



آزمایشگاه هوش مصنوعی کاربردی
دانشکده مهندسی مکانیک
دانشگاه تهران



دوره آشنایی با هوش مصنوعی

پروژه بخش یادگیری ماشین (قسمت دوم)

استاد مربوطه:

دکتر شریعت پناهی

مسئول آزمایشگاه:

دکتر کاشانی

دستیاران آموزشی:

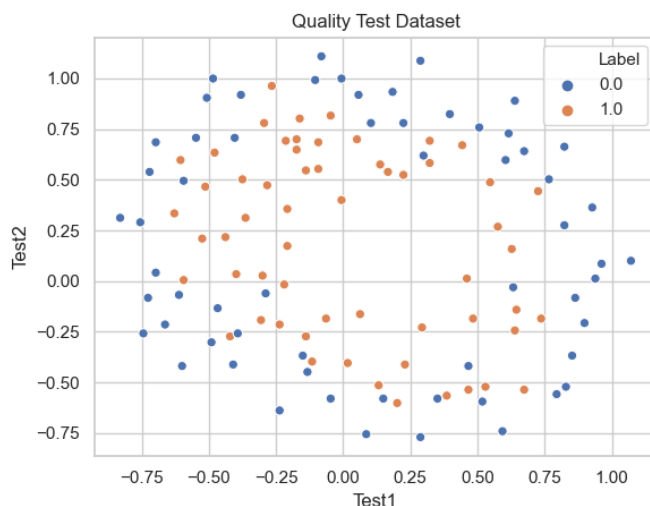
فاطمه مجاب

سعید دلیر

پاییز ۱۴۰۱

بخش دوم: دسته‌بندی قطعات به سالم و معیوب

در این بخش می‌خواهیم تراشه‌های تولید شده در یک خط تولید را بر پایه دو ویژگی آن‌ها به دو دسته سالم و معیوب دسته‌بندی کنیم. در مجموعه داده quality_test.csv دو ستون اول نشان دهنده نتایج تست تراشه و ستون سوم نشان‌دهنده کیفیت آن (قابل قبول یا مردود) است.



شکل ۱: نمایش مجموعه داده quality_test.csv

همانگونه که در شکل ۱ دیده می‌شود این مجموعه داده به صورت خطی جدایی‌پذیر نیست، به همین دلیل همانند توضیحات بخش قبل برای تفکیک دو کلاس باید فضای ویژگی‌ها را به مرتبه بالاتری برد.

مراحل اجرایی و خواسته‌ها:

- (۱) در مورد مفهوم Regularization و انواع آن تحقیق کنید و به اختصار توضیح دهید.
- (۲) فضای ویژگی‌های مجموعه داده را به درجه ۷ برده و با استفاده از الگوریتم Logistic Regression و L2 Regularization دو کلاس مختلف را از هم جدا کنید. برای پارامتر C (پارامتر Regularization) سه مقدار مختلف 1، 0.01 و 10000 در نظر بگیرید و در هر حالت دقت دسته‌بندی را گزارش کرده و مرز تصمیم‌گیری را رسم کنید.
- (۳) در مورد Cross Validation به اختصار توضیح دهید. با استفاده از روش K-fold CV درجه چندجمله‌ای را از بین اعداد ۳، ۷ و ۱۰ تعیین کنید.
- (۴) مجموعه داده اولیه (درجه ۱) را به دو بخش آموزش و تست تقسیم کنید. سپس با استفاده از روش kNN و برای مقادیر k برابر ۱ و ۷ و ۱۳ و ۱۹ دسته‌بندی را انجام دهید. برای هر مقدار k، پارامترهای Precision، Accuracy و Recall را روی داده‌های تست گزارش کرده و نتایج را تفسیر کنید.
- (۵) برای بهترین مقدار k از میان مقادیر بالا، نتیجه را با استفاده از فاصله منتهی نیز گزارش کنید.

توضیحات

- برای حل و پیاده‌سازی سوالات تنها استفاده از زبان برنامه‌نویسی پایتون مجاز است. همچنین شما مجاز به استفاده از کتابخانه‌های آماده مانند numpy, matplotlib, pandas و sklearn می‌باشید.
- تحویل گزارش برای این پروژه ضروری است. لزومی به توضیح جزئیات کدها نیست اما از آنجا که برای این پروژه از کتابخانه‌های آماده استفاده می‌شود، لطفاً تمامی پارامترهای تنظیم‌شده در هر قسمت از کد را گزارش کرده و فرض‌هایی را که برای پیاده‌سازی‌ها و محاسبات خود به کار برده‌اید ذکر کنید. ضمناً در فرایند ارزیابی کدهای شما لزوماً اجرا نخواهند شد، بنابراین همه‌ی نتایج و تحلیل‌های خود را به‌طور کامل ارائه کنید.
- گزارش شما باید به صورت تایپ شده و با فرمت pdf ارائه شود و کدهایی که به همراه گزارش تحویل می‌دهید باید قابل اجرا باشند. در انتها تمامی فایل‌های لازم را در یک فایل zip یا rar بارگذاری و برای دستیاران آموزشی ارسال کنید.
- پس از انجام و تحویل این تمرین، پاسخ خود را به دستیار آموزشی ارسال کرده و تمرین بخش بعدی را دریافت کنید.
- پرسش‌های خود را از طریق تلگرام از دستیاران آموزشی مربوطه بپرسید:

@saeed675

@SFatemehM