



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

دانشکده ریاضی و علوم کامپیوتر

استاد درس: دکتر مهدی قطعی

استاد کارگاه: بهنام یوسفی مهر

پاییز ۱۴۰۳

راهنمای تمرین ۶

درس هوش مصنوعی و کارگاه



مقدمه

در این تمرین شما باید با استفاده از الگوریتم ژنتیک روی دیتاست Superstore Sales اهداف بیان شده را برآورده کنید.

• Superstore Sales

این دیتاست بینش ارزشمندی در مورد فروش، رفتار مشتری و عملکرد محصول ارائه می دهد و شامل ۹۹۹۴ داده شامل ۲۱ فیچر است. با بررسی و تحلیل روی این دیتاست میتوان تأثیر عوامل مختلف مانند تخفیف ها، موقعیت های جغرافیایی و دسته بندی محصولات بر سودآوری را مورد بررسی قرار داد. به عنوان اهداف این تمرین، میتوان به ۳ مورد اشاره کرد:

۱. بهینه سازی سود: این هدف به دنبال یافتن بهترین ترکیب از محصولات، تخفیف ها و تعداد فروش است تا بیشترین سود ممکن حاصل شود. به عبارت دیگر، می خواهیم با استفاده از اطلاعات موجود، تحلیل کنیم که کدام محصولات باید با تخفیف بیشتری ارائه شوند یا کدام ترکیب از محصولات و قیمت ها بیشترین بازدهی را برای کسب و کار به همراه دارد.

۲. بهینه سازی تخفیف و فروش: بهینه سازی تخفیف و فروش به معنای یافتن تعادل مناسب بین میزان تخفیف ارائه شده و تعداد محصولاتی است که فروخته می شود. در واقع، هدف این است که با کاهش تخفیف ها و در عین حال افزایش فروش، کسب و کار بتواند سود بیشتری کسب کند. به این ترتیب، باید بررسی کرد که چگونه تخفیف های مختلف بر میزان تقاضا و سود نهایی تأثیر می گذارند.

۳. پیش بینی الگوهای خرید مشتریان: این هدف به شناسایی روندهای خرید مشتریان و پیش بینی رفتارهای آتی آنها مربوط می شود. به کمک این تحلیل ها، می توان الگوهای خرید مشتریان را شبیه سازی کرد و پیش بینی هایی در مورد اینکه چه محصولات یا خدماتی ممکن است در آینده برای آنها جذاب باشد، ارائه داد. (ممکن است به الگوریتم (خوشه بندی) دیگری به غیر از ژنتیک هم برای پیش بینی نیاز داشته باشید).

مراحل اجرای تمرین

پیش پردازش داده

- ساختار داده و اطلاعات موجود را بررسی کنید
- در صورت وجود داده ی null آن را حذف و یا تکمیل کنید.



- در صورت وجود داده‌ی اسمی بسته به نوع آن، آن را One-Hot Encode یا Label Encode کنید.
- برای کارایی بهتر مدل، می‌توانید داده‌های پرت و نویز را در دیتاست شناسایی کرده و حذف کنید.

توضیح الگوریتم

الگوریتم ژنتیک (GA) یک جستجوی اکتشافی است که از فرآیند انتخاب طبیعی الهام گرفته شده است. برای یافتن راه حل های تقریبی برای مسائل بهینه سازی و جستجو استفاده می شود.

تعاریف

- جمعیت: مجموعه ای از راه حل های بالقوه برای یک مشکل.
- کروموزوم ها: نمایشی از یک جواب که اغلب به صورت رشته کدگذاری می شود.
- ژن ها: بخش هایی از یک کروموزوم که پارامترها یا متغیرهای خاصی را نشان می دهند.
- تابع fitness: تابعی است که بر اساس نزدیکی هر کروموزوم به راه حل بهینه، آن را ارزیابی و مقداری به آن اختصاص می دهد.
- انتخاب: فرآیند انتخاب مناسب ترین کروموزوم ها برای ایجاد فرزندان.
- crossover: ترکیب بخشی از دو کروموزوم مادر برای تشکیل فرزندان.
- mutation: معرفی تغییرات تصادفی در یک کروموزوم برای حفظ تنوع در یک جمعیت.

مراحل اجرای الگوریتم

۱. مقداردهی اولیه: به طور تصادفی مجموعه ای اولیه از کروموزوم ها را تولید کنید.
۲. ارزیابی: تناسب هر کروموزوم را با استفاده از تابع fitness محاسبه کنید.
۳. انتخاب: جفت کروموزومی توانایی بالایی در تولید مثل دارند را انتخاب کنید.
۴. crossover: جفت کروموزوم ها با هم ترکیب کنید و فرزندان جدیدی را تشکیل دهید.
۵. mutation: در برخی از فرزندان تغییرات تصادفی اعمال کنید.



۶. جایگزینی: ساکنان قدیمی را با ساکنان جدید جایگزین کنید.

۷. خاتمه: مراحل را چندین بار تکرار کنید تا شرایط خاتمه برآورده شود (به عنوان مثال، یک راه حل پیدا شود و به حداکثر تعداد نسل برسد).

مثال

فرض کنید می‌خواهیم تابع $f(x) = x^2$ ($0 \leq x < 5$) را ماکسیم کنیم. مرحله‌ی اول الگوریتم مقداردهی اولیه است. بدین ترتیب چند رشته‌ی زیر را به صورت تصادفی تولید می‌کنیم: [10101, 01100, 11110, 00011]. از آنجایی که هم‌دامنه‌ی تابع در بازه‌ی $[0, 32]$ قرار دارد بنابراین از رشته‌های ۵ بیتی برای این کار استفاده کردیم. در ادامه نیاز است که هر جواب را ارزیابی کنیم. بنابراین داریم:

- برای $x = 101010 = 21$ $f(x) = 441$

- برای $x = 01100 = 12$ $f(x) = 144$

- برای $x = 11110 = 30$ $f(x) = 900$

- برای $x = 00011 = 3$ $f(x) = 9$

در مرحله‌ی بعد جفت‌های کروموزوم را بر اساس تناسب آنها انتخاب کنید. مقادیر بالاتر تناسب شانس بیشتری برای انتخاب دارند. بنابراین دو جفت [11110, 01100] & [11110, 10101] انتخاب می‌شوند. حال، crossover را در یک نقطه تصادفی انجام دهید. فرض کنید نقطه تقاطع بعد از بیت دوم قرار دارد. بنابراین

- [10101] & [11110] جفت [10110] & [11101] را تولید می‌کنند.

- [01100] & [11110] جفت [01110] & [11100] را تولید می‌کنند.

در مرحله‌ی بعد جهش را با یک احتمال کم اعمال کنید. حال جمعیت جدید حاصل شده را با جمعیت قدیمی جایگزین کنید و این مراحل را چندین بار تکرار کنید.

حال بدین ترتیب شما باید ارتباطی بین این الگوریتم و روش‌های خوشه‌بندی بیابید. برای مثال در پروسه‌ی مقداردهی اولیه می‌توانید یک centroid را مقداردهی کنید. هم‌چنین تابع fitness را می‌توانید مینیمم کردن فاصله‌ی اقلیدسی بین نقاط داخل یک خوشه قرار دهید.

هم‌چنین خروجی این الگوریتم را با الگوریتم k-means مقایسه نموده و نتایج را در قالب نمودارهای مختلف ارائه نمایید.



معیارهای سنجش تمرین

- درک درست از Evolutionary Algorithms و اجرا گام به گام این الگوریتم.
- درک نحوه حل مساله خوشه بندی با استفاده از این الگوریتم.
- مقایسه کوتاهی از الگوریتم k-means و خوشه بندی با استفاده از الگوریتم های تکاملی با روش بالا براساس نظر خودتان.
- از آنجایی که درصد درستی خوشه بندی به پیچیده تر و معنا دار بودن تعریف operator ها بستگی دارد درصد درستی خوشه بندی هدف و معیار این تمرین نمیباشد.
- پیاده سازی توابع fitness، mutation، crossover و selection اجباری می باشد و بخشی از نمره ی شما را تشکیل می دهد.
- گزارش کار باید در قالب پایان نامه ی رسمی دانشگاه که در گروه تلگرامی قرار گرفته به صورت فایل PDF فرستاده شود.
- عزیزان توجه بفرمایید که استفاده از مدل های زبانی مانند ChatGPT برای پاسخ دادن به تمرین غیرمجاز است.

مهلت تحویل

- مهلت تحویل تمرین تا ساعت ۲۳:۵۹ روز چهارشنبه ۱۹ دی می باشد.
- لطفا فایل pdf گزارش کار و فایل پیاده سازی کد را در قالب یک فایل zip در سامانه کورسز بارگذاری نمایید.
- در صورت هرگونه مشکل و یا ابهام می توانید با تدریسپاران خود در ارتباط باشید.