

بسمه تعالی

گزارش پروژه اول هوش مصنوعی - سوال دوم - رگرسیون

محمدباقر عابدی سقا - ۹۵۳۱۹۰۴

روند مدل سازی و اجرای کد:

ابتدا در تابع main() تمام ورودی های مسئله اعم از تعداد نسل هایی که ساخته می شود، تعداد جمعیت، سایز تورنومنت و mutation rate و سیگمای تابع نرمال داده می شود.

برای مدل کردن هر چند جمله ای یک کلاس Ind داریم که دارای آرایه ای ۴ تایی از ضرایب چند جمله ای و همچنین فیلد شایستگی fitness آن است. همچنین تابع تعیین شایستگی در این کلاس تعریف شده که مطابق فرمول داده شده در صورت پروژه، با گرفتن داده های ورودی و مقایسه هر individual با هر کدام از ۱۰۰ داده ی ورودی، شایستگی آن را تعیین می کند.

مرحله اول

در فایل app.py و تابع main()، ابتدا تمام ورودی ها در xs, ys و در مرحله بعد از آن به تعداد جمعیت، individual با ضرایب رندوم ساخته می شود.

مرحله دوم

در مرحله دوم شایستگی همه جمعیت ساخته شده با توجه به داده های ورودی سنجیده و ثبت می شود.

مرحله سوم تا ششم

از مرحله سوم به بعد درون تابع gen() به دفعات نسل انجام می شود. در این گزارش هر حالت ۱۰۰۰۰ بار آزمایش شده است.

در تابع gen() هر بار ابتدا

۱- پنجاه جفت والد به صورت رندوم از جمعیت انتخاب می شود

۲- پنجاه فرزند جدید از این ۱۰۰ والد توسط تابع crossover() ساخته می شود

۳- ضرایب چند جمله ای فرزندان جدید توسط تابع mutation() و با احتمال mutation rate به میزان خروجی تابع نرمال گوسی با sigma داده شده افزوده می شود.

۴- در نهایت fitness این ۵۰ فرزند جدید و جهش یافته محاسبه می شود

۵- با مرتب سازی فرزندان جدید همراه با جمعیت اولیه بر حسب fitness آنها، پنجاه تای اول که بزرگترین fitness را دارا هستند به عنوان جمعیت اولیه برگردانده شده و جایگزین می‌شوند. این رشته اعمال در این آزمایش ۵۰۰۰ بار تکرار شده و هر بار بهترین، پایین‌ترین و میانگین شایستگی جمعیت هر نسل ذخیره می‌شود. در پایان حلقه، نمودار چند جمله‌ای با بهترین ضرایب شایستگی همراه با نقاط اولیه داده شده در یک نمودار چاپ می‌شود تا دقت رگرسیون انجام شده قابل مشاهده باشد. (نمودار نتیجه، سبز رنگ و نقاط اولیه، دایره های آبی هستند.) سپس بهترین، بدترین و میانگین ضرایب شایستگی هر نسل طی زمان در نمودار دوم آورده می‌شود. (بهترین ضریب، مثلث سبز - میانگین ضریب، خط آبی - پایین‌ترین ضریب، مربع قرمز دارد.)

تاثیر ضرایب:

جمعیت اولیه: هرچه جمعیت اولیه بیشتر باشد، حالت های بیشتری مورد بررسی قرار گرفته و به تبع، نتیجه شباهت بیشتری به ورودی خواهد داشت.

تعداد نسل: با افزایش نسل، جمعیت ویرایش بیشتری گرفته و شبیه تر به خروجی مطلوب خواهد شد. به طور مثال در ۱۰ نسل، شباهت کمتری نسبت به ۱۰۰۰ نسل وجود دارد.

اندازه تورنومنت: افزایش اندازه تورنومنت باعث می‌شود تا فرزندان، صفات بیشتری از والدین خود به ارث برده و سریع تر به سمت نتیجه مطلوب جهش کنند. احتمال جهش (mutation rate): این کمیت در واقع احتمال جهش هر نسل را تعیین می‌کند. اگر برابر یک باشد، تمام نسل ها تغییر خواهند کرد و عملکردی مشابه random walk خواهیم داشت. هر چه بیشتر باشد تعداد جهش هر نسل نیز بالاتر می‌رود. لذا نمودار ضرایب شایستگی پیوسته تر خواهد بود. طبق نتایج آزمایش، کمترین مقدار داده شده در دادن نتیجه مطلوب، موثر تر خواهد بود.

سیگما: این مقدار باعث تعیین می‌کند که اندازه ی جهش در هر نسل چقدر باشد. لذا افزایش آن موجب افزایش تغییرات هر نسل نسبت به نسل قبلی می‌شود. در نمودار ضرایب شایستگی مشاهده می‌کنیم که فاصله ی بین پله های هر نسل رابطه مستقیم با سیگما دارد. لذا می‌توان نتیجه گرفت که کم بودن این مقدار، دیرتر تغییر می‌کند تا به پاسخ برسد و از طرفی مقدار بیش از اندازه بالای آن تغییرات زیادی ایجاد کرده و باز هم دقت پاسخ در تعداد نسل ثابت را پایین می‌آورد.

نمودارهای مقایسه پاسخ نهایی با هدف - بهترین، بدترین و میانگین ضرایب شایستگی هر نسل - خروجی ضرایب شایستگی هر آزمایش در ادامه آورده می‌شود.

MutRate = 0.1	MutRate = 0.05	MutRate = 0.02	MutRate = 0.01	fitness
0.39	0.52	0.47	0.50	Sigma = 0.001
0.87	0.87	0.84	0.85	Sigma = 0.01
0.91	0.89	0.84	0.90	Sigma = 0.1
0.03	0.8	0.85	0.91	Sigma = 1
0.03	0.07	0.08	0.06	Sigma = 10



