



به نام خدا

## پروژه اول درس مقدمه ای بر هوش محاسباتی و زیستی

### قوانین:

- توضیحات مربوط به هر بخش کد باید به صورت گزارش، همراه با کد نوشته شده ارسال گردد. کیفیت گزارش شما در فرآیند تصحیح از اهمیت ویژه ای برخوردار است، بنابراین تمامی نکات و فرضیات خود را در گزارش ذکر کنید.
- \* برای ارائه کد بدون توضیحات و یا توضیحات ناقص نمره ای در نظر گرفته نمی شود.
- در صورت مشاهده تقلب امتیاز تمامی افراد شرکت کننده در آن، 100- لحاظ می شود.
- می توانید در قالب گروه های دو نفره پروژه را انجام دهید.
- زبان های برنامه نویسی مجاز Python و MATLAB می باشند.
- کدها حتما باید در قالب نوتبوک با پسوند ipynb. و یا فایل پایتون با پسوند py. و یا فایل متلب با پسوند m. تهیه شوند.
- لطفا گزارش، کد ها و سایر ضمایم را در یک پوشه با نام زیر قرار داده و آن را فشرده سازید، سپس در سامانه بارگذاری نمایید.

## سوال اول:

### مقدمه

در این قسمت، با روش‌هایی که برگرفته از طبیعت و انتخاب طبیعی هستند، آشنا می‌شویم. در این روش‌ها که به طور کلی الگوریتم ژنتیک نامیده می‌شوند، ایده‌هایی برای مدل‌سازی جفت‌گیری، جهش و انتخاب طبیعی به کار گرفته می‌شود. در این گونه الگوریتم‌ها، ممکن است با انتخاب معیارهای ساده‌ی انتخاب طبیعی، نتایج مطلوب به دست نیاید و باید معیاری در نظر بگیریم که علاوه بر عملکرد فردی، به گوناگونی جمعیت نیز اهمیت دهد.

الگوریتم ژنتیک عموماً در مسئله‌هایی با فضای حالت بزرگ کاربرد دارند؛ این الگوریتم‌ها نمونه گرفتن از جمعیت و ترکیب و تغییر افراد و ارزیابی آن‌ها را انجام می‌دهند و سعی می‌کنند که نسل به نسل جواب‌ها را بهبود دهند تا به جواب مورد نظر برسند.

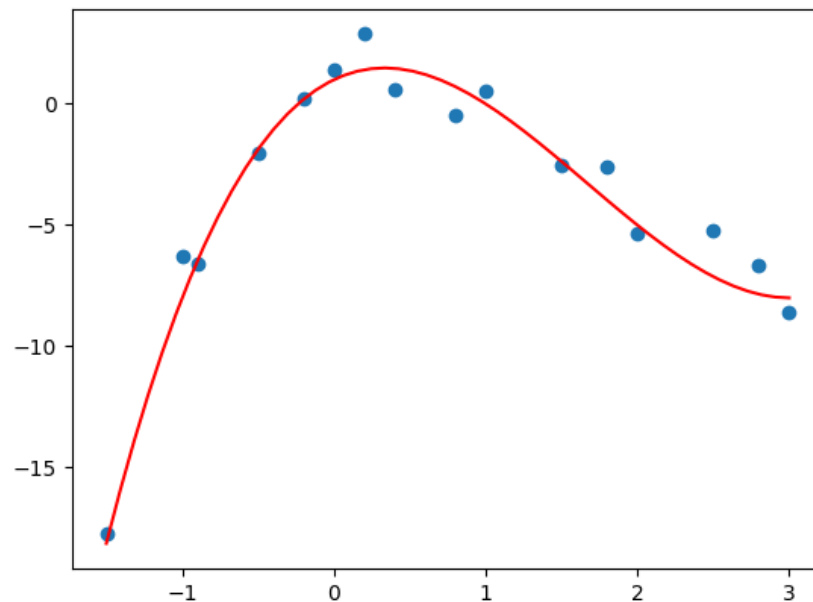
در این پروژه قصد داریم با استفاده از الگوریتم‌های ژنتیک، یکی از مسائل مربوط به بهینه‌سازی را پیاده‌سازی کنیم. هدف یک مسئله بهینه‌سازی، یافتن بهترین راه‌حل از مجموعه بزرگی از راه‌حل‌های ممکن است، جایی که "بهترین" راه‌حل با مجموعه‌ای از معیارها یا اهداف تعریف می‌شود. حل مسائل بهینه‌سازی می‌تواند به مزایای قابل توجهی مانند افزایش کارایی، کاهش هزینه‌ها، بهبود عملکرد و موارد دیگر منجر شود.

## توضیح مسئله اول

در این قسمت شما به مسئله curve fitting می‌پردازید. در این مسئله شما باید با توجه به نقاطی که به شما داده شده است، ضرایب مربوط به یک عبارت چند جمله‌ای را تعیین کنید. هدف این است که بر اساس نقاط داده شده، رابطه بین متغیر مستقل ( $x$ ) و متغیر وابسته ( $y$ ) تعیین شود. به نقاط داده شده زیر توجه کنید:

$x\_data = [0, 0.2, 0.4, 0.8, 1, 1.5, 1.8, 2, 2.5, 2.8, 3, -1, -0.2, -0.5, -0.9, -1.5]$

$y\_data = [1.431, 2.876, 0.564, -0.451, 0.504, -2.537, -2.607, -5.373, -5.222, -6.66, -8.585, -6.251, 0.24, -2.013, -6.621, -17.692]$



ورودی‌های مسئله شامل درجه چندجمله‌ای مورد نظر، محدوده ضرایب و تعدادی نقطه (نقاط بالا) نزدیک آن چند جمله‌ای خواهد بود. دقت کنید که تعداد نقاطی که به شما داده می‌شود، لزوماً به اندازه‌ای نیست که چندجمله‌ای را به صورت یکتا مشخص کند و ممکن است از تعداد نقاط لازم برای این کار کمتر باشد. همچنین، توجه کنید که ضرایب مطلوب مسئله، همواره اعداد صحیح هستند و حداکثر درجه چندجمله‌ای از درجه 6 می‌باشد.

## توضیح مسئله دوم

مسئله فروشنده دوره‌گرد (TSP) یکی از مسائل مشهور در حوزه بهینه‌سازی مسائل است. در این مسأله، یک فروشنده باید یک مجموعه از شهرها را دور زده و به هر یک از آن شهرها یک بار سفر کند و در نهایت به شهر اول بازگردد. هدف او یافتن مسیر کوتاه‌ترین ممکن است که تمام شهرها را شامل شود و هر شهر نیز تنها یک بار بازدید شود.

شهرها در صفحه مختصات، به صورت زیر واقع شده‌اند:

{[60, 200], [180, 200], [80, 180], [140, 180], [20, 160], [100, 160], [200, 160], [140, 140], [40, 120], [100, 120], [180, 100], [60, 80], [120, 80], [180, 60], [20, 40], [100, 40], [200, 40], [20, 20], [60, 20], [160, 20]}

## پیاده‌سازی مسائل

### بخش یک: مشخص کردن مفاهیم اولیه

در الگوریتم‌های ژنتیک ابتدا باید یک تعریف برای ژن ارائه دهید و سپس با استفاده از آن، یک کروموزوم بسازید. هر کروموزوم مجموعه‌ای از ژن‌ها است و این مجموعه یا همان کروموزوم، یک راه پیشنهادی برای حل مسئله مورد نظر می‌باشد.

توجه داشته باشید که در الگوریتم‌های ژنتیک باید اکثر کارها را با استفاده از تصادفی کردن وقایع انجام دهید، چرا که اگر فضای حالت بزرگ باشد، پیدا کردن شرطی که همه محدودیت‌ها را برقرار سازد بسیار دشوار است. به همین دلیل، تعریف کروموزوم‌ها اهمیت ویژه‌ای دارد و باید به گونه‌ای باشد که امکان اعمال تابع تناسب و توابع دیگر بر روی آن فراهم باشد.

## بخش دو: تولید جمعیت اولیه

پس از تعریف و پیاده‌سازی کروموزوم‌ها، باید جمعیت اولیه‌ای از کروموزوم‌ها را به صورت تصادفی بسازید. تعداد این جمعیت می‌تواند به عنوان یک پارامتر حل مسئله باشد و به انتخاب‌های شما بستگی دارد.

## بخش سه: پیاده‌سازی و مشخص کردن تابع معیار سازگاری

بعد از تولید جمعیت اولیه، نیاز داریم تا تابع معیاری تعریف کنیم که بتواند برای شناسایی کروموزوم‌های برتر که شرایط و محدودیت‌های مسئله را بهتر مدل می‌کنند استفاده شود. ابتدا یک تعریف مناسب برای این تابع معیار ارائه دهید، سپس آن را برای این مسئله پیاده‌سازی کرده و میزان سازگاری جمعیت خود را بدست آورید.

## بخش چهار: پیاده‌سازی crossover و mutation و تولید نسل بعدی

حال برای اینکه به یک پاسخ از مسئله داده شده نزدیک شویم، نیاز است در هر نسل، جمعیت جدیدی با استفاده از جمعیت نسل قبل تولید گردد. برای این کار باید از روش‌های crossover و mutation استفاده گردد. در این بخش از دو نوع روش crossover و دو نوع روش mutation از میان روش‌های تدریس شده استفاده شود.

تابع crossover بر روی دو کروموزوم اعمال می‌شود و آن‌ها را ترکیب می‌کند تا به کروموزوم‌هایی از ترکیب آن دو که در حالت ایده‌آل بهترین ویژگی‌های دو ژن اولیه را دارند برسد. این ترکیب و نرخ ایجاد آن باید به عنوان پارامترهای مسئله در نظر گرفته شوند.

تابع mutation بر روی یک کروموزوم اعمال می‌شود و آن را جهش یا تغییر می‌دهد؛ به این امید که بتواند به کروموزوم‌های بهتری جهش پیدا کند. می‌توانید درصد معقولی از ژن‌های برتر را نیز برای انتقال مستقیم به نسل‌های آینده در نظر بگیرید.

## بخش پنج: ایجاد الگوریتم ژنتیک روی مسئله

پس از انجام بخش‌های بالا، باید این توابع پیاده‌سازی شده را در یک الگوریتم استفاده کنید. توجه کنید که نیاز است هاپر پارامترهایی برای میزان randomness و نحوه نزدیک شدن به پاسخ نهایی خود داشته باشید که با تغییر آن‌ها به جواب بهتری برسید.

## بخش شش: ارزیابی نتایج

1. جمعیت اولیه‌ی بسیار کم یا بسیار زیاد چه مشکلاتی را به وجود می‌آورند؟
2. اگر تعداد جمعیت در هر دوره افزایش یابد، چه تأثیری روی دقت و سرعت الگوریتم می‌گذارد؟
3. تأثیر هر یک از عملیات های crossover و mutation را بیان و مقایسه کنید. آیا می‌توان فقط یکی از آن‌ها را استفاده کرد؟ چرا؟
4. در مسئله اول از هر دو روش کد گذاری binary-coded و real-coded استفاده کنید و نتایج را مقایسه کنید.
5. به نظر شما چه راهکارهایی برای سریع‌تر به جواب رسیدن در این مسئله وجود دارد؟
6. با وجود استفاده از این روش‌ها، باز هم ممکن است که کروموزوم‌ها پس از چند مرحله دیگر تغییر نکنند. دلیل این اتفاق و مشکلاتی که به وجود می‌آورد را شرح دهید. برای حل آن چه پیشنهادی می‌دهید؟
7. چه راه حلی برای تمام شدن برنامه در صورتی که مسئله جواب نداشته باشد ارائه می‌دهید؟
8. با افزایش درجه چندجمله‌ای در مسئله اول، زمان پیدا کردن ضرایب به چه صورت تغییر می‌کند؟
9. به نظر شما افزایش و یا کاهش تعداد نقاط به چه صورت در اجرای الگوریتم تأثیر می‌گذارد؟
10. نحوه انتخاب کدگذاری تان (binary-coded, real-coded, tree-coded) برای حل مسائل را توضیح دهید.
11. **امتیازی:** در مسئله اول کروموزوم‌ها را به گونه‌ای طراحی کنید که شامل درجه چندجمله‌ای (مثلاً به عنوان ژن اول) نیز باشند و بر اساس درجه چندجمله‌ای، تعداد بیت‌های هر ژن مشخص شوند. توجه کنید که باید مفهوم جدیدی برای fitness تعریف کنید که با درجه چندجمله‌ای رابطه غیرمستقیم داشته باشد و در ازای افزایش درجه چندجمله‌ای، دقت برنامه بیشتر از حالات با درجه پایین‌تر باشد.

\* جواب این سوالات را به صورت کامل در برنامه خود اعمال نموده و توضیحات لازم را در گزارش خود

بنویسید.

سوال دوم:

[https://drive.google.com/drive/folders/1dH7sYqzF3wg-B4PpK7\\_2cgisnssQvq84?usp=sharing](https://drive.google.com/drive/folders/1dH7sYqzF3wg-B4PpK7_2cgisnssQvq84?usp=sharing)