix} 7.5 \\ 0 \end{bmatrix} \simeq \begin{bmatrix} 5.0 \\ 0.0 \end{bmatrix} \Rightarrow f_2 = 25$

$C_3 = -0.5 P_1 + 1.5 P_2 = \begin{bmatrix} -6.5 \\ 4.0 \end{bmatrix} \simeq \begin{bmatrix} -5.0 \\ 4.0 \end{bmatrix} \Rightarrow f_3 = 17$

$\therefore$ children solutions ~~are~~ are $\begin{bmatrix} 5 \\ 0 \end{bmatrix}, \begin{bmatrix} -5 \\ 4 \end{bmatrix}$

N.B.: Those who have answered

$\begin{bmatrix} 7.5 \\ 0 \end{bmatrix}$, $\begin{bmatrix} -6.5 \\ 4.0 \end{bmatrix}$

are awarded +2.

16. In Mult-class classification, if we have 10 instances to classify, how many binary classifiers we need for

(i) One vs All
(ii) One vs One 2M

(i) 10 (ii) 45

17. What is another name of Associative memory ? 1M

CONTENT ADDRESSABLE MEMORY

18. A memory which associates pencil with rubber and vice versa is called as what? 1M

BI-DIRECTIONAL HETRO-ASSOCIATIVE OR BIDIRECTIONAL ASSOCIATIVE MEMORY

19. Virus software on your computer identifies a harmless program as a malicious one. (i) What does it represent (TP/TN/FP/FN) ? and (ii) What type of error it is? 2M

FP/ TypeI

20. For IBM Watson, initial goal was to defeat champion of which game? 1M

JEOPARDY

21. Name one of Turing awardee 2018 who also helped to found MILA(the Montreal Institute for Learning Algorithms) . 1M

Yoshuo Bengio

22. Name the cateogory/technique of ANN applications which was used by Dr. John Snow to convince people that Cholera is spread by dirty water not by bad air during outbreak of cholera in 1831-45 in London. 1M

CLUSTERING

23. What kind of learning is used for cases where unlabeled data is abundant, free/cheap & labeled data is expensive and scarce. 1M

SEMI-SUPERVISED

***