

پیاده‌سازی Regression با استفاده از الگوریتم ژنتیک

در این تمرین شما باید با استفاده از الگوریتم genetic ضرایب یک تابع درجه دو به فرم $ax^2 + bx + c$ را بیابید. در فایل‌های `x_train.csv` و `y_train.csv` مختصات (x, y) متناظر با تابع موردنظر به ازای 1000 نقطه داده شده است. دقت کنید که نقاط داده شده دارای offset نسبت به تابع هستند در نتیجه الگوریتم شما باید ضرایب را به گونه‌ای پیدا کند که تابع حاصل بهترین نماینده نقاط داده شده باشد. در نهایت باید پیشرفت نسل به نسل در الگوریتم شما دیده شود. دقت کنید که ممکن است همه به یک جواب نرسند اما جواب‌ها باید به هم نزدیک باشند.

داده‌های `x_train` در بازه‌ی $(-500, 500)$ می‌باشد همچنین ضرایب تابعی که داده‌ها طبق آن تولید شده بین $(-50, 50)$ و offset داده‌ها بین $(-20, 20)$ می‌باشد.

برای این تمرین از هر زبان برنامه‌نویسی دلخواه می‌توانید استفاده کنید.

راهنمایی :

سعی کنید برای راحتی تحویل و دیباگ پس از اتمام هر نسل `average fitness/cost` و `best individual fitness` و `best individual values` را چاپ کنید.

اگر از زبان python استفاده می‌کنید می‌توانید با استفاده از تابع `np.genfromtxt('x_train.csv', delimiter=',')` داده را به صورت `numpy array` بخوانید. هم چنین برای درک بهتر داده‌ها می‌توانید با استفاده از `matplotlib.pyplot.scatter` ابتدا نمودار آن‌ها را بکشید.

سعی کنید ضرایب را طوری انتخاب کنید که فاصله تابع حاصل به ازای هر داده در `x_train` نسبت به `y` داده شده در `y_train` کمترین مقدار شود.

امتیازی :

در زبان python کتابخانه‌های متعددی الگوریتم genetic را به روش‌های مختلف پیاده‌سازی کرده‌اند. به عنوان مثال کتابخانه gplearn با استفاده از مفهوم syntax tree تلاش می‌کند تا یک رابطه‌ی ریاضی با استفاده از function set تعیین شده میان نقاط برقرار کند. با استفاده از این کتابخانه و یا کتابخانه‌های مشابه سعی کنید مساله فوق را حل کنید.

دقت کنید که هدف این قسمت آشنایی شما با ابزارها می باشد در نتیجه نیازی نیست که حتما بهترین جواب را پیدا کنید اما باید بتوانید چگونگی عملکرد کتابخانه و پارامترهای اصلی آن را توضیح بدهید.

تکرات:

۱- توجه داشته باشید که برای گرفتن نمره امتیازی باید ابتدا بدون استفاده از کتابخانه الگوریتم را پیاده‌سازی کنید.

۲- در صورت داشتن هرگونه ابهام یا سوال، آن را از راه **درس‌افزار** مطرح کنید تا بقیه دانشجویها هم از مشکلات احتمالی باخبر و از سوالات شما استفاده کنند.

موفق باشید