



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

دانشکده ریاضی و علوم کامپیوتر

استاد درس: دکتر مهدی قطعی

استاد کارگاه: بهنام یوسفی مهر

بهار ۱۴۰۳

راهنمای تمرین ۶

درس هوش مصنوعی و کارگاه



مقدمه

در این تمرین هدف کلاسترینگ (خوشه‌بندی) دیتاست DryBean با استفاده از الگوریتم ژنتیک می باشد. در ادامه الگوریتم گام به گام توضیح داده شده است.

• Dry Bean Dataset

در این تحقیق از هفت نوع مختلف دانه‌ی خشک با در نظر گرفتن ویژگی‌هایی مانند شکل، نوع و ساختار بسته به وضعیت بازار استفاده شده است. دیتاست شامل ۱۳۶۱۱ دانه شامل ۱۶ فیچر است. همچنین کلاس دانه نیز مشخص به عنوان برچسب برای شما قرار داده شده که بتوانید نظارت‌شده خوشه‌بندی خود را ارزیابی کنید.

مراحل اجرای تمرین

پیش پردازش داده

- ساختار داده و اطلاعات موجود را بررسی کنید
- در صورت وجود داده‌ی 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 می خواهیم تابع $f(x) = x^2$ ($0 \leq x < 5$) را ماکسیم کنیم. مرحله ای اول الگوریتم مقداردهی اولیه است. بدین ترتیب چند رشته ی زیر را به صورت تصادفی تولید می کنیم: [10101, 01100, 11110, 00011]. از آن جایی که هم دامنه ی تابع در بازه ی $[0, 32]$ قرار دارد بنابراین از رشته های 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.
- علاوه بر بخش عملی، پاسخ دادن به سوالات تئوری این تمرین نیز بخشی از ارزیابی شما عزیزان خواهد بود و در ارائه مجازی تمرین ششم، باید به سوالات تئوری تسلط بالایی داشته باشید. این سوالات در گروه تلگرامی در اختیارتان قرار گرفته است.



- گزارش کار باید در قالب پایان‌نامه‌ی رسمی دانشگاه که در گروه تلگرامی قرار گرفته به صورت فایل PDF فرستاده شود.
- عزیزان توجه بفرمایید که استفاده از مدل‌های زبانی مانند ChatGPT برای پاسخ دادن به تمرین غیرمجاز است.

مهلت تحویل

- مهلت تحویل تمرین تا ساعت ۲۳:۵۹ روز ۱۸ خرداد می باشد.
- با استفاده از این **لینک** میتوانید به درایو پروژه ششم دسترسی داشته باشید. سوالات تئوری و دیتاست در این درایو می باشند.
- لطفا فایل pdf گزارش کار و فایل پیاده سازی کد و همچنین فایل حل سوالات تئوری را در قالب یک فایل zip در سامانه کورسز و کوئرا بارگذاری نمایید.
- در صورت هرگونه مشکل و یا ابهام می‌توانید با تدریسپاران خود در ارتباط باشید و اینکه به ایمیل درس سوالات خود را ارسال نمایید:

aut.ai.spring@gmail.com