Chaithanya Chikkannaswamy

SUID : 387781563

Evolutionary Machine Learning – HW2

HW2: For the same problem, use an Evolution Strategies approach, and compare with the previous results (of HW1).

Dataset: Bank Marketing Dataset

Link:
https://www.kaggle.com/chaithanya96/bankmarketing

The bank marketing dataset is the csv file with users details used to predict the marketing decision. The decision yes or no is represented with binary numbers 1 and 0. The sample columns included are:

1. 'age',
2. 'job',
3. 'marital',
4. 'education',
5. 'default',
6. 'balance',
7. 'housing',
8. 'loan',
9. 'contact',
10. 'day',
11. 'month',
12. 'duration',
13. 'campaign',
14. 'pdays',
15. 'previous',
16. 'poutcome',
17. 'market?'

The screenshot of the sample dataset is attached below:

Chaithanya Chikkannaswamy

bank marketing 2

age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	market?
30	unemployed	married	primary	0	1787	0	0	cellular	19	oct	79	1	-1	0	unknown	0
33	services	married	secondary	0	4789	1	1	cellular	11	may	220	1	339	4	failure	0
35	management	single	tertiary	0	1350	1	0	cellular	16	apr	185	1	330	1	failure	0
30	management	married	tertiary	0	1476	1	1	unknown	3	jun	199	4	-1	0	unknown	0
59	blue-collar	married	secondary	0	0	1	0	unknown	5	may	226	1	-1	0	unknown	0
35	management	single	tertiary	0	747	0	0	cellular	23	feb	141	2	176	3	failure	0
36	self-employed	married	tertiary	0	307	1	0	cellular	14	may	341	1	330	2	other	0
39	technician	married	secondary	0	147	1	0	cellular	6	may	151	2	-1	0	unknown	0
41	entrepreneur	married	tertiary	0	221	1	0	unknown	14	may	57	2	-1	0	unknown	0
43	services	married	primary	0	-88	1	1	cellular	17	apr	313	1	147	2	failure	0
39	services	married	secondary	0	9374	1	0	unknown	20	may	273	1	-1	0	unknown	0
43	admin.	married	secondary	0	264	1	0	cellular	17	apr	113	2	-1	0	unknown	0
36	technician	married	tertiary	0	1109	0	0	cellular	13	aug	328	2	-1	0	unknown	0
20	student	single	secondary	0	502	0	0	cellular	30	apr	261	1	-1	0	unknown	1
31	blue-collar	married	secondary	0	360	1	1	cellular	29	jan	89	1	241	1	failure	0
40	management	married	tertiary	0	194	0	1	cellular	29	aug	189	2	-1	0	unknown	0
56	technician	married	secondary	0	4073	0	0	cellular	27	aug	239	5	-1	0	unknown	0
37	admin.	single	tertiary	0	2317	1	0	cellular	20	apr	114	1	152	2	failure	0
25	blue-collar	single	primary	0	-221	1	0	unknown	23	may	250	1	-1	0	unknown	0
31	services	married	secondary	0	132	0	0	cellular	7	jul	148	1	152	1	other	0
38	management	divorced	unknown	0	0	1	0	cellular	18	nov	96	2	-1	0	unknown	0
42	management	divorced	tertiary	0	16	0	0	cellular	19	nov	140	3	-1	0	unknown	0
44	services	single	secondary	0	106	0	0	unknown	12	jun	109	2	-1	0	unknown	0
44	entrepreneur	married	secondary	0	93	0	0	cellular	7	jul	125	2	-1	0	unknown	0
26	housemaid	married	tertiary	0	543	0	0	cellular	30	jan	169	3	-1	0	unknown	0

Code execution steps:

CMA – Evolutionary Strategy

I have implemented the CMA Evolution strategy to train a shallow feedforward neural network for a 2-class classification task. The accuracy is then compared with the accuracy calculated using Genetic Algorithm implementation obtained in HW1.

The snapshot of the model summary is attached below:

```
model = Sequential()
model.add(Dense(20, activation='sigmoid', input_shape=(10,)))
model.add(Dense(1, activation='sigmoid'))
model.summary()
```

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense_1 (Dense)              (None, 20)                220
_________________________________________________________________
dense_2 (Dense)              (None, 1)                 21
=================================================================
Total params: 241
Trainable params: 241
Non-trainable params: 0
_________________________________________________________________
```

+ Code + Markdown

1. Cal_fitness function is defined to calculate the fitness of the individuals and to take the maximum fitness to find out the best solution.
2. The fitness calculation is done for several 100 generation and each time, the max fitness values are compared to store the maximum fitness value in each generation.
3. The program executed until the number of generations is 100.
4. The obtained accuracy is 0.71.

```
print(accuracy_score)
```

```
this is the Mean Squared Error: 0.19037290790570838
this is the confusion matrix
[[768 252]
 [ 70  41]]
Accuracy Score 0.7152961980548187
[0.7152961980548187]
```

+ Code + Markdown

Chaithanya Chikkannaswamy

Conclusion:

The **CMA-ES** resulted in 0.715 test accuracy. **Genetic Neural Network** resulted in a test accuracy of 0.89 The **Sequential Neural Network** resulted in a test accuracy of 0.88 and the

References

1. http://cma.gforge.inria.fr/apidocs-pycma/cma.evolution_strategy.CMAEvolutionStrategy.html
2. https://github.com/hardmaru/estool/blob/master/simple_es_example.ipynb
3. Discussed with classmates.