# هوش مصنوعی — تمرین کامپیوتری شماره ۲: ژنتیک

مدرس: **دکتر فدایی و دکتر یعقوب‌زاده** طراحان: سید پارسا حسینی‌نژاد، پریا خوش‌تاب

مهلت تحویل: یکشنبه ۲۱ فروردین ۱۴۰۱، ساعت ۲۳:۵۵

## مقدمه

در این پروژه، با روش‌هایی که برگرفته از طبیعت و انتخاب طبیعی هستند، آشنا می‌شویم. در این روش‌ها که به طور کلی الگوریتم‌های 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.

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

در این پروژه با استفاده از الگوریتم‌های ژنتیک، مسئله‌ی رمزگشایی که یکی از مسائل پرکاربرد در حوزه‌ی علوم کامپیوتری می‌باشد را پیاده‌سازی خواهیم کرد.

## مسئله رمزگشایی

هدف از رمزنگاري، تبدیل متن خام (پیام) به متن رمزشده است تا هیچ‌کس جز مقصد پیام آن را نفهمد. هر چه بازگرداندن متن رمزشده به پیام اصلی از نظر زمانی پیچیده تر باشد، رمزنگاری ارزشمندتر است.

رمزنگاری استفاده شده در این پروژه با استفاده از یک کلید رمزنگاری[1] و یک جدول صورت می‌گیرد. به این منظور، ابتدا یک کلید رمزنگاری دلخواه انتخاب می‌شود. سپس یک رشته‌ی جدید با استفاده از این کلید ساخته می‌شود و در انتها و با استفاده از جدول رمزنگاری، متن رمز شده ساخته می‌شود.

برای شروع عملیات رمزنگاری، در ابتدا یک کلید دلخواه انتخاب می‌شود و این کلید به تعداد حروف متنی که می‌خواهیم رمز کنیم تکرار می‌شود. برای مثال اگر متنی که می‌خواهیم آن را رمز کنیم عبارت

"WE ARE DISCOVERED SAVE YOURSELF"

[1] Encryption Key

باشد و کلید انتخاب شده کلمه‌ی "RUN" باشد، رشته‌ی اولیه‌ای که تولید می‌شود تا در مرحله‌ی بعدی رمزنگاری که با استفاده از جدول صورت می‌گیرد مطابق عبارت زیر است:

"RU NRU NRUNRUNRUN RUNR UNRUNRUN"

سپس این عبارت تولید شده با استفاده از جدول رمزنگاری به رشته‌ی رمز اصلی تبدیل می‌شود. به این عبارت، کلید نیز می‌گویند (دقت کنید این کلید با کلیدی که در ابتدا استفاده شد متفاوت است و فقط تشابه اسمی دارند).

جدول رمزنگاری استفاده شده در این پروژه جدول مشخص شده در تصویر (۱) است. این جدول شامل حروف الفبای انگلیسی می‌باشد که ۲۶ بار در ردیف‌های مختلف نوشته شده است و هر الفبا در مقایسه با الفبای ردیف قبلی به صورت چرخه‌ای، یک واحد به سمت چپ انتقال یافته‌است. به عنوان مثال، هنگامی که B به موقعیت اول در ردیف دوم شیفت پیدا می‌کند، حرف A به انتهای ردیف دوم منتقل می‌شود. حال، برای تبدیل هر حرف رشته‌ی محاسبه شده در مرحله‌ی قبل (کلید) به عبارت رمز، باید ستونی که مربوط به حرف مورد نظر در پیام اصلی است و سطری که مربوط به حرف مورد نظر در کلید محاسبه شده است انتخاب شود و از تقاطع این دو، حرف رمز به دست آید. برای مثال، با توجه به عبارتی که در قسمت‌های قبل دیده‌شد، اولین حرف رمز N خواهد بود، چون تقاطع ستون W و سطر R برابر N است.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z |
| B | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A |
| C | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B |
| D | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C |
| E | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D |
| F | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E |
| G | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F |
| H | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G |
| I | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H |
| J | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I |
| K | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J |
| L | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K |
| M | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L |
| N | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M |
| O | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
| P | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
| Q | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P |
| R | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q |
| S | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R |
| T | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S |
| U | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T |
| V | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U |
| W | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V |
| X | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W |
| Y | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X |
| Z | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y |

تصویر ۱ - جدول رمزنگاری

با توجه به توضیحات بالا، عبارت رمزشده‌ی مثال ذکر شده در قسمت‌های بالا به شکل زیر خواهد بود.

| | |
|---|---|
| متن خام: | W E A R E D I S C O V E R E D S A V E Y O U R S E L F |
| کلید: | R U N R U N R U N R U N R U N R U N R U N R U N R U N |
| متن رمزشده: | N Y N I Y Q Z M P F P R I Y Q J U I V S B L L F V F S |

در این نوع فرآیند رمزنگاري حروف بزرگ به حروف بزرگ متناظر خود و حروف کوچک به حروف کوچک متناظر خود نگاشت می‌شوند.

## پیاده‌سازی مسئله

در این پروژه یک پیام رمزشده و یک متن مرجع به شما داده می‌شود و شما باید با استفاده از الگوریتم ژنتیک، کلید موردنظر را پیدا کنید و پیام رمزگشایی‌شده را بازگردانید. **دقت شود که کلید رمزنگاری در این پروژه ۱۴ حرف دارد.** بنابراین فضای حالت کلید حدود $۱۰^{۲۰}$ است و محاسبه‌ی آن با روش‌های آزمون و خطا با کامپیوترهای شخصی بسیار طولانی خواهد شد و استفاده از الگوریتم‌های ژنتیک بسیار موثر است. شما باید مراحل زیر را همانطور که در درس نیز آموخته‌اید پیاده‌سازی کنید و سپس با تجمیع تمام این مراحل یک الگوریتم کلی برای حل مسئله پیاده‌سازی کنید.

### بخش صفر: تمیز کردن داده‌ها و ایجاد لغت‌نامه

در اولین مرحله شما باید متن مرجعی که به شما داده شده است را پردازش کنید و از آن یک لغت‌نامه استخراج کنید تا بتوانید از آن در بخش‌های بعد و برای محاسبه‌ی امتیاز تناسب کروموزوم از آن استفاده کنید. یکی از مراحل پردازش مجموعه‌داده، تمیزکردن داده‌هاست. تمیزکردن در این پروژه چگونه باید انجام شود؟ (به کاراکترهای غیر کارآمد و ایست‌واژه‌ها[2] فکر کنید). روند تولید لغت‌نامه از متن داده شده را توضیح دهید.

### بخش یک: مشخص‌کردن 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 باید با استفاده از تصادفی کردن وقایع انجام دهید چرا که اگر فضای حالت بزرگ باشد، پیدا کردن شرطی که همه‌ی محدودیت‌ها را برقرار سازد بسیار دشوار است. به همین دلیل، تعریف کروموزوم اهمیت ویژه ای دارد و باید به گونه ای باشد که امکان اعمال تابع تناسب و توابع دیگر بر روی آن فراهم باشد.

[2] Stop Words

## بخش دو: تولید جمعیت اولیه

پس از تعریف و پیاده‌سازی ساخت یک کروموزوم، باید جمعیت اولیه‌ای از کروموزوم‌ها به صورت کاملا رندوم و بدون هیچ جهت‌گیری خاصی بسازید. تعداد این جمعیت می‌تواند به عنوان یک پارامتر حل مسئله باشد و به انتخاب‌های شما بستگی دارد.

## بخش سه: پیاده‌سازی و مشخص‌کردن تابع معیار سازگاری[3]

بعد از تولید جمعیت اولیه، نیاز داریم تا تابع معیاری تعریف کنیم که بتواند برای تشخیص کروموزوم‌های برتر از این نظر که شرایط و محدودیت‌های مسئله را فراهم کنند استفاده شود. در این مسئله برای محاسبه امتیاز تناسب کروموزوم، از لغت‌نامه‌ای که در بخش صفر به وجود آمد استفاده می‌کنیم. تضمین می‌شود که تمامی لغات پیام‌ها در متن پیوست آمده‌اند (در پروژه هاي واقعي خودتان بايد لغت‌نامه خود را ايجاد و گسترش دهيد تا پوشش مناسب را ايجاد كنيد).

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

حال برای اینکه به کلید رمزگشایی هدف مسئله نزدیک شویم، نیاز به ایجاد جمعیت جدید از جمعیت‌های نسل قبل خود داریم. این کار را باید با روش‌های معروفی که در الگوریتم ژنتیک وجود دارد و در درس نیز با آن‌ها آشنا شده‌اید، یعنی mutation و crossover انجام دهید.

تابع crossover بر روی دو کروموزوم اعمال می شود و آن‌ها را ترکیب می کند تا به کروموزوم‌های بهتری برسد. این ترکیب و نرخ ایجاد آن می تواند به عنوان پارامترهای مسئله باشد. تابع mutation بر روی یک کروموزوم اعمال می‌شود و با استفاده از روشی آن را جهش داده و تغییر می‌دهد، به امید آن که بتواند به کروموزوم بهتری دست یابد. همچنین می‌توانید از درصد معقولی از ژن‌های برتر برای انتقال مستقیم به نسل‌های آینده نیز استفاده کنید.

## بخش پنج: ایجاد الگوریتم ژنتیک روی مسئله

در آخر باید این توابع پیاده‌سازی شده را در یک الگوریتم استفاده کنید. توجه کنید که می‌توانید پارامترهایی برای راه خود داشته باشید که با تغییر آن به جواب بهتری برسید.

## بخش شش: سوالات

1. جمعیت اولیه‌ی بسیار کم یا بسیار زیاد چه مشکلاتی را به وجود می‌آورند؟
2. اگر تعداد جمعیت در هر دوره افزایش یابد، چه تاثیری روی دقت و سرعت الگوریتم می‌گذارد؟
3. چرا در الگوریتم‌های ژنتیک هم از crossover و هم از mutation استفاده می‌شود؟ اگر از هر کدام از آن‌ها استفاده نشود چه مشکلی ممکن است پیش بیاید؟
4. کدام یک از عملیات‌های crossover و mutation تاثیر بیشتری در بالاتر رفتن دقت دارند؟ کدام یک این تاثیر را با سرعت بالاتری می‌گذارند؟

[3] Fitness Function

5. با استفاده از روش‌های crossover و mutation همچنان ممکن است پس از چند مرحله جمعیت تغییر نکند. این مشکل را چگونه می‌توان از بین برد؟
6. اگر قرار بود تنها از یکی از روش‌های crossover و mutation استفاده شود، به نظر شما کدام موثرتر واقع می‌شد؟ چرا؟
7. به نظر شما چه راهکارهایی برای سریعتر به جواب رسیدن در این مسئله‌ی خاص وجود دارد؟

## نکات پایانی

- موعد تحویل غیرحضوری تا پایان روز ۲۱ فروردین میباشد.
- تمامی نتایج باید در یک فایل فشرده با عنوان AI-CA2-Genetic-#SID.zip تحویل داده شود. این فایل باید شامل موارد زیر باشد:
  - یک پوشه به نام code شامل کدهای تمام قسمت‌هایی از تمرین که پیاده‌سازی کرده‌اید.
  - گزارش پروژه با فرمت PDF و شامل شرح تمامی کارهای انجام شده، نتایج به دست آمده و تحلیلها و بررسیهای خواسته شده در صورت پروژه.
  - در صورتی که از Jupyter Notebook استفاده میکنید نیازی به ارسال جداگانه کدها و گزارش نیست و هر دو را میتوانید در یک فایل Notebook قرار دهید. حتما خروجی html فایل Notebook خود را نیز همراه فایل Notebook ارسال کنید.
- پروژه‌ی خود را در قالب یک کلاس Decoder پیاده‌سازی کنید. این کلاس باید تمامی ویژگی‌های لازم برای رمزگشایی را داشته باشد. سازنده‌ی این کلاس متن مرجع، متن کد شده و طول کلید را به عنوان ورودی دریافت می‌کند. این کلاس تابع decode را پیاده‌سازی می‌کند که ورودی ندارد و خروجی آن متن رمزگشایی شده است. دقت کنید که تست‌ها به صورت اتوماتیک اجرا خواهند شد. کد شما باید بتواند به صورت زیر تست شود:

```
encodedText = open('encoded_text.txt').read()
globalText = open('global_text.txt').read()
d = Decoder(globalText, encodedText, keyLength = CONSTANT)
decodedText = d.decode()
```

- توجه داشته باشید علاوه بر ارسال فایل‌های پروژه، این پروژه تحویل نیز گرفته خواهد شد. بنابراین لازم است بر تمامی قسمتهای کدتان تسلط کافی داشته باشید و تمام بخش‌های پروژه باید قابلیت اجرای مجدد در زمان تحویل را داشته باشند. همچنین در صورت عدم حضور در زمان تحویل، نمره‌ای دریافت نخواهید کرد.
- هیچگونه شباهتی در انجام این پروژه بین افراد مختلف پذیرفته نمیشود. در صورت کشف هرگونه تقلب برای همه افراد متقلب نمره ۱۰۰- در نظر گرفته میشود.
- استفاده از مراجع با ارجاع به آنها بلامانع است. اما در صورتی که گزارش شما ترجمه عینی از آنها باشد یا از گزارش افراد دیگر استفاده کرده باشید کار شما تقلب محسوب میشود.

- در صورتی که سوالی در مورد پروژه داشتید بهتر است در فروم یا گروه تلگرامی درس مطرح کنید تا بقیه از آن استفاده کنند، در غیر این صورت می‌توانید به طراحان پروژه ایمیل بزنید و سوالتان را از یکی از آنها بپرسید. ایمیل طراحان نیز در ابتدای تمرین مشخص شده‌است.

**موفق باشید!**