

## به نام خدا

پدرام رستمی - ۹۵۳۱۰۳۲

### گزارش پروژه رگراسیون

در این پروژه قصد داریم با استفاده از الگوریتم ژنتیک، با یک معادله‌ی خط درجه ۳ یک معادله را رگراسیون کنیم. در این مسئله، **individual** ها همان ضرایب اند. در این گزارش به بررسی تأثیر متغیرهای مسئله‌ی ژنتیک در میزان همگرایی دو معادله خط می‌پردازیم.

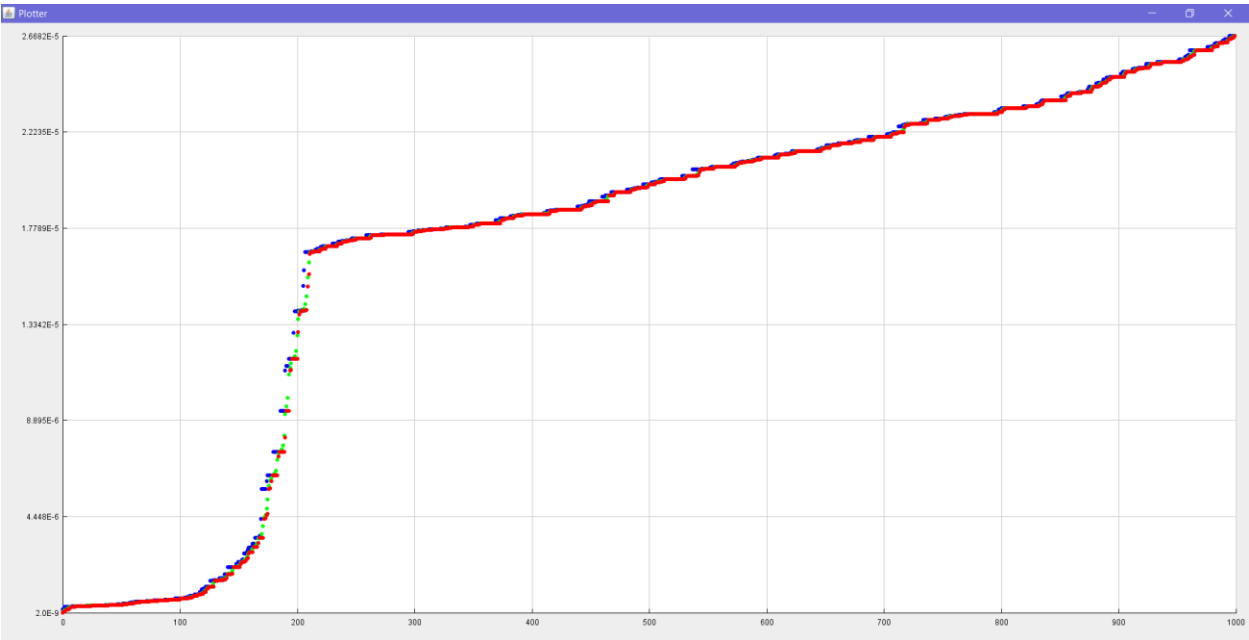
### تأثیر متغیر تعداد نسل‌ها (Number of Generations)

برای تشخیص شناسایی تأثیر تعداد نسل‌ها بر این مسئله، کفایت متغیر **NumberOfGeneration** را برابر مقادیر ۱۰۰۰، ۲۵۰۰ و ۵۰۰۰ قرار دهیم و بقیه‌ی متغیرها را ثابت فرض کنیم. برای کاهش تأثیر جمعیت اولیه‌ی تصادفی، برای هر ۳ مقدار ۱۰۰۰، ۲۵۰۰ و ۵۰۰۰ از جمعیت اولیه‌ی تصادفی یکسانی استفاده می‌کنیم.

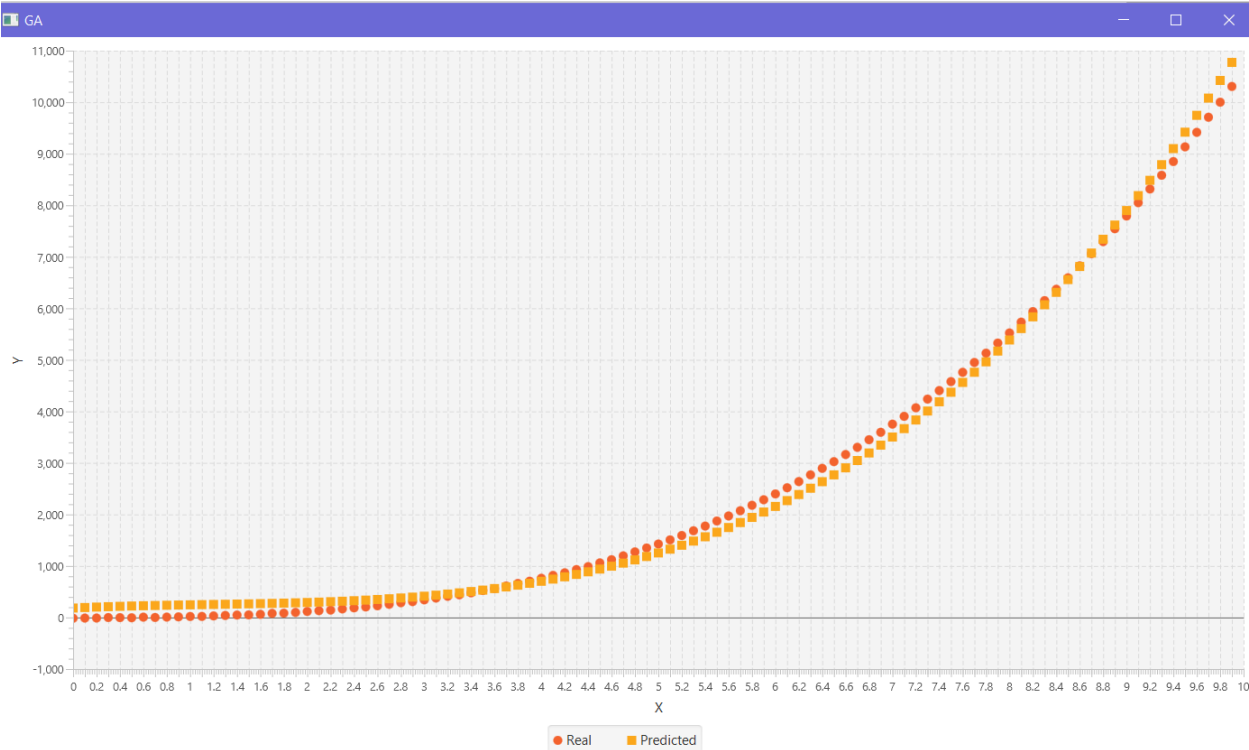
تست (۱) تأثیر تعداد نسل‌ها

NumberOfGenerations = 1000 , PopulationSize = 50, TournamentSize = 2, MutationRate = 0.02, Variance = 0.01

در این تست بازه‌ی انتخاب تصادفی ضرایب بین -۲۰۰ تا ۲۰۰ است.



شکل ۱ – نمودار بهترین، بدترین و میانگین شایستگی تست ۱



شکل ۲ – مقایسه‌ی حالت پیشبینی شده با واقعیت تست ۱

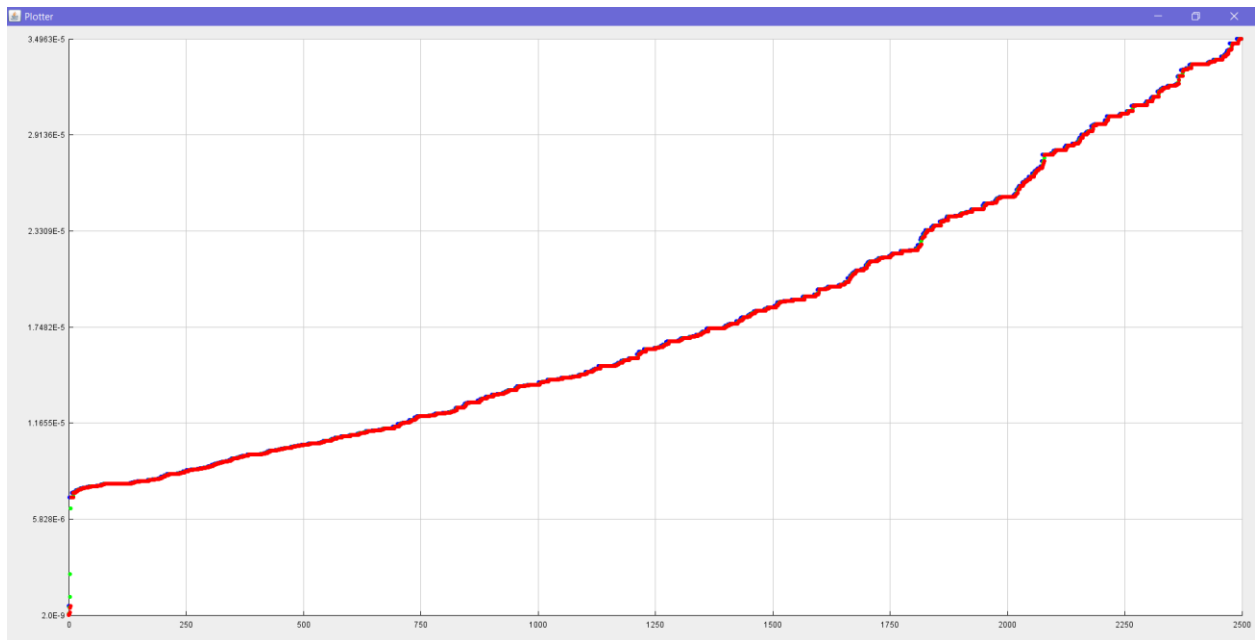
معادله‌ی پیشبینی شده :

$$y = 12.057x^3 - 27.604x^2 + 154.131x - 145.417$$

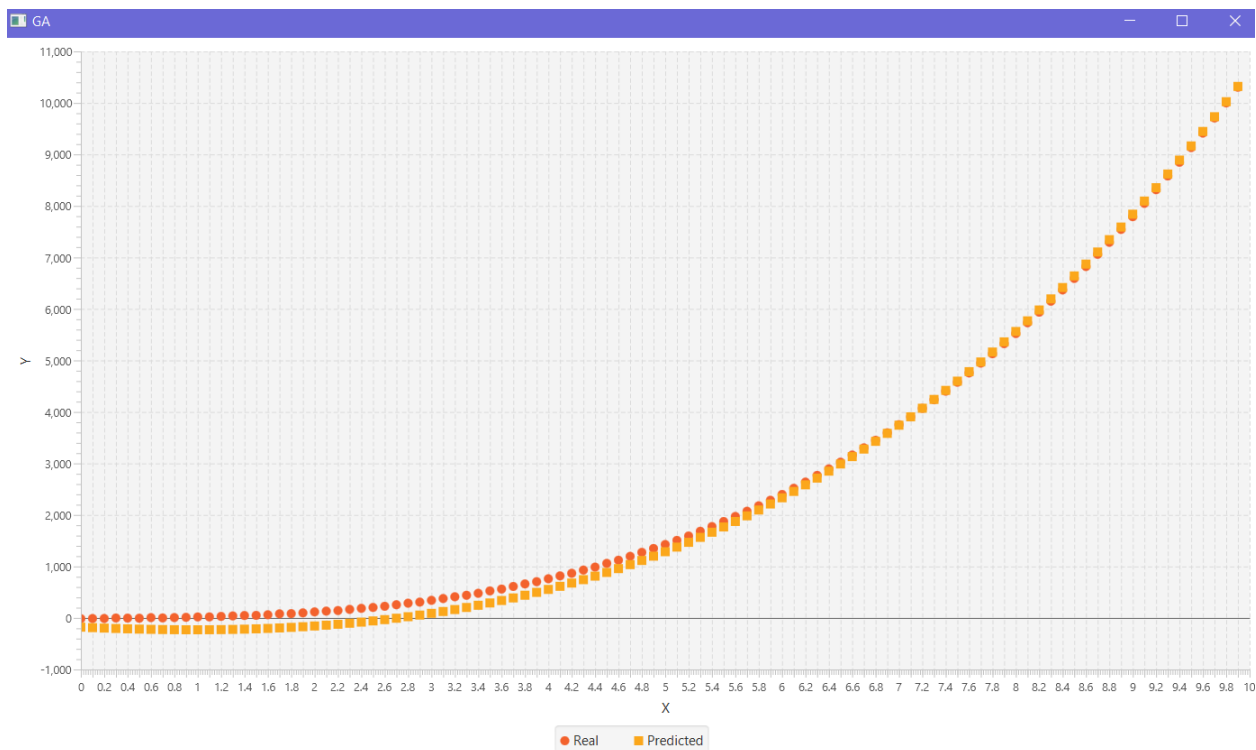
## تست (۲) تأثیر تعداد نسل‌ها

*NumberOfGenerations = 2500 , PopulationSize = 50, TournamentSize = 2, MutationRate = 0.02, Variance = 0.01*

در این تست بازه‌ی انتخاب تصادفی ضرایب بین -۲۰۰ تا ۲۰۰ است.



شکل ۳ – نمودار بهترین، بدترین و میانگین شایستگی تست ۲



شکل ۴ – مقایسه‌ی حالت پیش‌بینی شده با واقعیت تست ۲

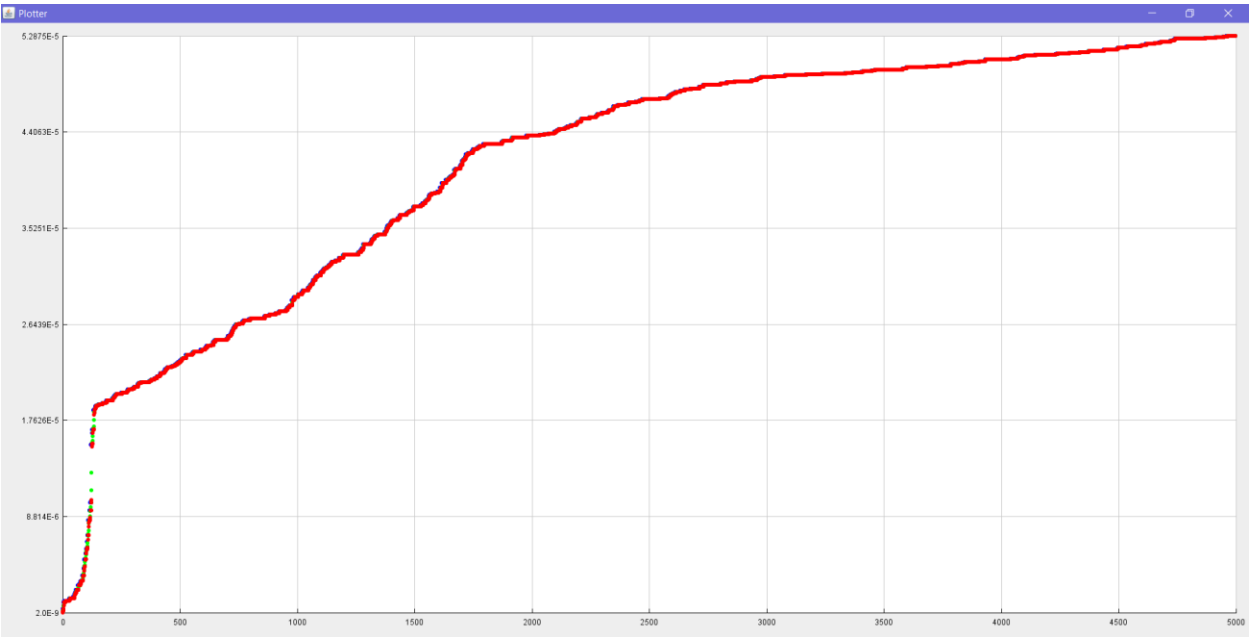
معادله‌ی پیش‌بینی‌شده:

$$y = 7.947x^3 + 38.217x^2 - 96.761x - 178.633$$

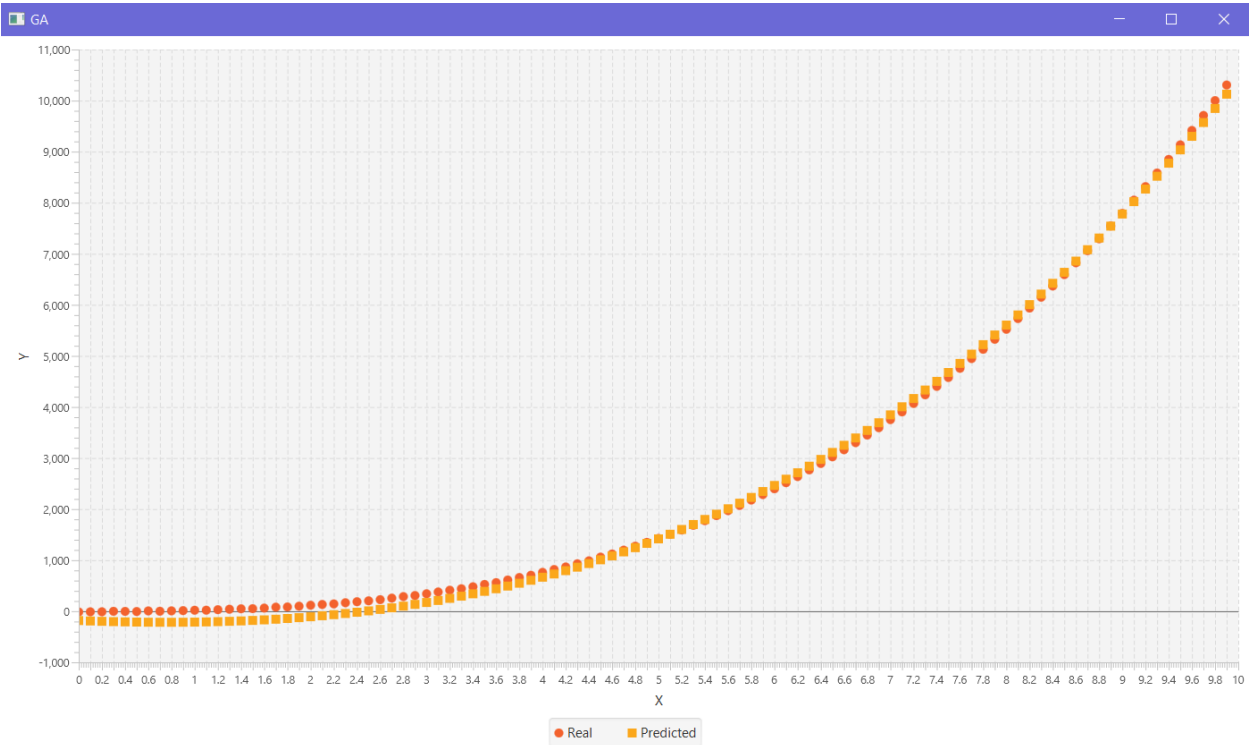
تست (۳) تأثیر تعداد نسل‌ها

NumberOfGenerations = 5000 , PopulationSize = 50, TournamentSize = 2, MutationRate = 0.02, Variance = 0.01

در این تست بازه‌ی انتخاب تصادفی ضرایب بین -۲۰۰ تا ۲۰۰ است.



شکل ۵ – نمودار بهترین، بدترین و میانگین شایستگی تست ۳



شکل ۶ – مقایسه‌ی حالت پیش‌بینی شده با واقعیت تست ۳

معادله‌ی پیش‌بینی‌شده:

$$y = 6.708x^3 + 47.238x^2 - 83.942x - 182.423$$

نتیجه گیری

همانطور که در تست های (۱)، (۲) و (۳) و جدول ۱ مشخص است، با افزایش تعداد نسل ها ضرایب بهتری پیدا می شوند.

تعداد نسل ها	فیتنس آخرین نسل
1000	2.6682E-5
2500	3.4963E-5
5000	5.2875E-5

جدول ۱ – تأثیر تعداد نسل ها بر شایستگی آخرین نسل

با افزایش تعداد نسل ها، تعداد ژن های قویتر در جمعیت بیشتر شده و جمعیت قوی تر می شود. در نتیجه با افزایش تعداد نسل ها، جمعیت شایسته تر می شود.

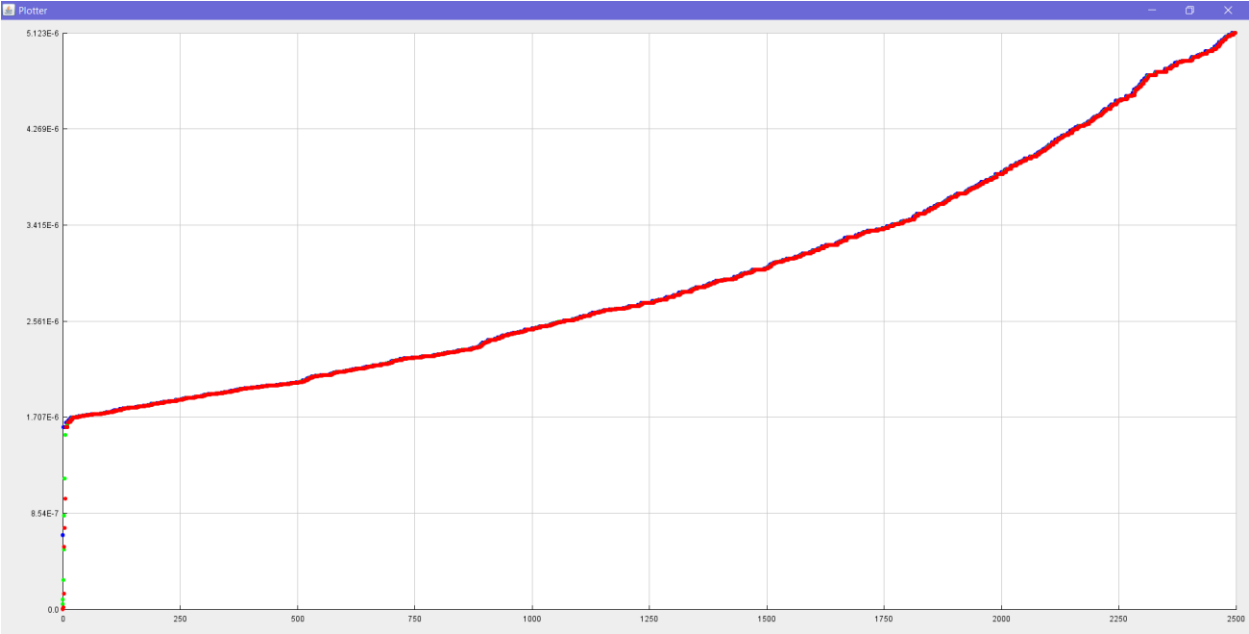
تأثیر متغیر اندازه ی جمعیت (Population Size)

برای تشخیص شناسایی تأثیر اندازه ی جمعیت، کافیت متغیر PopulationSize را برابر مقادیر ۵۰، ۱۰۰ و ۲۰۰ قرار می دهیم و بقیه ی متغیرها را ثابت فرض می کنیم.

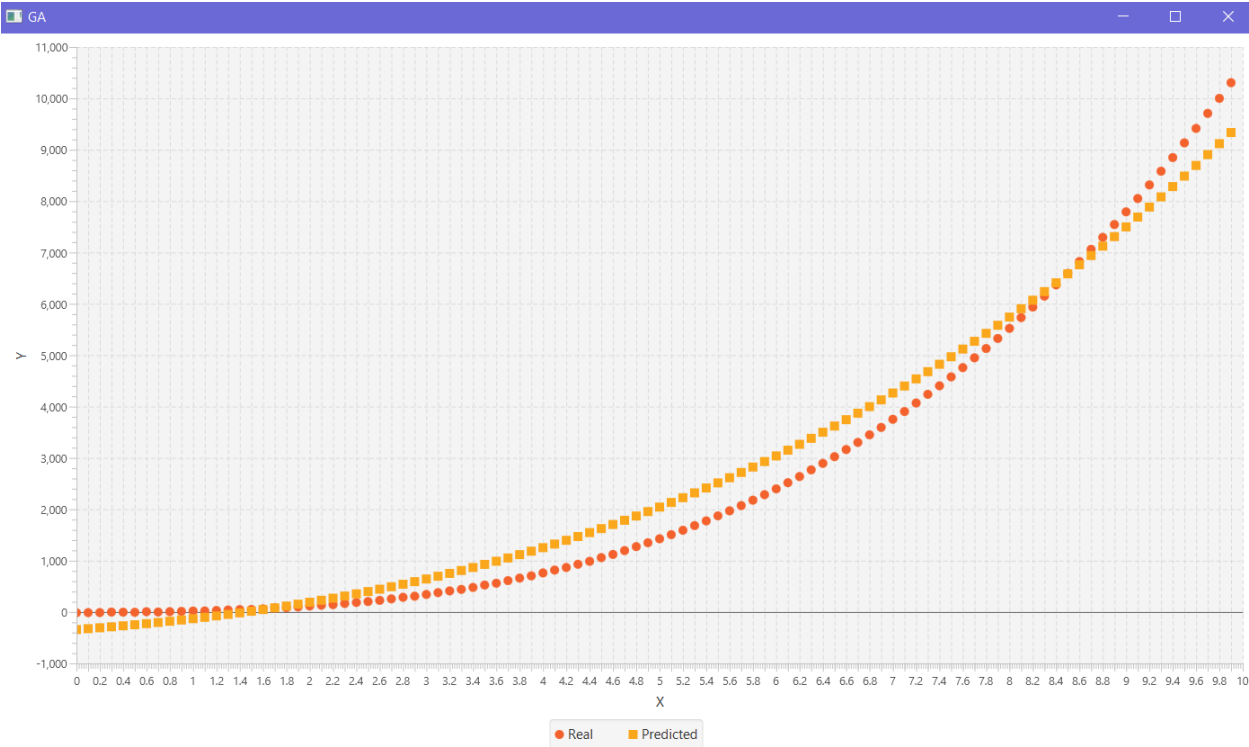
تست (۴) تأثیر اندازه‌ی جمعیت

*NumberOfGenerations = 2500 , PopulationSize = 50, TournamentSize = 2, MutationRate = 0.02, Variance = 0.01*

در این تست بازه‌ی انتخاب تصادفی ضرایب بین -۴۰۰ تا ۴۰۰ است.



شکل ۷ – نمودار بهترین، بدترین و میانگین شایستگی تست ۴



شکل ۸ – مقایسه‌ی حالت پیشبینی شده با واقعیت تست ۴

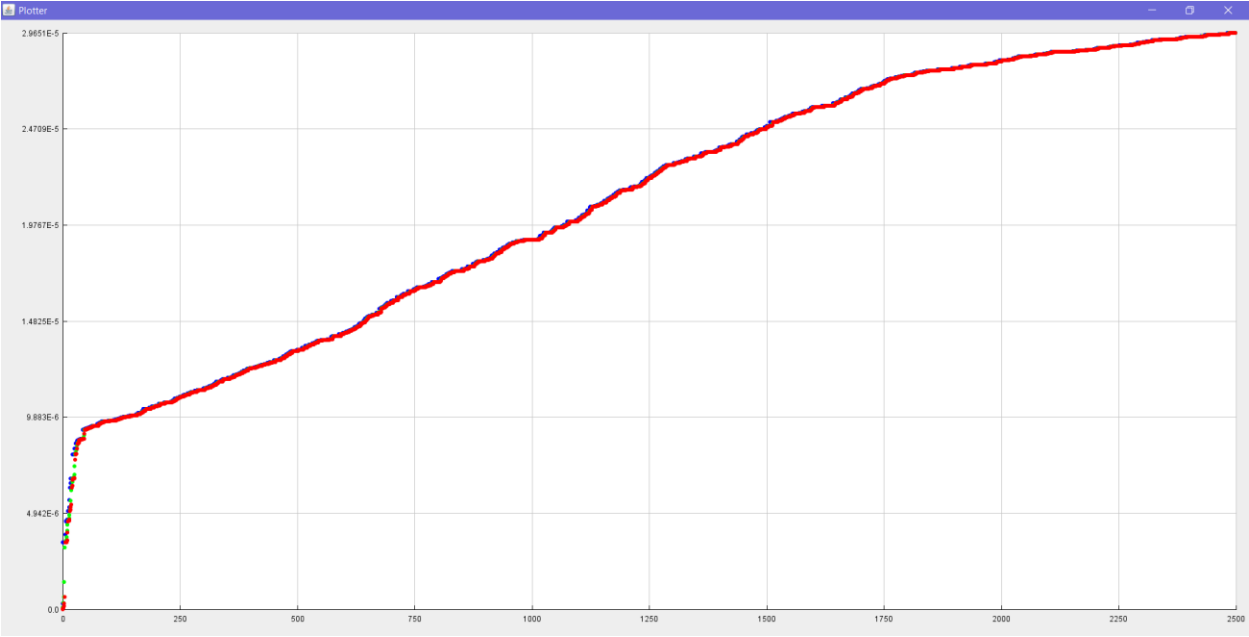
معادله‌ی پیشبینی‌شده:

$$y = 4.019x^3 + 42.229x^2 + 165.258x - 340.864$$

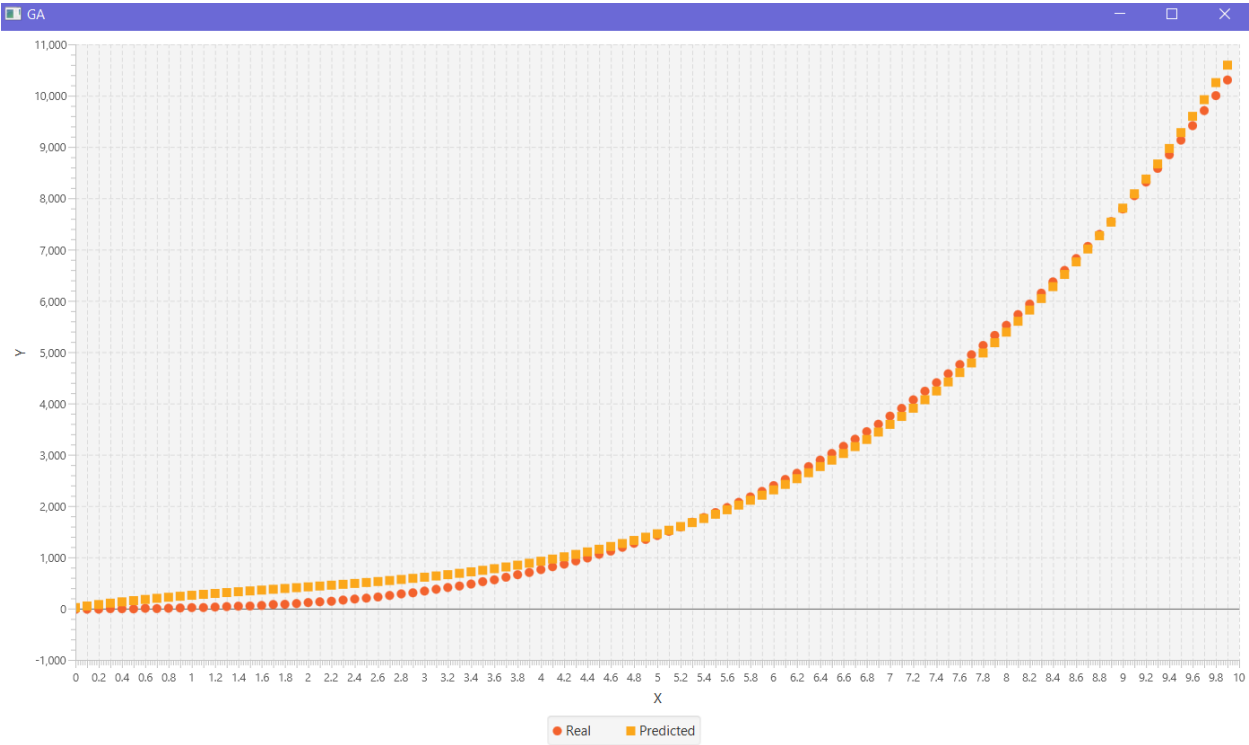
تست (۵) تأثیر اندازهی جمعیت

NumberOfGenerations = 2500 , PopulationSize = 100, TournamentSize = 2, MutationRate = 0.02, Variance = 0.01

در این تست بازه‌ی انتخاب تصادفی ضرایب بین -۴۰۰ تا ۴۰۰ است.



شکل ۹ – نمودار بهترین، بدترین و میانگین شایستگی تست ۵



شکل ۱۰ – مقایسه‌ی حالت پیشبینی شده با واقعیت تست ۵

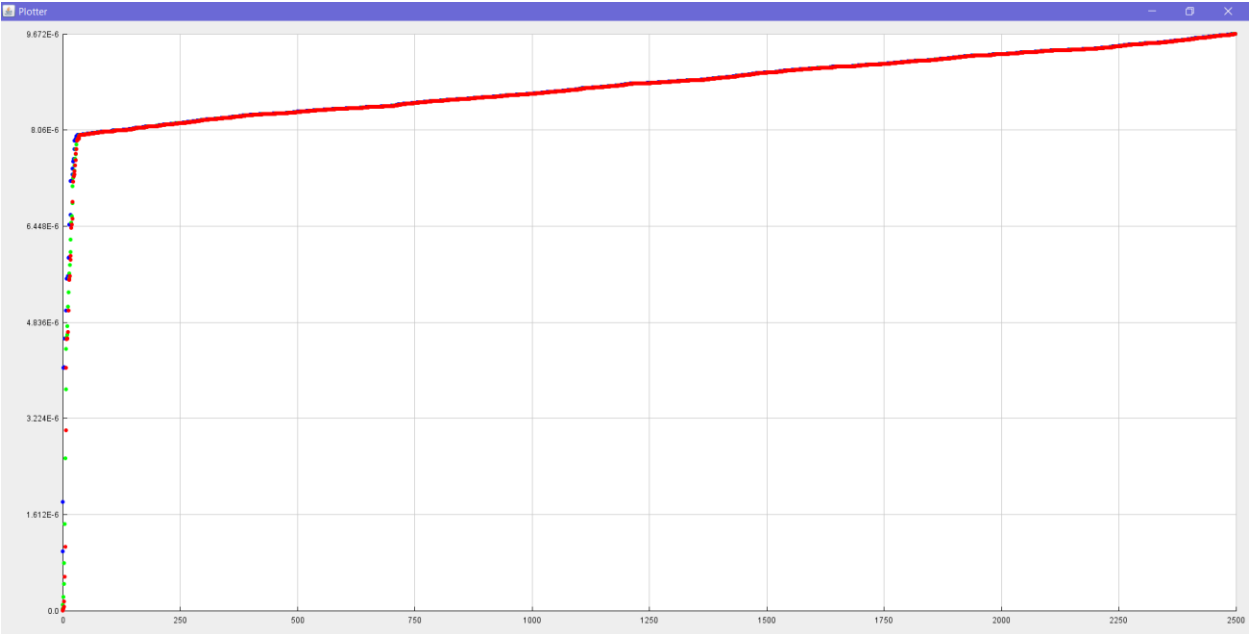
معادله‌ی پیشینی‌شده:

$$y = 16.541x^3 - 87.182x^2 + 309.849x + 21.136$$

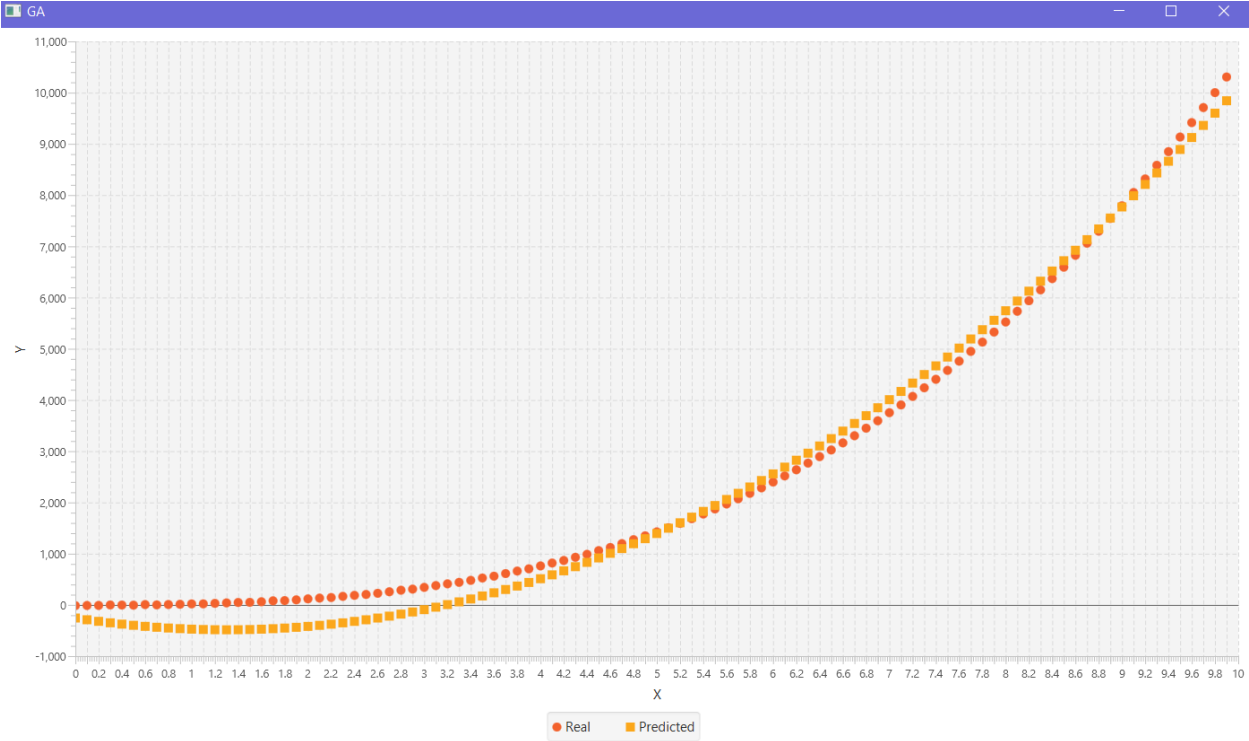
تست (۶) تأثیر اندازهی جمعیت

NumberOfGenerations = 2500 , PopulationSize = 200, TournamentSize = 2, MutationRate = 0.02, Variance = 0.01

در این تست بازه‌ی انتخاب تصادفی ضرایب بین -۴۰۰ تا ۴۰۰ است.



شکل ۱۱ – نمودار بهترین، بدترین و میانگین شایستگی تست ۶



شکل ۱۲ – مقایسه‌ی حالت پیشبینی شده با واقعیت تست ۶

معادله‌ی پیشینی‌شده:

$$y = 0.455x^3 + 134.029x^2 - 351.671x - 255.750$$



نتیجه گیری

همانطور که از نتایج تست های (۴)، (۵) و (۶) و جدول ۲ مشخص است، می توان گفت اندازه ی جمعیت نسبت مستقیمی با ضرایب تولید شده نداشت.

اندازه ی جمعیت	فیتنس آخرین نسل
50	5.123E-6
100	2.9651E-5
200	9.672E-6

جدول ۲ – تأثیر اندازه ی جمعیت بر شایستگی نسل آخر

احتمال انتخاب تصادفی یک individual با شایستگی خوب ربطی به اندازه ی جمعیت ندارد. در نتیجه نمی توان تأثیر مستقیمی بین این متغیر با میزان همگرایی رابطه ی مستقیمی پیدا کرد.

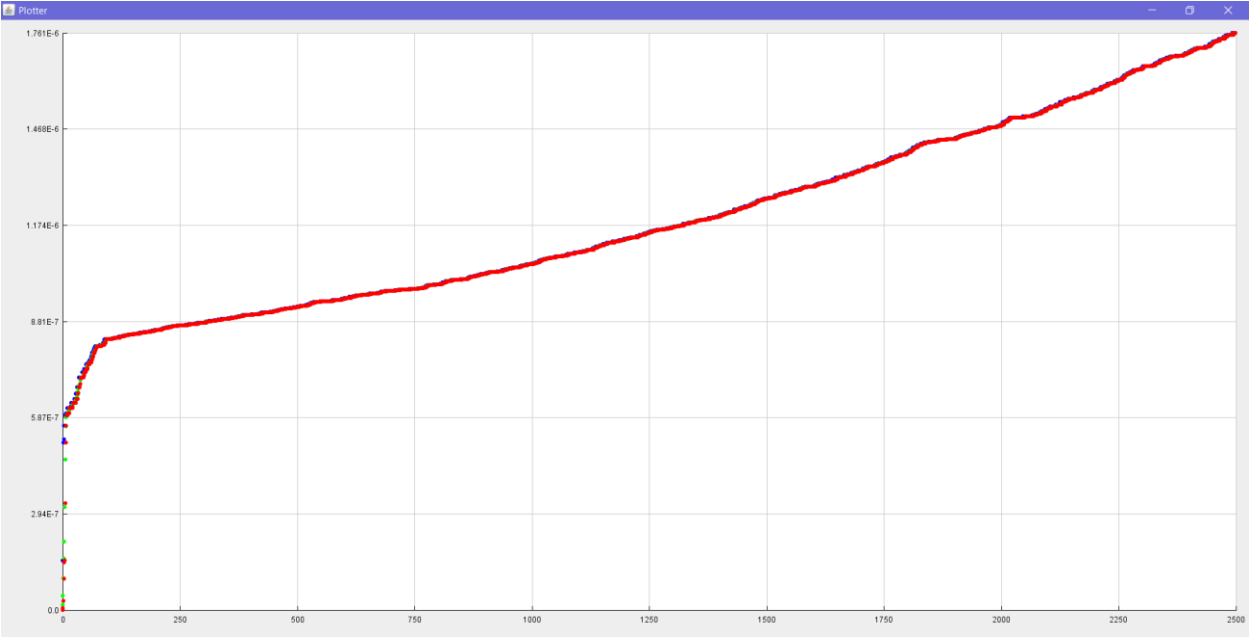
تأثیر متغیر Tournament Size

برای تشخیص تأثیر Tournament Size بر این مسئله، کفایت این متغیر را برابر مقادیر ۲، ۵ و ۱۰ قرار دهیم و بقیه ی متغیرها را ثابت فرض کنیم. برای کاهش تاثیر جمعیت اولیه ی تصادفی، برای هر ۳ مقدار از جمعیت اولیه ی تصادفی یکسانی استفاده می کنیم.

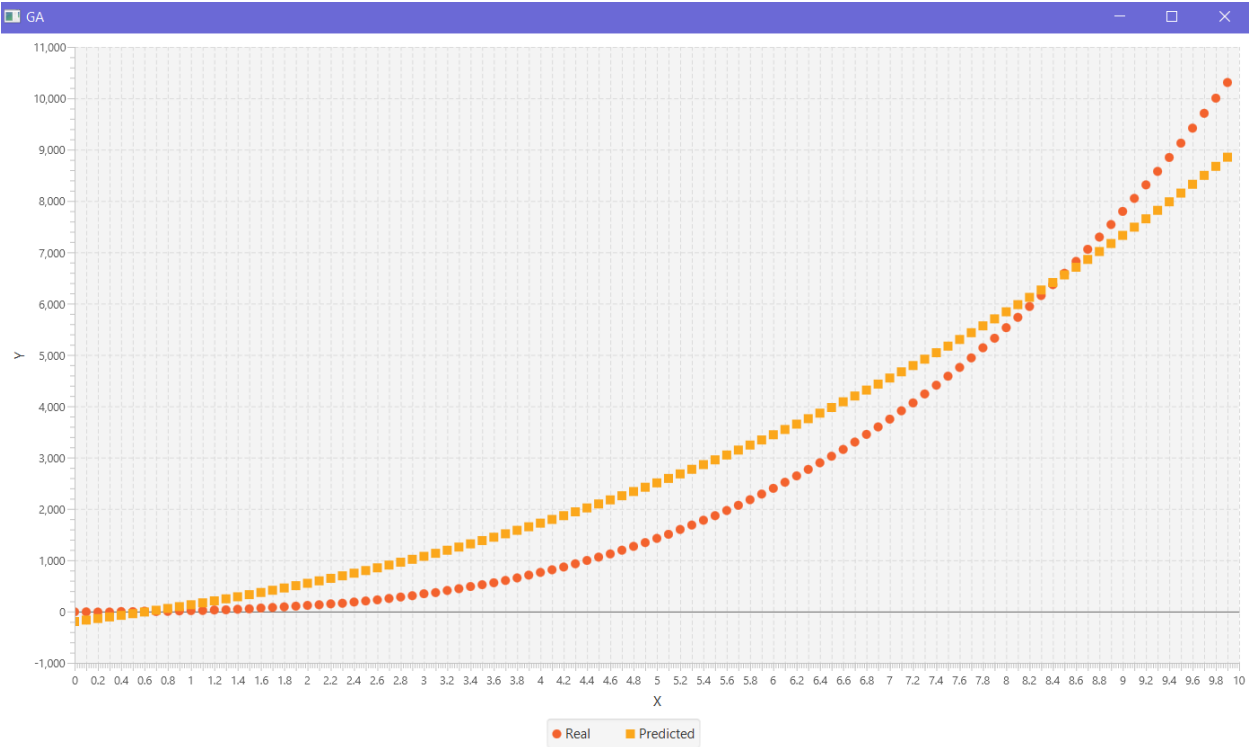
تست (۷) تأثیر Tournament Size

NumberOfGenerations = 2500 , PopulationSize = 50, TournamentSize = 2, MutationRate = 0.02, Variance = 0.01

در این تست بازه‌ی انتخاب تصادفی ضرایب بین -۴۰۰ تا ۴۰۰ است.



شکل ۱۳ – نمودار بهترین، بدترین و میانگین شایستگی تست ۷



شکل ۱۴ – مقایسه‌ی حالت پیشبینی شده با واقعیت تست ۷

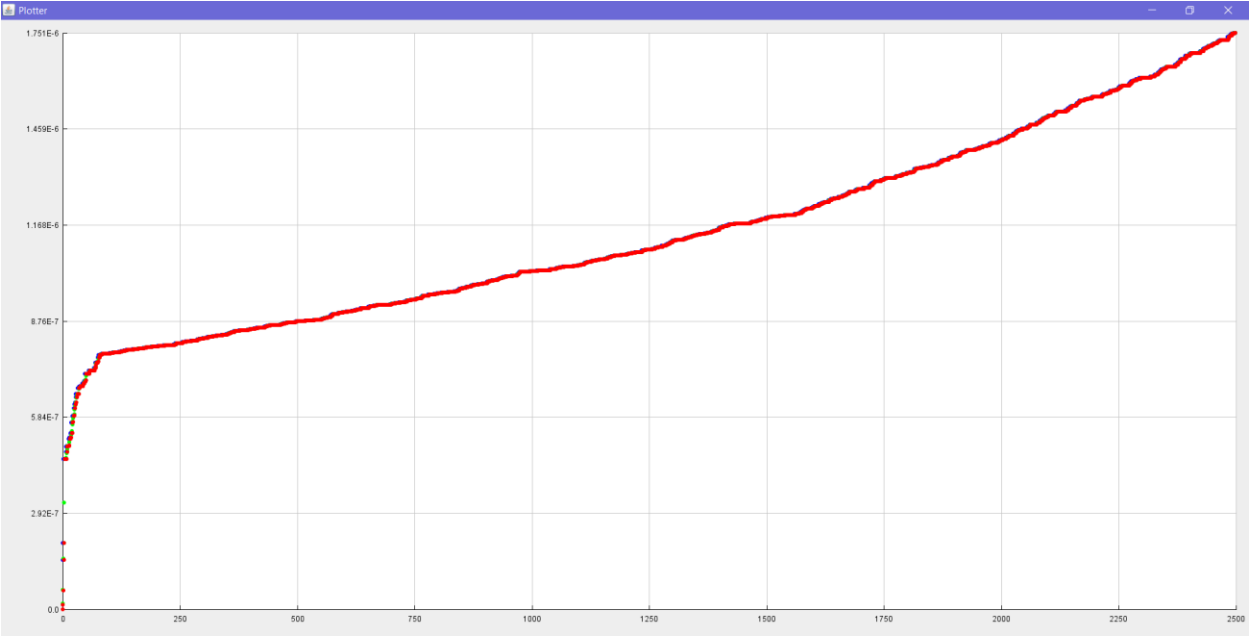
معادله‌ی پیشینی شده:

$$y = 2.575x^3 + 37.847x^2 + 286.670x - 194.469$$

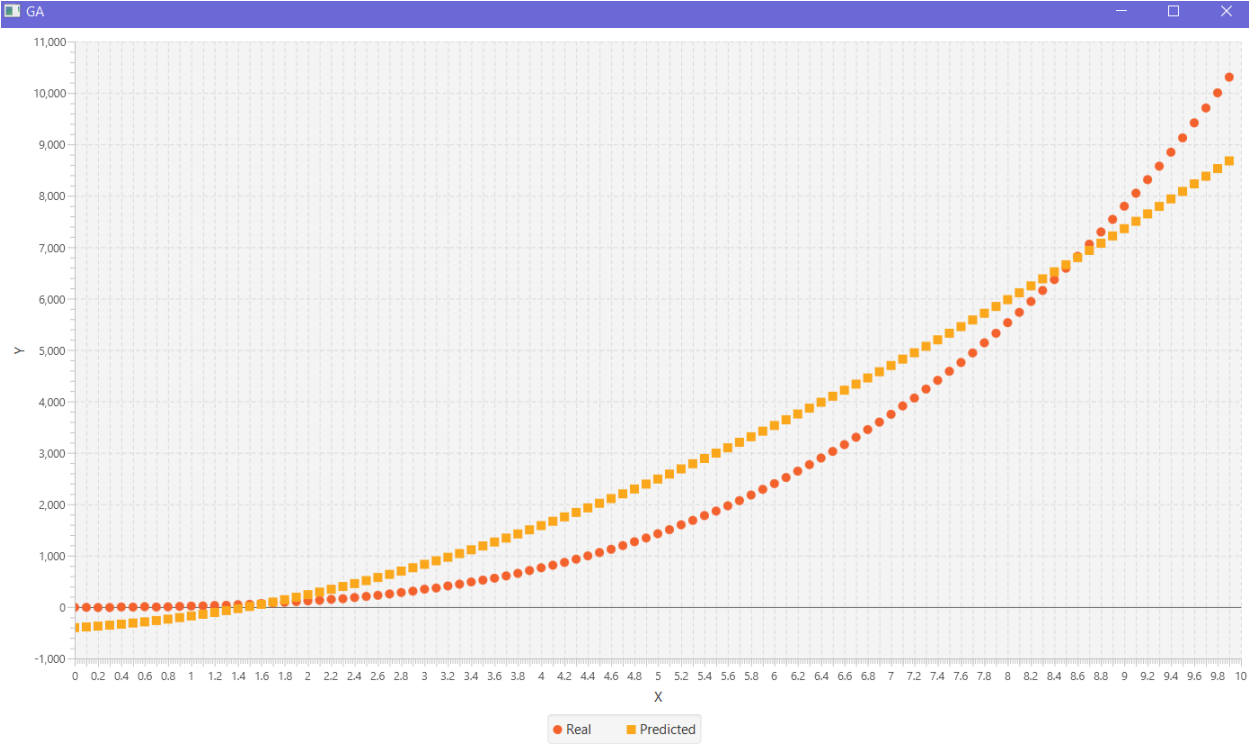
تست (۸) تأثیر *Tournament Size*

*NumberOfGenerations* = 2500 , *PopulationSize* = 50, *TournamentSize* = 5, *MutationRate* = 0.02, *Variance* = 0.01

در این تست بازه‌ی انتخاب تصادفی ضرایب بین -۴۰۰ تا ۴۰۰ است.



شکل ۱۵ – نمودار بهترین، بدترین و میانگین شایستگی تست ۸



شکل ۱۶ – مقایسه‌ی حالت پیشبینی شده با واقعیت تست ۸

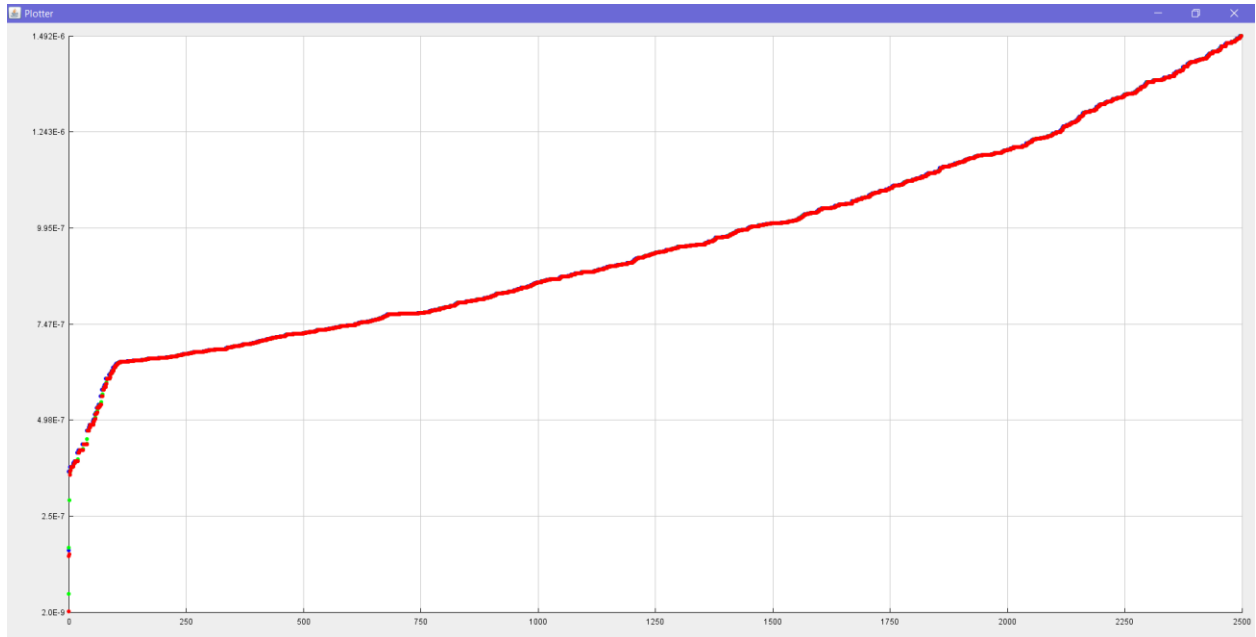
معادله‌ی پیشبینی‌شده:

$$y = -2.095x^3 + 100.458x^2 + 127.968x - 400.648$$

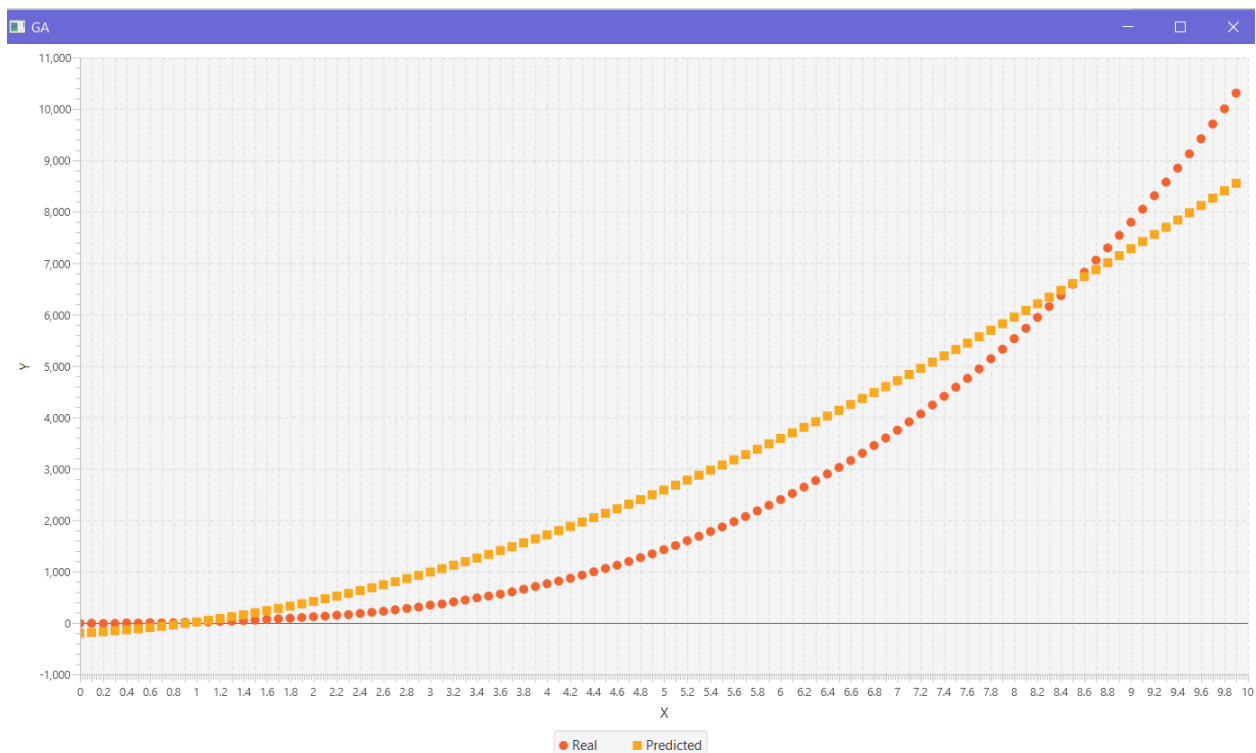
## تست (۹) تأثیر *Tournament Size*

*NumberOfGenerations = 2500 , PopulationSize = 50, TournamentSize = 10, MutationRate = 0.02, Variance = 0.01*

در این تست بازه‌ی انتخاب تصادفی ضرایب بین -۴۰۰ تا ۴۰۰ است.



شکل ۱۷ – نمودار بهترین، بدترین و میانگین شایستگی تست ۹



شکل ۱۸ – مقایسه‌ی حالت پیشبینی شده با واقعیت تست ۹

معادله‌ی پیشبینی‌شده:

$$y = -1.970x^3 + 96.095x^2 + 125.950x - 199.634$$

## نتیجه‌گیری

با توجه به تست‌های (۷)، (۸) و (۹) و شکل‌های ۱۳ و ۱۵ و ۱۷ می‌توان نتیجه گرفت با افزایش Tournament Size جمعیت سریعتر همگرا می‌شوند و این متغیر تأثیر مستقیمی بر جواب ما ندارد.

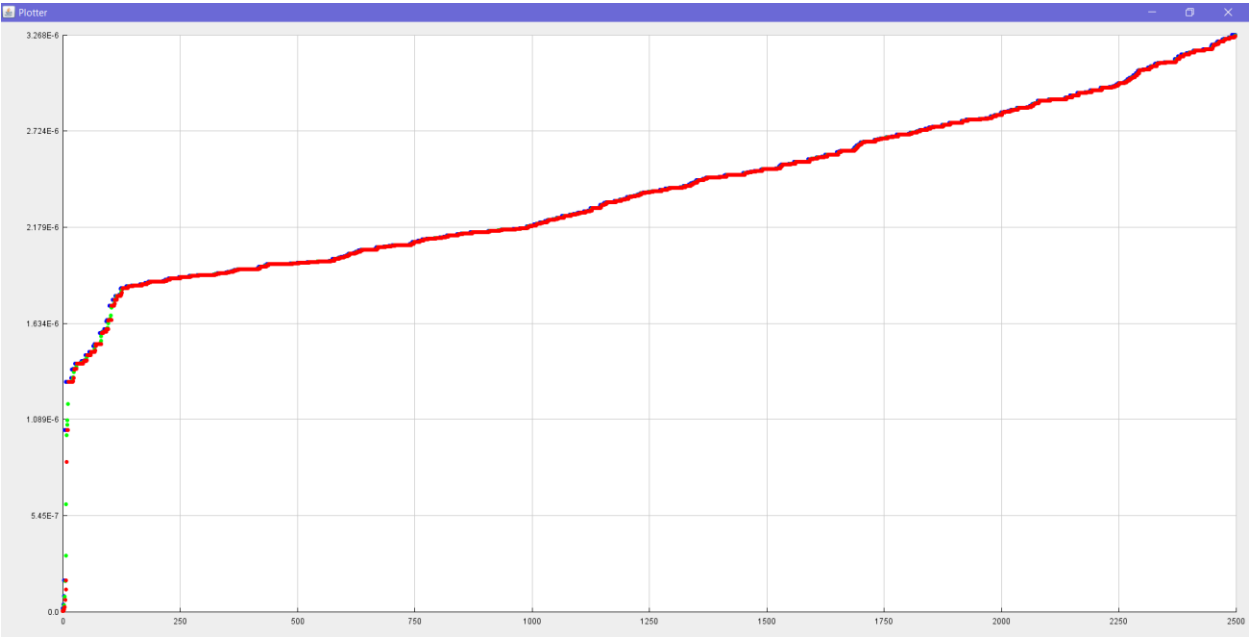
## تأثیر متغیر Mutation Rate

برای تشخیص تأثیر متغیر Mutation Rate بر مسئله، این متغیر را برابر مقادیر ۰,۰۱، ۰,۰۲، ۰,۰۵ و ۰,۱ قرار دهیم و بقیه‌ی متغیرها را ثابت فرض کنیم. برای کاهش تأثیر جمعیت اولیه‌ی تصادفی، برای هر ۴ مقدار از جمعیت اولیه‌ی تصادفی یکسانی استفاده می‌کنیم.

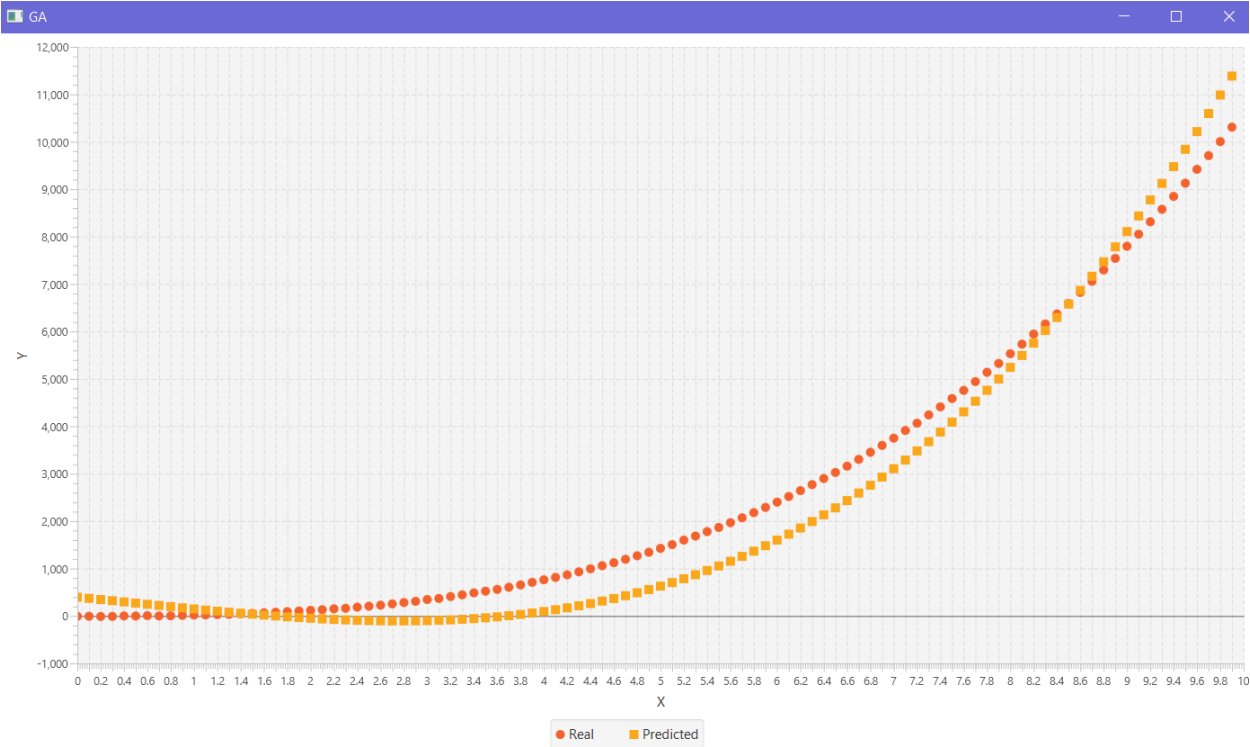
تست (۱۰) تأثیر *Mutation Rate*

*NumberOfGenerations = 2500 , PopulationSize = 50, TournamentSize = 2, MutationRate = 0.01, Variance = 0.01*

در این تست بازه‌ی انتخاب تصادفی ضرایب بین -۴۰۰ تا ۴۰۰ است.



شکل ۱۹ – نمودار بهترین، بدترین و میانگین شایستگی تست ۱۰



شکل ۲۰ – مقایسه‌ی حالت پیشبینی شده با واقعیت تست ۱۰

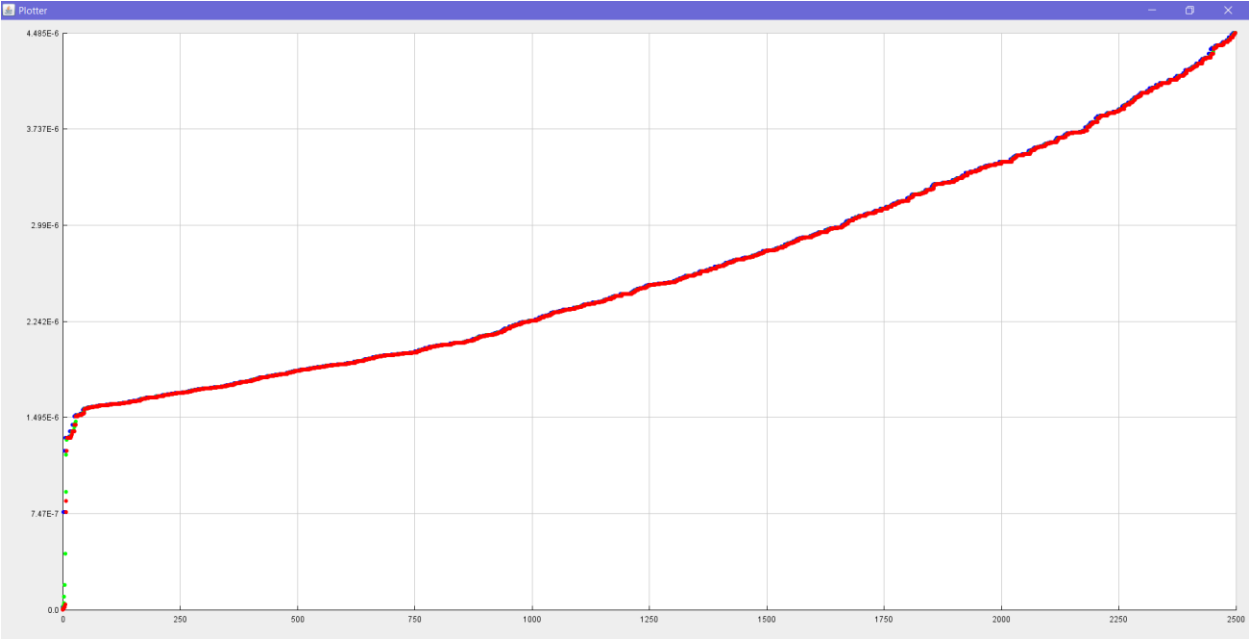
معادله‌ی پیشینی‌شده:

$$y = 16.141x^3 - 23.487x^2 - 239.374x + 395.372$$

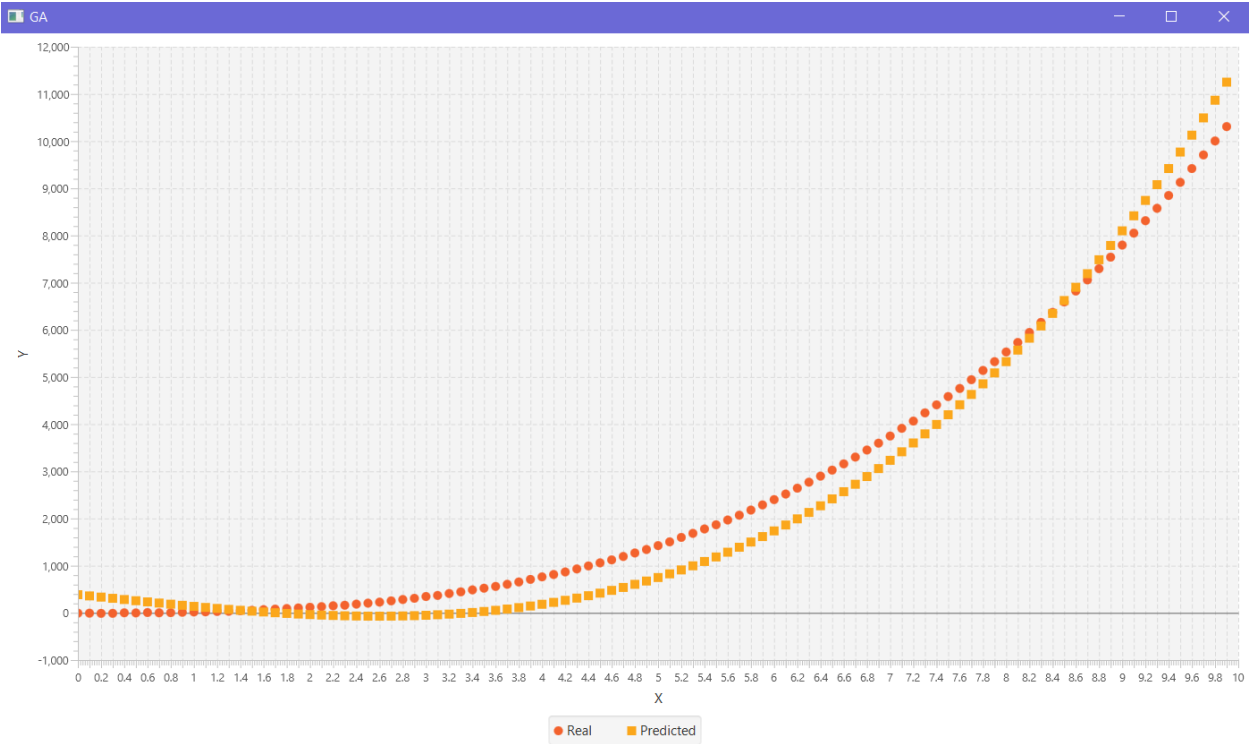
تست (۱۱) تأثیر *Mutation Rate*

*NumberOfGenerations = 2500 , PopulationSize = 50, TournamentSize = 2, MutationRate = 0.02, Variance = 0.01*

در این تست بازه‌ی انتخاب تصادفی ضرایب بین -۴۰۰ تا ۴۰۰ است.



شکل ۲۱ – نمودار بهترین، بدترین و میانگین شایستگی تست ۱۱



شکل ۲۲ – مقایسه‌ی حالت پیشبینی شده با واقعیت تست ۱۱

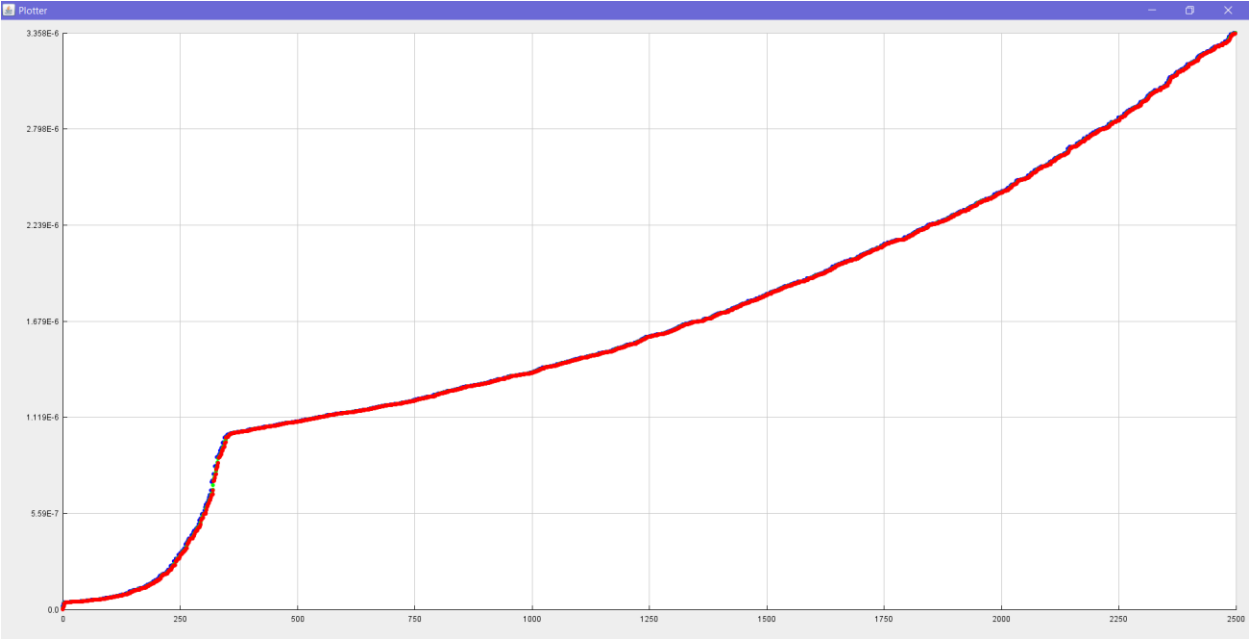
معادله‌ی پیشینی‌شده:

$$y = 14.499x^3 - 6.937x^2 - 254.992x + 385.202$$

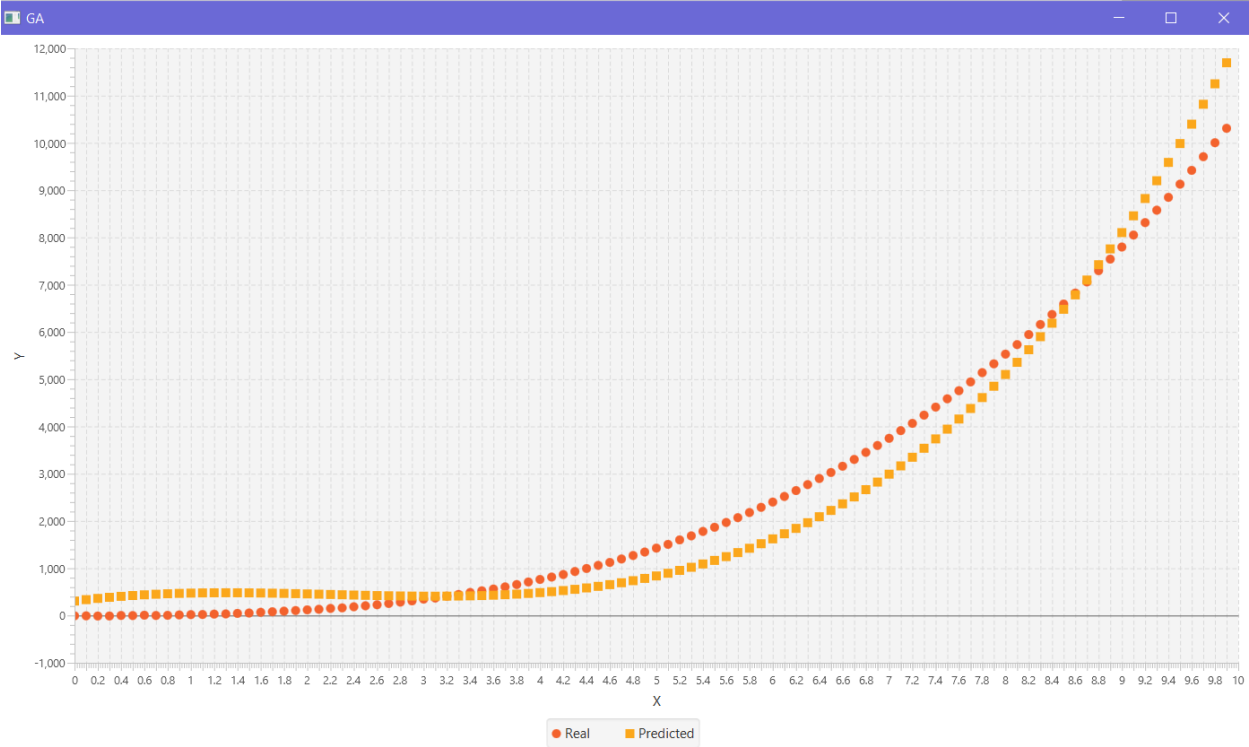
تست (۱۲) تأثیر *Mutation Rate*

*NumberOfGenerations = 2500 , PopulationSize = 50, TournamentSize = 2, MutationRate = 0.05, Variance = 0.01*

در این تست بازه‌ی انتخاب تصادفی ضرایب بین -۴۰۰ تا ۴۰۰ است.



شکل ۲۳ – نمودار بهترین، بدترین و میانگین شایستگی تست ۱۲



شکل ۲۴ – مقایسه‌ی حالت پیشبینی شده با واقعیت تست ۱۲

معادله‌ی پیشبینی‌شده:

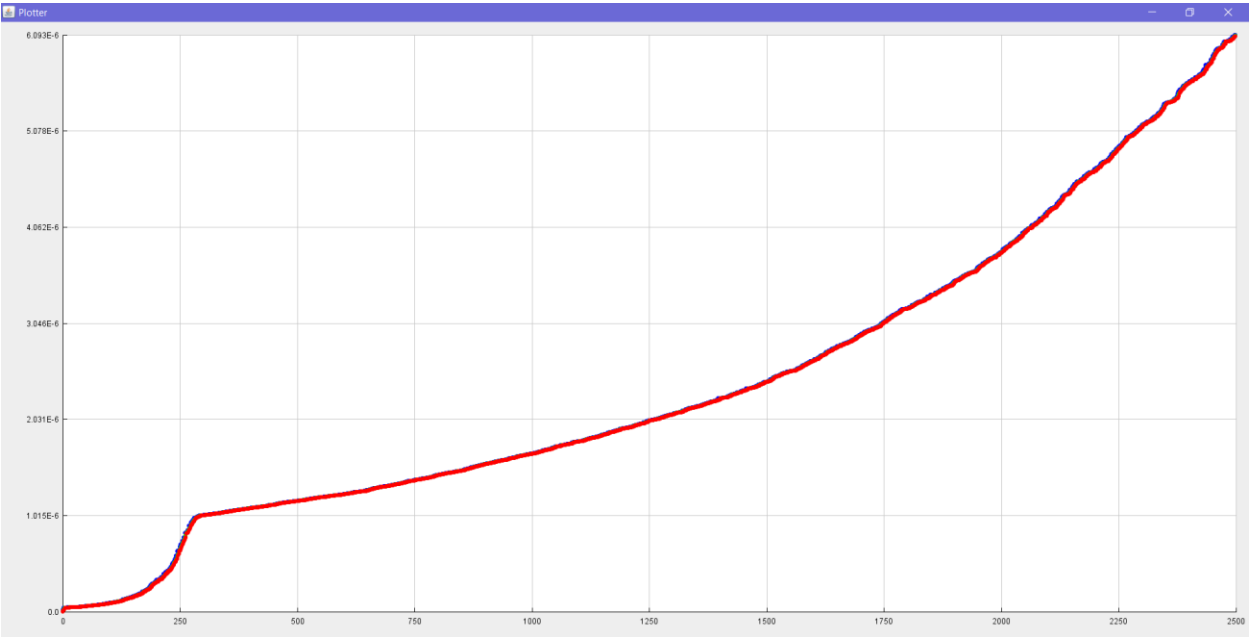
$$y = 25.670x^3 - 169.419x^2 + 311.555x + 306.313$$



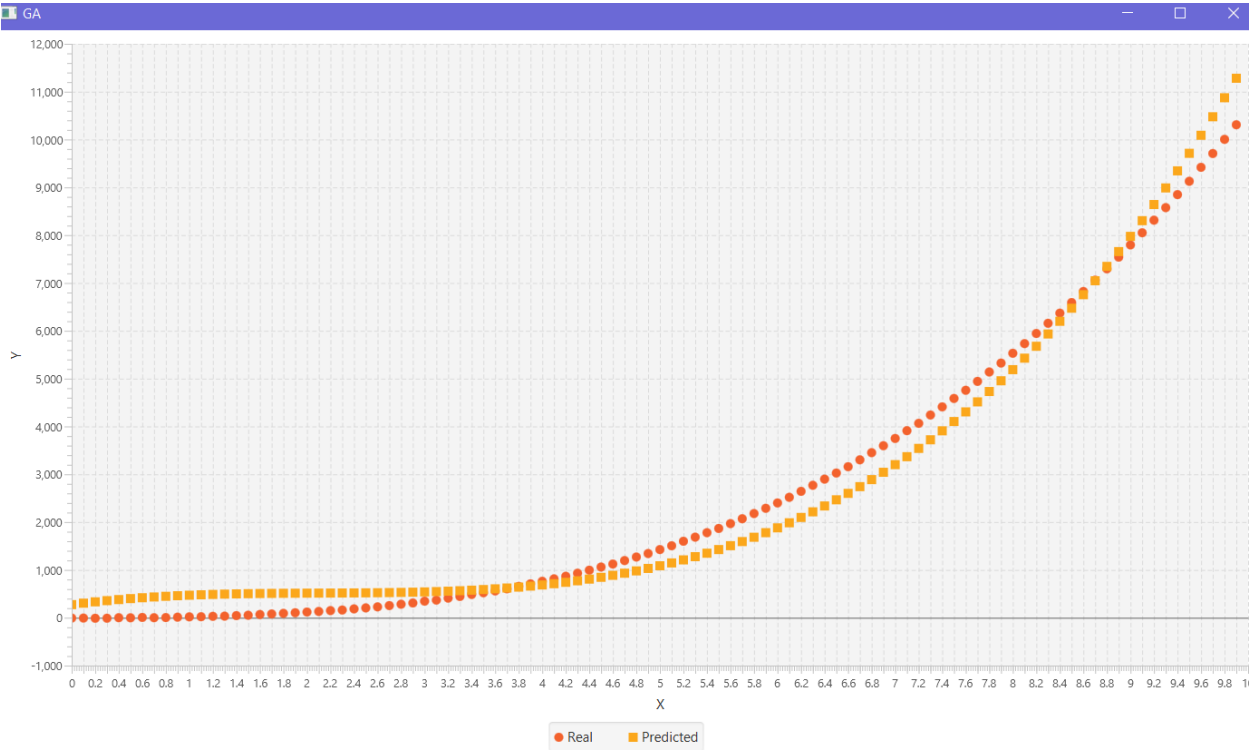
تست (۱۳) تأثیر *Mutation Rate*

*NumberOfGenerations = 2500 , PopulationSize = 50, TournamentSize = 2, MutationRate = 0.1, Variance = 0.01*

در این تست بازه‌ی انتخاب تصادفی ضرایب بین -۴۰۰ تا ۴۰۰ است.



شکل ۲۵ – نمودار بهترین، بدترین و میانگین شایستگی تست ۱۳



شکل ۲۶ – مقایسه‌ی حالت پیشبینی شده با واقعیت تست ۱۳

معادله‌ی پیشینی شده:

$$y = 22.724x^3 - 144.871x^2 + 318.728x + 274.690$$

نتیجه گیری

با توجه به تست های (۱۰)، (۱۱)، (۱۲) و (۱۳) می توان نتیجه گرفت که با افزایش نرخ جهش، جواب ها نیز بهتر می شوند. هر چه نرخ جهش بیشتر باشد، جهش هم بیشتر می شود. در نتیجه امکان اتفاق افتادن جهش مناسب هم بیشتر می شود. همچنین جهش هایی که موجب ضعیف شدن individual ها میشوند در مرحله ی بازگشت از بین می روند.

فیتنس آخرین نسل	نرخ جهش
3.268E-6	0.01
4.485E-6	0.02
3.358E-6	0.05
6.093E-6	0.1

جدول ۳

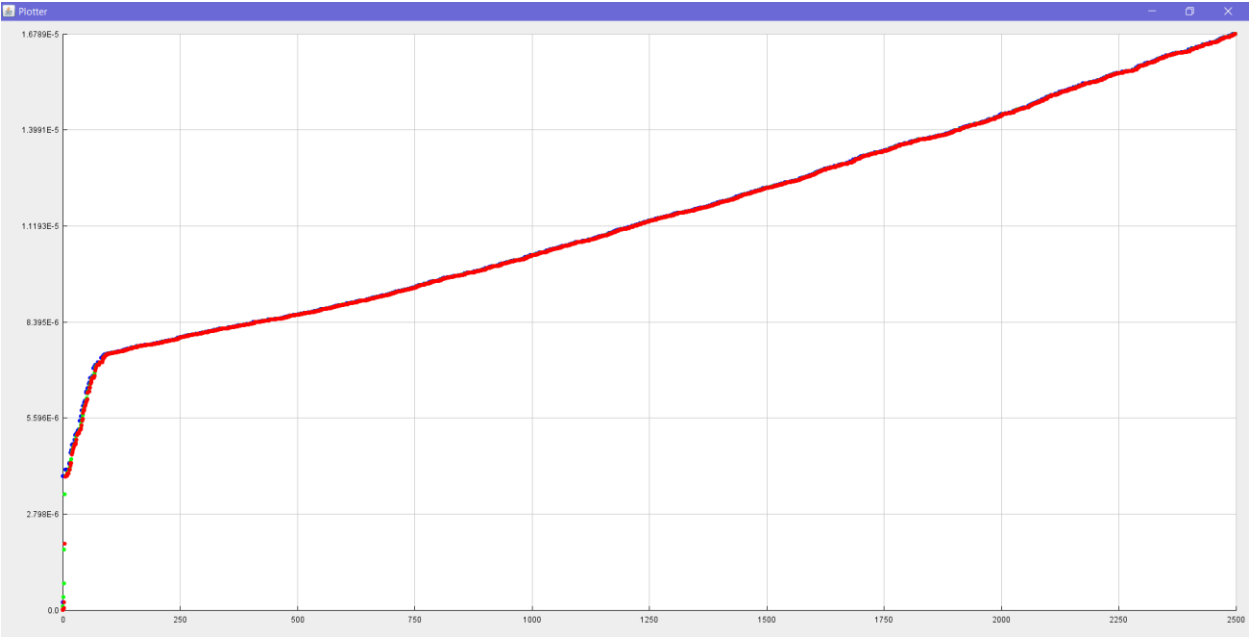
تأثیر متغیر Variance

برای تشخیص تأثیر متغیر Variance بر مسئله، این متغیر را برابر مقادیر ۰,۰۰۱، ۰,۰۱، ۰,۱ و ۱ و ۱۰ قرار می دهیم و بقیه ی متغیرها را ثابت فرض می کنیم. برای کاهش تأثیر جمعیت اولیه ی تصادفی، برای هر ۴ مقدار از جمعیت اولیه ی تصادفی یکسانی استفاده می کنیم.

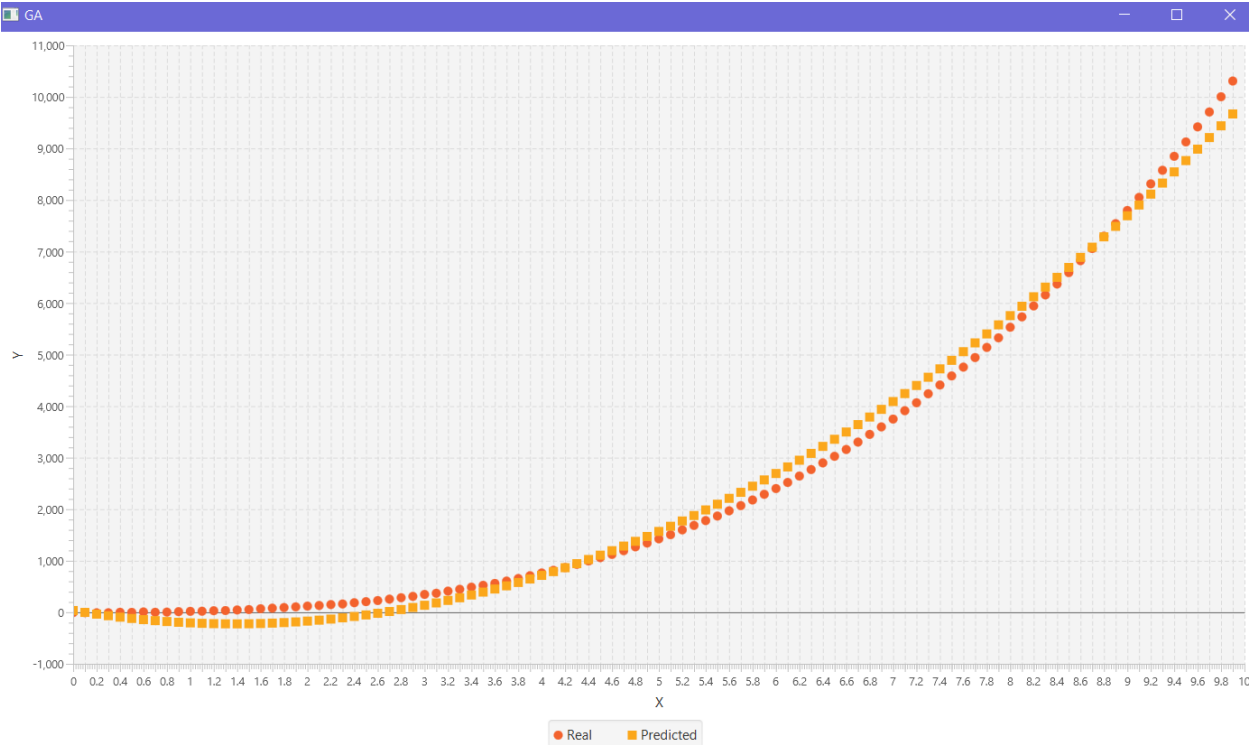
تست (۱۴) تأثیر *Variance*

*NumberOfGenerations = 2500 , PopulationSize = 50, TournamentSize = 2, MutationRate = 0.05, Variance = 0.001*

در این تست بازه‌ی انتخاب تصادفی ضرایب بین -۴۰۰ تا ۴۰۰ است.



شکل ۲۷ – نمودار بهترین، بدترین و میانگین شایستگی تست ۱۴



شکل ۲۸ – مقایسه‌ی حالت پیشبینی شده با واقعیت تست ۱۴

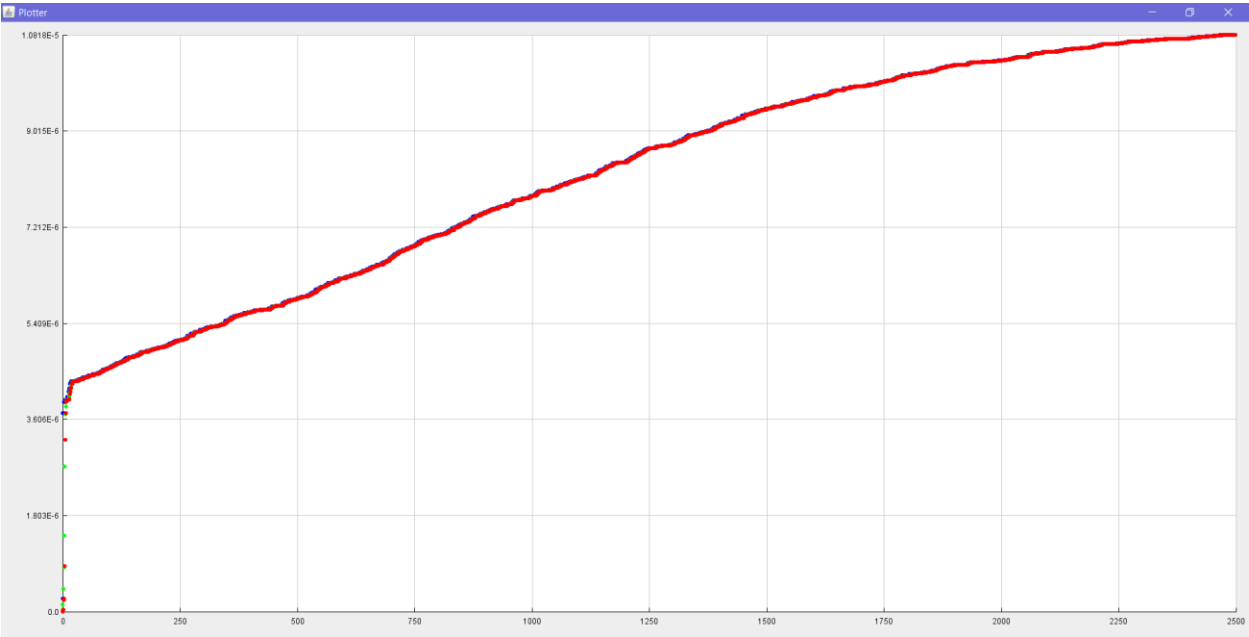
معادله‌ی پیشینی‌شده:

$$y = -0.065x^3 + 136.948x^2 - 376.178x + 34.072$$

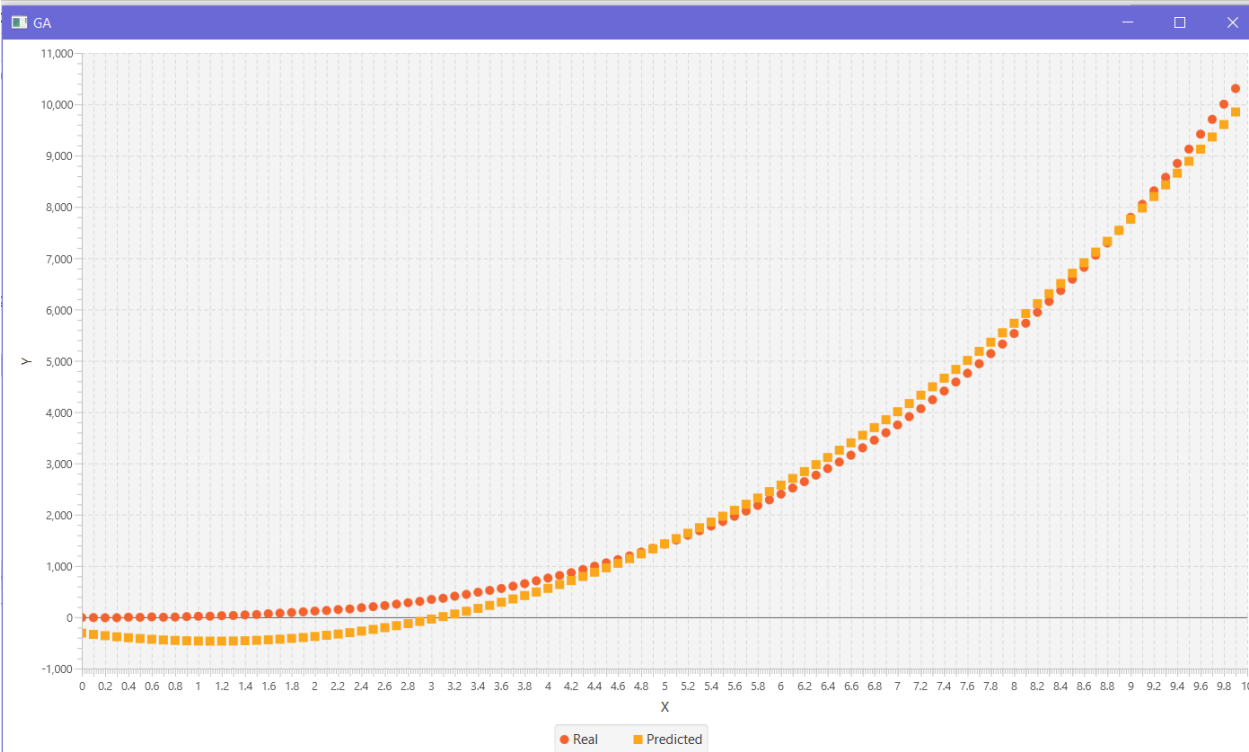
تست (۱۵) تأثیر *Variance*

*NumberOfGenerations = 2500 , PopulationSize = 50, TournamentSize = 2, MutationRate = 0.05, Variance = 0.01*

در این تست بازه‌ی انتخاب تصادفی ضرایب بین -۴۰۰ تا ۴۰۰ است.



شکل ۲۹ – نمودار بهترین، بدترین و میانگین شایستگی تست ۱۵



شکل ۳۰ – مقایسه‌ی حالت پیشبینی شده با واقعیت تست ۱۵

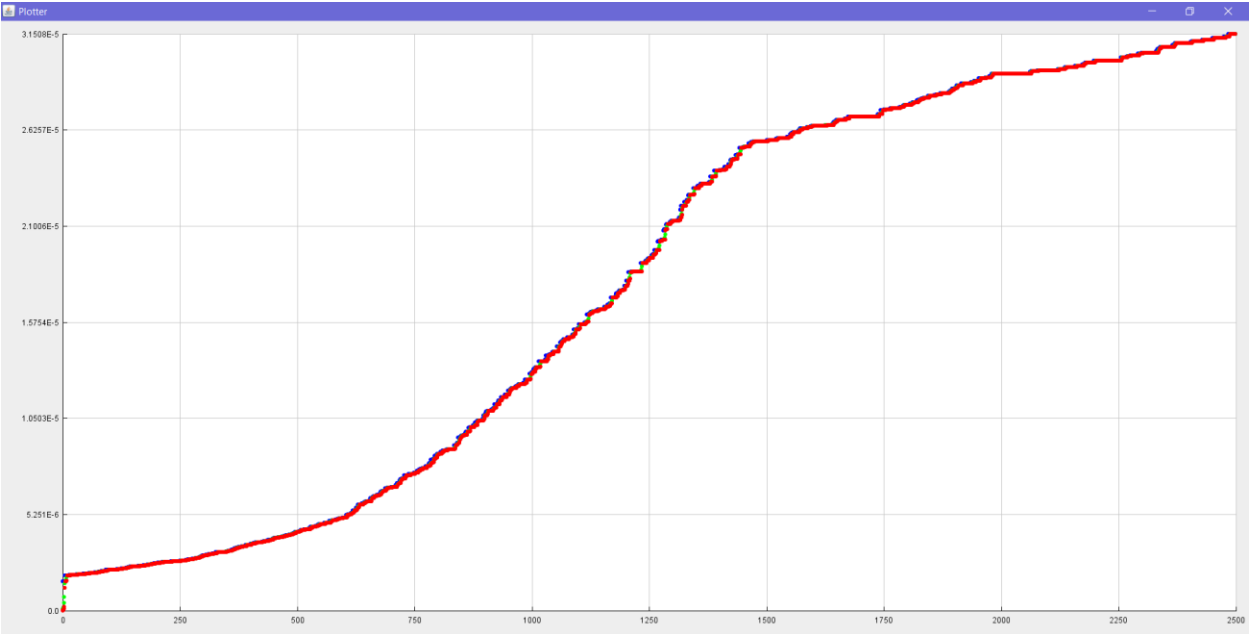
معادله‌ی پیشینی شده:

$$y = 1.434x^3 + 117.017x^2 - 272.878x - 309.059$$

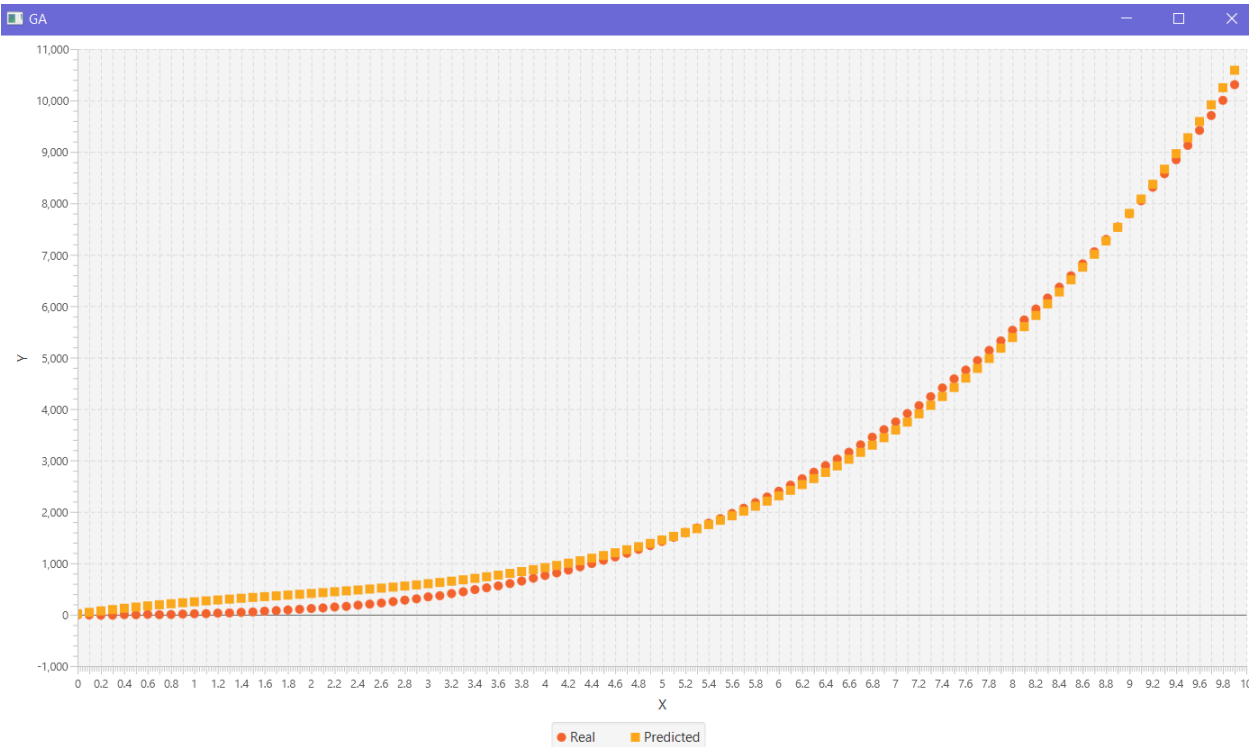
تست (۱۶) تأثیر *Variance*

*NumberOfGenerations* = 2500 , *PopulationSize* = 50, *TournamentSize* = 2, *MutationRate* = 0.05, *Variance* = 0.1

در این تست بازه‌ی انتخاب تصادفی ضرایب بین -۴۰۰ تا ۴۰۰ است.



شکل ۳۱ – نمودار بهترین، بدترین و میانگین شایستگی تست ۱۶



شکل ۳۲ – مقایسه‌ی حالت پیشبینی شده با واقعیت تست ۱۶

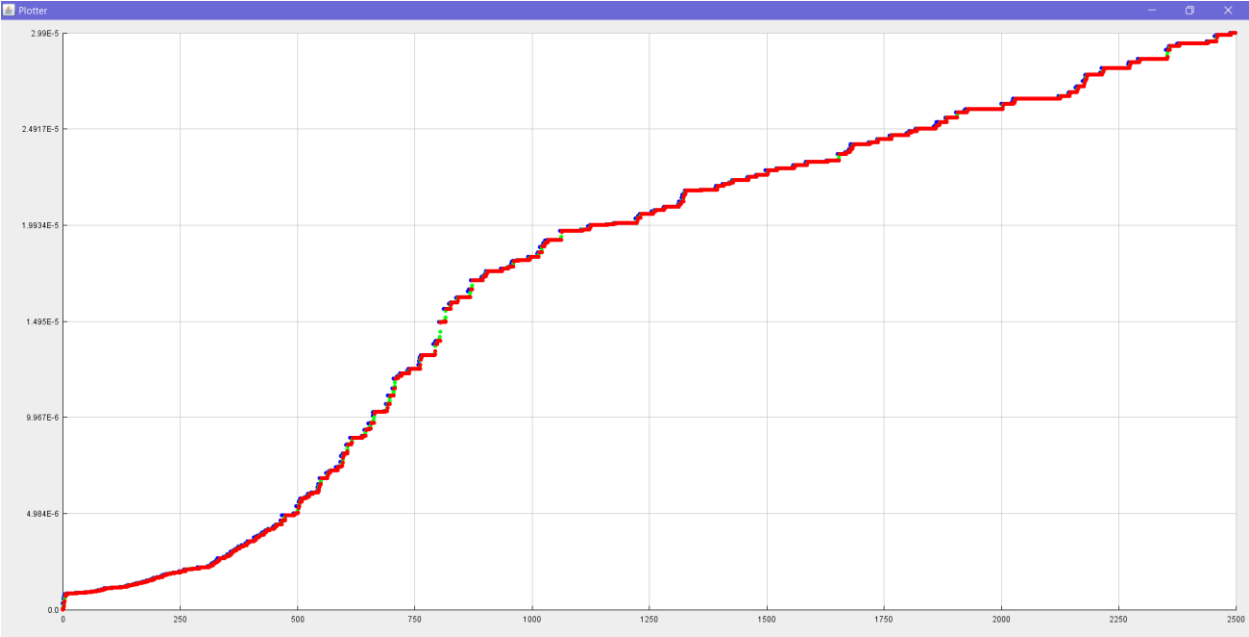
معادله‌ی پیشبینی‌شده:

$$y = 16.414x^3 - 85.290x^2 + 303.335x + 16.634$$

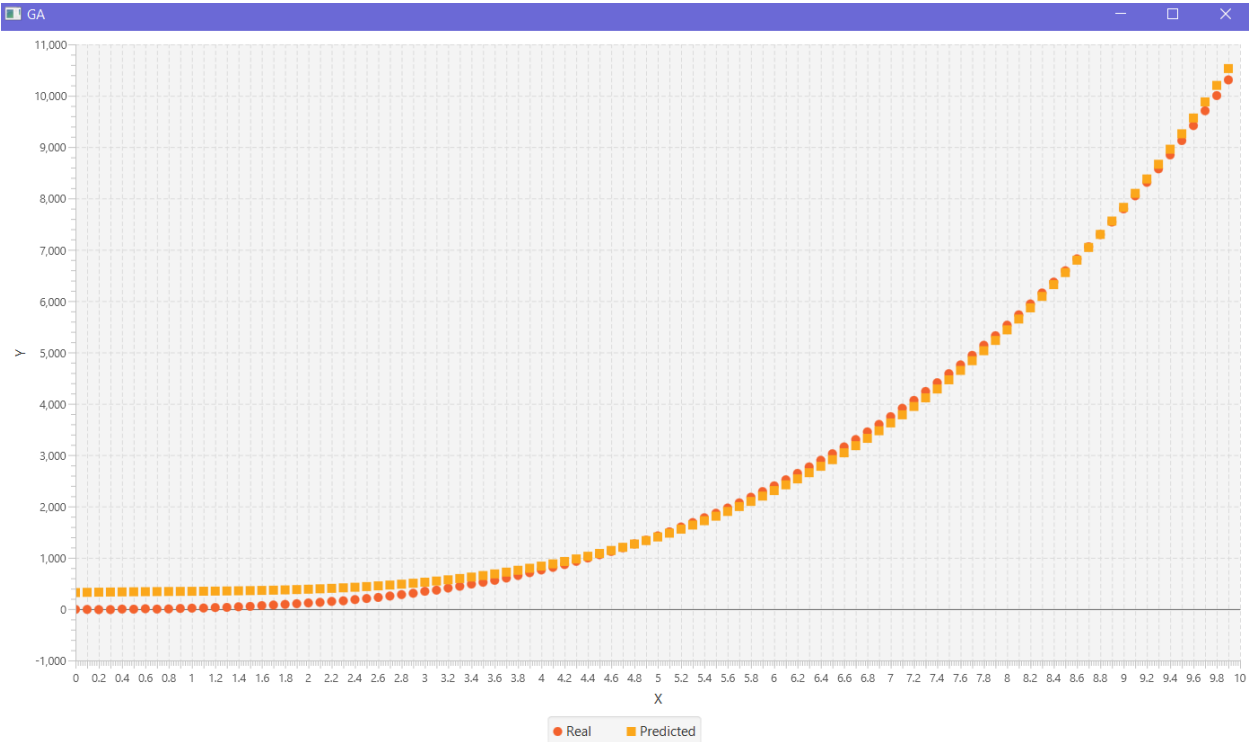
تست (۱۷) تأثیر *Variance*

*NumberOfGenerations = 2500 , PopulationSize = 50, TournamentSize = 2, MutationRate = 0.05, Variance = 1*

در این تست بازه‌ی انتخاب تصادفی ضرایب بین -۴۰۰ تا ۴۰۰ است.



شکل ۳۳ – نمودار بهترین، بدترین و میانگین شایستگی تست ۱۷



شکل ۳۴ – مقایسه‌ی حالت پیشبینی شده با واقعیت تست ۱۷

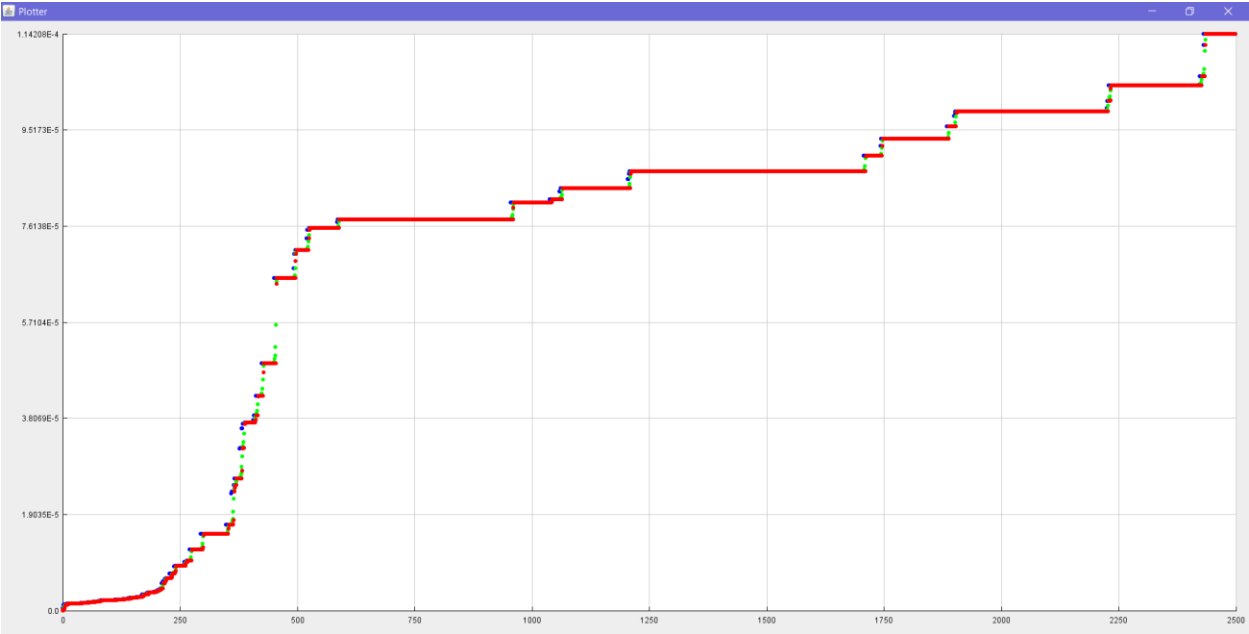
معادله‌ی پیشبینی‌شده:

$$y = 13.264x^3 - 31.448x^2 + 42.210x + 322.710$$

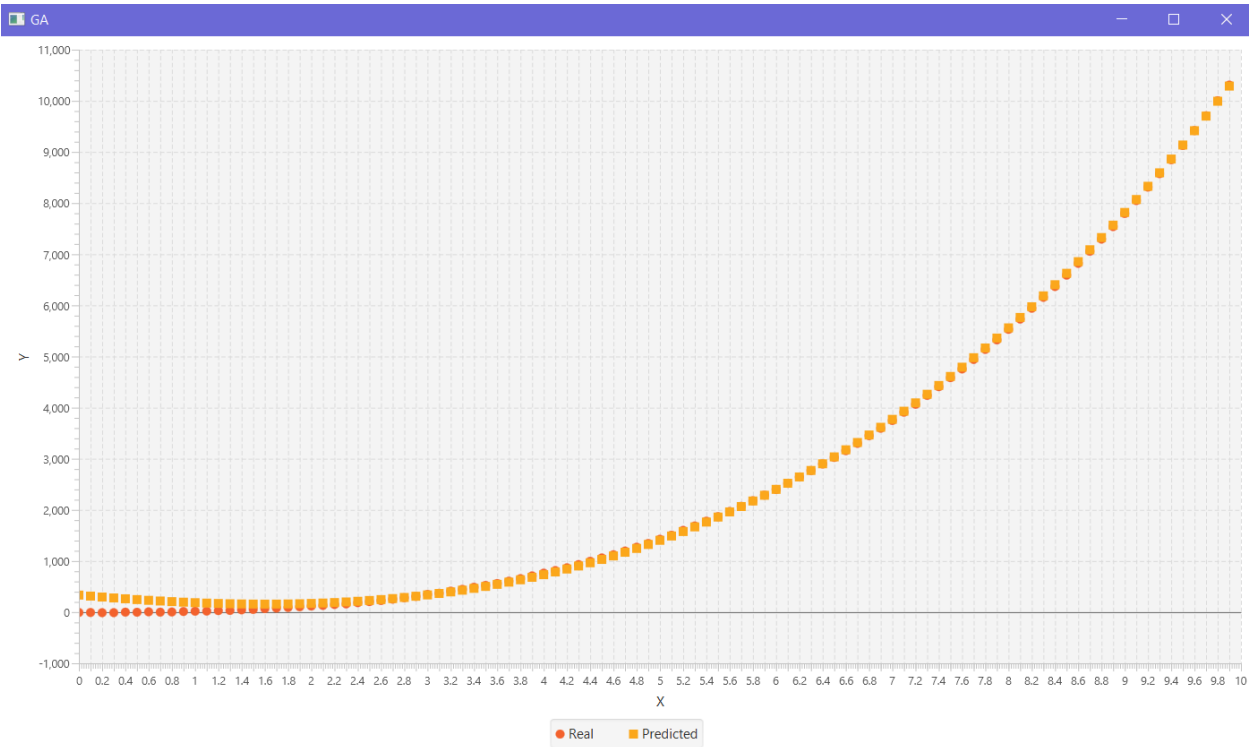
تست (۱۸) تأثیر *Variance*

*NumberOfGenerations = 2500 , PopulationSize = 50, TournamentSize = 2, MutationRate = 0.05, Variance = 10*

در این تست بازه‌ی انتخاب تصادفی ضرایب بین -۴۰۰ تا ۴۰۰ است.



شکل ۳۵ – نمودار بهترین، بدترین و میانگین شایستگی تست ۱۸



شکل ۳۶ – مقایسه‌ی حالت پیشبینی شده با واقعیت تست ۱۸

معادله‌ی پیشینی شده:

$$y = 7.963x^3 + 42.826x^2 - 199.002x + 334.104$$

نتیجه گیری

با توجه به نتایج تست های (۱۴)، (۱۵)، (۱۶)، (۱۷) و (۱۸) و جدول ۴ می توان نتیجه گرفت که با افزایش Variance نتایج بهتری می توان کسب کرد.

Variance	فیتنس آخرین نسل
0.001	1.6789E-5
0.01	1.0818E-5
0.1	3.1508E-5
1	2.99E-5
10	1.14208E-4

جدول ۴

همچنین با توجه به شکل های ۲۷، ۲۹، ۳۱، ۳۳ و ۳۵ می توان فهمید که با افزایش Variance مقدار نويز گاوسی زیاد می شود. در نتیجه جهش ها با مقدار بیشتری انجام می شوند. به طوری که در شکل ۳۵ و در شرایطی که Variance ۱۰ لست نمودار حالتی پله گونه به خود گرفته است.