

## تمرین رگرسیون خطی

مونا ثقفی

9912762290

**تابع ( fitness\_func )** امتیاز فیتنس را برای هر راه حل در محیط ژنها محاسبه می کند. این کار را با حلقه زدن از طریق هر نمونه در داده های آموزشی انجام می دهد و برای هر نمونه، خروجی را با استفاده از ضرایب حل، محاسبه می کند و آن را با مقدار خروجی واقعی آن نمونه مقایسه می کند. سپس مجذور خطا را بین مقادیر خروجی پیش بینی شده و واقعی محاسبه کرده و خطا را در تمام نمونه های آموزشی جمع آوری می کند. سپس این مقدار خطای انباشته شده معکوس می شود (1/err) برای محاسبه امتیاز فیتنس. هر چه خطا کمتر باشد، نمره فیتنس بیشتر است.

**تابع ( test )** برای ارزیابی بهترین راه حل یافت شده توسط الگوریتم ژنتیک استفاده می شود. با استفاده از بهترین ضرایب حل (b0 تا b10) از طریق هر نمونه در داده های آزمایشی در حلقه بررسی میکند و مقدار خروجی را محاسبه می کند. سپس مجذور خطا بین مقادیر خروجی پیش بینی شده و واقعی را در تمام نمونه های آزمایشی جمع آوری می کند که نشان دهنده میزان توانایی مدل برای تعمیم است.

```
num_generations = 100
num_parents_mating = 4

sol_per_pop = 10
num_genes = 11

init_range_low = -2
init_range_high = 5

parent_selection_type = "sss"
keep_parents = 1

crossover_type = "single_point"

mutation_type = "random"
mutation_percent_genes = 10
```

### Parameters of the best solution :

[-0.10748551 -1.27147218 -1.9723521 0.73137597 -0.0518103 1.88474409 0.75541696 -  
1.33153793 1.04750576 -0.41986103 1.62998818]

**Fitness value of the best solution = 1.0273662529270258e-07**

**test function output : 38636.123254089805**

دلیل دقیق بودن جواب این است که با جنریشن 100 و محاسبه با 10 تا ستون جدول، اگر نتیجه را تقسیم بر 10 کنیم (میانگین بگیریم) و بدلیل مجذور کردن اختلاف ها، از میانگین جذر میگیریم که برای این مقادیر داریم:

$$38636/10 \approx 3800$$

$$\sqrt{3800} \approx 61.6 \Rightarrow 60$$

یعنی اختلاف جواب بدست آمده از تابع برنامه  $\pm 60$  از داده های جدول میباشد که این رقم با توجه به جنریشن 100 عدد مناسب و معقولی میباشد. در مسئله بیت کوین یعنی اختلاف حدود ۶۰ دلار با قیمت واقعی در جدول داده.