



مرحله اول پروژه

مرحله اول پروژه درس از دو بخش تشکیل شده است. در بخش اول باید از الگوریتم‌های جستجو برای پیاده‌سازی یک بازیکن هوشمند در بازی معروف Pacman استفاده کنید و در بخش دوم با پیاده‌سازی الگوریتم کاهش گرادیان یک مدل رگرسیون خطی برای مسأله داده شده آموزش داده و آن را مورد ارزیابی قرار دهید. توجه کنید که باید برای پروژه خود بعد از پیاده‌سازی گزارش تهیه کرده و در آن پیاده‌سازی خود را توضیح داده و نتایج را نمایش دهید. فایل‌های پروژه در این [لینک](#) قرار دارند.

۱- بازی Pacman

در این بخش از پروژه می‌خواهیم عامل هوشمندی پیاده‌سازی کنیم که در مورد نحوه دستیابی به غذاهای موجود در بازی معروف Pacman تصمیم‌گیری کند. در این مسئله که Corner Problem نام دارد، از الگوریتم‌های جستجوی آگاهانه و ناآگاهانه برای هدایت عامل Pacman به سمت غذاها استفاده می‌شود. الگوریتم‌های در نظر گرفته شده برای پیاده‌سازی UCS، DFS و A^* هستند. شرط برنده شدن این است که عامل Pacman همه غذاهای موجود در محیط را مصرف کند.

۱-۱- مراحل پیاده‌سازی

برای تسهیل پیاده‌سازی عامل، ساختار کد عامل از قبل ایجاد شده و شما باید فقط بخش‌های مرتبط با الگوریتم‌های جستجو را که در زیر مشخص شده است تکمیل کنید.

۱. به فایل `search.py` مراجعه کنید. در این فایل سه تابع وجود دارد که شما باید پیاده‌سازی آن را تکمیل کنید:

- تابع `depthFirstSearch` محل پیاده‌سازی الگوریتم DFS
- تابع `uniformCostSearch` محل پیاده‌سازی الگوریتم UCS
- تابع `aStarSearch` محل پیاده‌سازی الگوریتم A^*

۲. تابع اکتشافی مورد استفاده در الگوریتم A^* در تابع `cornersHeuristic` در ابتدای فایل `searchAgents.py` تعریف می‌شود که باید آن را پیاده‌سازی کنید. در پیاده‌سازی این تابع می‌توانید از توابع تعامل با محیط که در اختیار شما قرار گرفته است استفاده کنید.

توجه کنید که توابع پیاده‌سازی شده باید بتوانند در نمونه‌های مختلفی از مسأله Corner Problem بکار گرفته شوند. نمونه‌هایی که می‌توانند از نظر ابعاد محیط، تعداد غذا و موقعیت اولیه Pacman متفاوت باشند. به همین دلیل سه محیط اولیه در اختیار شما قرار داده شده است تا برای آزمایش توابع پیاده‌سازی شده مورد استفاده قرار گیرد و در ادامه توضیح داده شده است.

۱-۲- محیط‌ها

- محیط Simple Corner: این محیط، ساده‌ترین محیط است و می‌توانید برای ارزیابی صحت پیاده‌سازی الگوریتم‌ها از آن استفاده کنید. جهت اجرای این محیط با هر یک از الگوریتم‌های جستجوی پیاده‌سازی شده می‌توانید از دستورات زیر استفاده کنید:

```
python pacman.py -l simpleCorner -p SearchAgent -a
fn=dfs,prob=CornersProblem

python pacman.py -l simpleCorner -p SearchAgent -a
fn=ucs,prob=CornersProblem

python pacman.py -l simpleCorner -p SearchAgent -a
fn=astar,prob=CornersProblem,heuristic=cornersHeuristic
```

- **Hard Corner:** این محیط نسبت به محیط simple corner پیچیده تر است و می تواند برای ارزیابی تابع اکتشافی مورد استفاده در الگوریتم A^* به صورت زیر مورد استفاده قرار گیرد:

```
python pacman.py -l hardCorner -p SearchAgent -a
fn=astar,prob=CornersProblem,heuristic=cornersHeuristic
```

- **Big Corner:** این محیط دارای ابعاد بزرگتری نسبت به محیط های قبلی بوده و به صورت زیر قابل فراخوانی است:

```
python pacman.py -l BigCorner -p SearchAgent -a
fn=astar,prob=CornersProblem,heuristic=cornersHeuristic
```

۱-۳- نحوه ارزیابی

هر بار پس از اجرای محیط، یک گزارش در ترمینال چاپ می شود که نشان دهنده امتیاز کسب شده، زمان اجرا و تعداد گره های بسط داده شده است. ارزیابی کیفیت تابع اکتشافی پیاده سازی شده براساس تعداد گره های بسط داده شده توسط الگوریتم جستجوی A^* خواهد بود، که در جدول زیر امتیاز مرتبط برای محیط های Hard Corner و Big Corner نشان داده شده است:

Score	Expanded Nodes in BigCorner	Expanded Nodes in HardCorner
100%	Less than 2000	Less than 800
80%	Less than 3000	Less than 1200
60%	Less than 4500	Less than 1500

همانطور که قبلاً گفته شد در ارزیابی پیاده سازی انجام شده از نمونه محیط های دیگری نیز استفاده می شود که انتظار می رود تعداد گره های بسط داده شده در حد قابل قبول باشد. بنابراین علاوه بر صحت پیاده سازی الگوریتم های DFS، UCS و A^* در فایل search.py، تابع اکتشافی پیاده سازی شده در searchAgents.py نیز از نظر تعداد گره های بسط داده شده مورد بررسی قرار خواهد گرفت.

۲- رگرسیون خطی

در این بخش باید الگوریتم آموزش یک مدل رگرسیون خطی (Linear Regression) از روی داده های داده شده در مورد وقوع سیل را پیاده سازی کنید. مدل آموزش داده شده باید بتواند با دریافت ورودی، پیش بینی کند که چقدر احتمال دارد سیل رخ دهد.

۲-۱- بررسی مجموعه داده

مجموعه داده به دو بخش آموزش و آزمایش تقسیم بندی شده است. از بخش آموزش (Train) برای آموزش مدل و از بخش آزمایش (Test) برای ارزیابی مدل آموزش دیده استفاده می شود. هر نمونه (سطر) در این مجموعه داده نشان دهنده اطلاعات یک منطقه بوده و ستون های آن نمایانگر ویژگی های یک منطقه می باشند که عبارتند از:

- Id: برای هر ناحیه، یک id منحصر به فرد وجود دارد.
- Monsoon Intensity: این ویژگی برای هر نمونه، نشان دهنده شدت بادهای موسمی در آن منطقه می باشد.
- Topography Drainage: این ویژگی میزان تخلیه آب از یک حوزه زهکشی شده را نشان می دهد.
- River Management: این ویژگی امتیازی برای مدیریت منابع رودخانه است.
- Deforestation: امتیازی که به میزان جنگل زدایی از آن منطقه اشاره دارد.
- Urbanization: این ویژگی میزان شهرنشینی در آن منطقه را نشان می دهد.
- Climate Change: این ویژگی امتیازی است که به میزان تاثیر تغییرات اقلیمی در این منطقه داده شده است.
- Dams Quality: این ویژگی امتیازی است که به کیفیت سدهای آن منطقه اشاره دارد.
- Siltation: این ویژگی امتیازی است که به میزان آلوده شدن آب بوسیله خاک، ماسه و سایر مواد تعلق می گیرد.
- Agricultural Practices: امتیازی که به کیفیت شیوه های کشاورزی در آن ناحیه تعلق می گیرد.
- Encroachments: امتیازی که به میزان تجاوز به حریم رودخانه در یک ناحیه تعلق گرفته است.
- Ineffective Disaster Preparedness: امتیازی که ناکارآمدی آن منطقه در مواجهه با بلایای طبیعی را نشان می دهد.
- Drainage Systems: امتیازی که به کیفیت سیستم های زهکشی یک منطقه اشاره می کند.
- Coastal Vulnerability: این ویژگی نشان دهنده آسیب پذیری ساحلی یک منطقه می باشد.
- Landslides: امتیازی که به رانش زمین در آن ناحیه تعلق می گیرد.
- Watersheds: امتیازی که به حوزه های آبخیزداری یک منطقه اشاره دارند.
- Deteriorating Infrastructure: امتیازی که به میزان وخامت زیرساخت های اساسی یک منطقه اشاره می کند.
- Population Score: امتیازی که به میزان جمعیت ساکن در یک منطقه اشاره می کند.
- Wetland Loss: امتیازی که به میزان از دست رفتن تالاب های یک منطقه اشاره دارد.
- Inadequate Planning: امتیازی که نشان دهنده عدم سطح آمادگی و برنامه ریزی در برابر بلایای طبیعی می باشد.
- Political Factors: امتیازی که به تاثیر اهمیت سیاسی یک منطقه در مواجهه با سیل به آن منطقه اختصاص یافته است.
- Flood Probability: احتمال وقوع سیل در یک منطقه (همان عددی است که قرار است مدل شما پیش بینی کند).

۲-۲- مراحل پیاده سازی

برای پیاده سازی این بخش باید گام های زیر را انجام شود:

۱. خواندن مجموعه داده های آموزش و آزمایش.
۲. (اختیاری) بررسی وجود مقادیر ناموجود (missing) در مجموعه داده. در صورت وجود چنین مقادیری، از یک تکنیک دلخواه برای جایگزینی آنها استفاده کنید.
۳. (اختیاری) مهندسی ویژگی. در این مرحله شما می توانید در ویژگی های مجموعه داده تغییراتی ایجاد کنید. مثلاً بعضی از ویژگی ها را حذف کرده یا از روی ویژگی های فعلی، ویژگی های جدیدی استخراج کنید.
۴. (اختیاری) پیاده سازی یک تکنیک Feature Scaling و اعمال بر روی مجموعه داده.

۵. تعیین ابرپارامترهای آموزش از قبیل Learning Rate و Max Epochs
۶. پیاده‌سازی الگوریتم بهینه‌سازی Stochastic Gradient Descent (از Scratch) برای آموزش مدل رگرسیون خطی
۷. آموزش مدل با مجموعه داده آموزش
۸. رسم نمودار تغییر مقادیر تابع زیان در هنگام آموزش برحسب Epoch
۹. ارزیابی مدل با مجموعه داده آزمایش

۲-۳- ابزارهای پیشنهادی

برای پیاده‌سازی این بخش در زبان Python می‌توانید در صورت نیاز از کتابخانه‌های زیر بهره ببرید:

- Pandas: جهت اعمال تغییرات در مجموعه داده
 - Numpy: جهت انجام عملیات‌های ریاضی و جبرخطی
 - Seaborn و Matplotlib: جهت نمایش نمودارهای مختلف
- همچنین می‌توانید از محیط‌های Data Spell, Jupiter Notebook یا Google colab برای پیاده‌سازی استفاده کنید.

۲-۴- نحوه ارزیابی

برای هر یک از مراحل مشخص شده برای پیاده‌سازی امتیازی در نظر گرفته شده است. بخش‌های اختیاری برای آن تعریف شده‌اند که بتوانید مدلی آموزش دهید که در زمان ارزیابی نتایج بهتری را از خود نشان بدهد (امکان دارد که در صورت استفاده نادرست، نتایج بدتری را شاهد باشیم). در پیاده‌سازی این بخش استفاده از کتابخانه scikit learn یا سایر کتابخانه یادگیری ماشین در هیچ مرحله‌ای (به غیر از ارزیابی) مجاز نیست. تمام گام‌هایی که در مرحله پیاده‌سازی قید شده‌اند، باید بدون استفاده از کتابخانه‌های آماده پیاده‌سازی شوند. در صورت مشاهده استفاده از کتابخانه‌های آماده بجز برای ارزیابی، نمره منفی به شما تعلق می‌گیرد. در صورت صحت پیاده‌سازی و دستیابی به امتیاز $R^2 \text{ Score} \geq 0.8$ کامل برای این بخش در نظر گرفته می‌شود. در گزارش نتایج ارزیابی باید معیارهای Mean Squared Error و Mean Absolute Error روی مجموعه داده آموزش و آزمایش نیز گزارش شود. برای گزارش این موارد می‌توانید از ماژول sklearn.metrics استفاده کنید.

۲-۵- موارد اضافی

انجام هر یک از موارد زیر نمره اضافه در این مرحله از پروژه خواهد داشت:

- بکارگیری تکانه (momentum) در بهینه‌ساز
- تغییر نرخ یادگیری در حین آموزش با یک سیاست خاص. در صورت پیاده‌سازی باید نمودار تغییر نرخ یادگیری برحسب Epoch را نمایش دهید.
- استفاده از توقف زود هنگام در بهینه‌سازی.
- استفاده از یک تکنیک Regularization در تخمین پارامترهای مدل رگرسیون خطی

در صورت پیاده‌سازی هر یک از این موارد اضافی، تمام اعضای تیم باید دلیل استفاده از آن و منطق استفاده از آنرا بدانند. این موضوع باید هم در گزارش قید شود و هم در زمان تحویل پروژه کاملاً بر آن مسلط باشند.

موفق باشید