

Homework Week 13

سوالات تحلیلی

- ۱- موضوع Concept Drift را بررسی کنید و توضیح دهید که چیست؟ و بیشتر در چه زمینه هایی مورد مطالعه قرار میگیرد؟
- ۲- یک نوع از مدل‌هایی که برای بحث Concept Drift مورد استفاده قرار میگیرد، مدل های Ensemble هستند. توضیحی در مورد مدل‌های Ensemble ارائه دهید. در گام بعدی، ارتباط بین Ensemble و Catastrophic Forgetting را بررسی کنید. (موضوع Catastrophic Forgetting خارج از مطالب مباحث آموزش کاروندان هست، اما بدلیل اینکه یک موضوع مهم مورد بررسی در حوزه هوش مصنوعی هست، در حد بررسی کوتاه میتواند کمک کننده باشد. با توجه به اینکه اکثر مدل‌ها از forgetting یا weighting در مدلسازی خودشان استفاده میکنند، مدل‌ها در گذر زمان دچار مشکلات مرتبط با حافظه میشوند. این یک موضوع برجسته در حوزه علمی هست و نیاز هست که مدلساز در زمان ساخت مدل خودش این مهم را در نظر بگیرد.)
- ۳- در بحث Imbalance بودن داده‌ها، ابتدا بررسی کنید که چه زمانی این اتفاق می افتد، و اینکه راه های پیشنهاد شده برای مقابله با این موضوع چی هستند؟

Homework Week 13

سوالات برنامه نویسی

سوال 1

تحلیل ROC + Threshold + Imbalance با تمرکز روی تصمیم‌گیری

صورت سؤال

شما یک مسئله‌ی دسته‌بندی دودویی با داده‌های نامتوازن دارید. هدف، تحلیل رفتار مدل تحت تغییر threshold است، نه صرفاً گزارش Accuracy.

مراحل

بخش ۱ - ساخت داده

۱. با استفاده از numpy یک دیتاست مصنوعی بسازید:

- کلاس ۰: ۹۰۰ نمونه
 - کلاس ۱: ۱۰۰ نمونه
 - حداقل یک feature عددی که تا حدی قابلیت تفکیک داشته باشد
۲. داده را به train/test با نسبت ۳۰/۷۰ و stratify=y تقسیم کنید.

بخش ۲ - آموزش مدل پایه

۱. یک مدل LogisticRegression آموزش دهید.

۲. مقادیر زیر را روی test محاسبه کنید:

- Accuracy
- Precision
- Recall
- F1-score

Homework Week 13

Confusion Matrix ○

۳. خروجی‌ها را تفسیر کنید (به خصوص FN و FP).

بخش ۳ - تحلیل threshold

۱. به جای predict از predict_proba استفاده کنید.

۲. برای threshold های مختلف (مثلاً از ۰,۱ تا ۰,۹):

○ y_pred را به صورت دستی بسازید

○ Precision و Recall را محاسبه کنید

۳. نمودار Precision-Recall tradeoff رسم کنید.

بخش ۴ - ROC و AUC

۱. منحنی ROC را رسم کنید.

۲. مقدار AUC را گزارش دهید.

۳. توضیح دهید:

○ چرا AUC به threshold وابسته نیست؟

○ آیا ممکن است مدلی Accuracy بالاتر ولی AUC پایین‌تر داشته باشد؟

Homework Week 13

سوال ۲

SMOTE + Drift + ارزیابی صحیح مدل

صورت سؤال

در این تمرین بررسی می‌کنید که SMOTE همیشه راه‌حل نیست و ترکیب آن با drift و threshold چه اثری دارد.

بخش ۱ - داده‌ی اولیه

۱. یک دیتاست دودویی نامتوازن بسازید (مشابه تمرین قبل).
۲. یک مدل LogisticRegression یا RandomForest آموزش دهید.

۳. معیارهای زیر را گزارش دهید:

- Confusion Matrix
- Recall کلاس مثبت
- Balanced Accuracy
- ROC AUC

بخش ۲ - اعمال SMOTE

۱. فقط روی train set از SMOTE استفاده کنید.

۲. مدل را مجدداً آموزش دهید.

۳. همان معیارها را دوباره گزارش دهید.

۴. مقایسه‌ی عددی و مفهومی انجام دهید:

- چه چیزی بهتر شده؟
- چه چیزی بدتر شده؟

بخش ۳ - تغییر threshold بعد از SMOTE

Homework Week 13

۱. threshold را از ۰.۵ به مقدار دلخواه تغییر دهید.

۲. نشان دهید:

- چگونه می‌توان Recall را بالا برد
 - و هم‌زمان Precision را کنترل کرد
۳. نتیجه را با Confusion Matrix نشان دهید.

بخش ۴ - شبیه‌سازی Drift

۱. داده‌ی test را کمی تغییر دهید (مثلاً):

- شیفت میانگین یک feature
- یا تغییر واریانس

۲. بدون retrain:

- عملکرد مدل را دوباره ارزیابی کنید

۳. پاسخ دهید:

- این drift از چه نوعی است؟ (Feature / Concept)
- چرا SMOTE کمکی به این مشکل نمی‌کند؟