

**Quiz — 3**  
**MIR — Spring 2026**  
**Instructor: Dr. Amin Eskandari**  
**Head Teaching Assistants: Hamid Namjoo, Amir Hossein Hemati**  
**Teaching Assistants: Ali Mojahed, Ali Nikvan, Amir Javvi**

**سوال 1**

در یک سیستم بازیابی اطلاعات پزشکی، مجموعه تستی با مشخصات زیر داریم:

تعداد کل اسناد: (N) (نامعلوم)

تعداد اسناد مثبت (مرتبط): (P = 500)

تعداد اسناد منفی (نامرتبط): (N - P)

برای یک مدل، نتایج زیر گزارش شده است:

$$\text{Accuracy} = 0.95$$

$$\text{TPR} = \text{Recall} = 0.80$$

$$\text{TNR} = 0.96$$

$$\text{Precision} > 0.9$$

در عین حال، تیم ارزیابی ادعا می‌کند که:

$$\text{FPR} < \text{FNR}$$

الف) بررسی کنید آیا وجود مدلی با این مشخصات اصولاً ممکن است یا نه؟

باید یکی از دو کار زیر را انجام دهید (و هر دو نیازمند استدلال عمیق هستند، نه صرفاً چند محاسبه):

1. اگر ممکن است:

یک بازه‌ی معتبر برای (N) (تعداد کل نمونه‌ها) پیدا کنید که در آن تمام این متریک‌ها بتوانند به‌طور همزمان برقرار باشند

نشان دهید که برای حداقل یک مقدار مشخص از (N) می‌توان یک confusion matrix سازگار ساخت (نیازی نیست ماتریس دقیق بسازید، اما باید حدود TP, FP, FN, TN را منطقی و سازگار توصیف کنید)

2. اگر غیرممکن است:

دقیقاً نشان دهید تناقض از کجا می‌آید؟

باید از روابط بین  $TPR$ ,  $TNR$ ,  $Accuracy$  و نسبت کلاس‌ها استفاده کنید و نشان دهید شرط

$$FPR < FNR$$

با بقیه‌ی اعداد در تضاد است.

ب) فرض کنید شما فقط گزارش متریک‌ها را دارید و هیچ‌کس تعداد کل نمونه‌ها ( $N$ ) را در گزارش ننوشته است!

توضیح دهید چگونه می‌توان با استفاده از  $MCC$  (ضریب هم‌بستگی ماتیو) و ترکیب آن با متریک‌های بالا، یک تست «بررسی سازگاری» طراحی کرد که بتواند:

- یا تناقض متریک‌ها را آشکار کند،

- یا نشان دهد که گزارش متریک‌ها منطقی است

## سوال 2

دو سیستم رتبه‌بندی اسناد ( $A$  و  $B$ ) روی یک دیتاست یکسان ارزیابی شده‌اند

تعداد اسناد مثبت واقعی: ( $P$ ) (نامعلوم)،

تعداد کل اسناد: ( $N$ ) (نامعلوم)

گزارش متریک‌ها (واقعی ولی ناقص) به شکل زیر است:

(صفحه بعد)

مدل A

$$\text{Precision}_A = 0.95$$

$$\text{TPR}_A = 0.20$$

$$\text{FPR}_A = 0.02$$

مدل B

$$\text{Precision}_B = 0.60$$

$$\text{TPR}_B = 0.80$$

$$\text{FPR}_B = 0.30$$

در گزارش رسمی، فقط این را نوشته‌اند:

Accuracy هر دو مدل تقریباً برابر است

الف) بدون دانستن (P) و (N) بررسی کنید آیا جمله‌ی «Accuracy دو مدل تقریباً برابر است» می‌تواند برای بعضی نسبت‌های کلاس مثبت/منفی درست و برای بعضی دیگر غلط باشد یا نه؟  
باید:

1. نشان دهید که Accuracy هر مدل را می‌توان به صورت تابعی از نسبت کلاس مثبت

$$\Pi = P/N$$

نوشت؛

2. بررسی کنید آیا برای بعضی بازه‌های  $\Pi$  اختلاف Accuracy بین دو مدل می‌تواند بسیار کوچک (تقریباً برابر) باشد و برای بازه‌های دیگر به‌طور معنی‌دار از هم جدا شود؟

3. به صورت کیفی (نه لزوماً عددی دقیق) توضیح دهید در چه نوع توزیع کلاس (بسیار نامتوازن / نسبتاً متعادل) مدل A در Accuracy سود می‌برد و در چه حالتی مدل B؟

(ب)

مدیر پروژه می‌گوید:

چون هر دو مدل Accuracy مشابهی دارند، ما باید مدلی را انتخاب کنیم که MCC آن بیشتر است، چون MCC معیار عادلانه‌تری است، شما هنوز (P) و (N) را ندارید!

از شما خواسته شده تصمیم بگیرید در بازه‌های مختلف  $\pi = P/N$  کدام مدل «به‌طور ساختاری» می‌تواند MCC بالاتری داشته باشد، بدون آنکه مقدار دقیق MCC را محاسبه کنید.

کار شما:

1. یک تحلیل کیفی ارائه دهید که:

وقتی کلاس مثبت خیلی کم است کدام مدل از نظر MCC پتانسیل بیشتر/کمتر دارد؟ و چرا؟  
وقتی کلاس‌ها نسبتاً متعادل هستند، کدام مدل از نظر MCC احتمالاً برتری دارد.

2. توضیح دهید چرا در بعضی بازه‌های  $\pi$  ممکن است مدلی که Accuracy پایین‌تری دارد، MCC بالاتر داشته باشد؟ و این چه چیزی درباره‌ی ساختار خطاها (FP, FN) می‌گوید؟

سوال 3

یک سیستم فیلتر اسپم ایمیل داریم، نسبت کلاس مثبت (اسپم) در داده‌ی واقعی حدوداً:

$$\pi = 0.2$$

برنامه‌نویس سیستم، برای گزارش به مدیر، یک مدل به ظاهر عالی طراحی می‌کند که متریک‌های زیر را دارد:

$$\text{Accuracy} > 0.95$$

$$\text{Precision} > 0.9$$

$$\text{NPV} > 0.95$$

اما در عمل کاربران بسیار ناراضی‌اند و می‌گویند «ایمیل‌های مهم زیادی از دست می‌رود»

الف) نشان دهید که می‌توان یک خانواده‌ای از confusion matrix ها طراحی کرد که:

- تمام متریک‌های بالا را برآورده کند،

- اما در آن‌ها FNR بسیار بالا باشد (یعنی بسیاری از ایمیل‌های مثبت [اسپم واقعی یا هر معنای مثبت که تعریف می‌کنید] نادیده گرفته شوند)

وظیفه شما:

1. بدون دادن یک ماتریس خاص، ساختار کیفی این خانواده را توصیف کنید، مثلاً:

- نسبت TP بسیار پایین، FN بالا، TN بسیار بالا، FP متوسط/کم؛

- توضیح دهید چرا با این ساختار، هم Precision و هم NPV می‌توانند بالا بمانند اما FNR فاجعه‌بار شود؟

2. سپس با استفاده از MCC توضیح دهید چرا احتمالاً MCC در چنین سناریوهایی پایین خواهد بود، حتی اگر Accuracy و Precision و NPV بالا باشند؟

ب) حالا فرض کنید شما مسئول طراحی یک معیار ارزیابی ساده (بدون فرمول پیچیده)، بر اساس همین متریک‌ها هستید تا جلوی چنین تقلب‌هایی را بگیرید.

هدف شما:

- یک تابع کیفی (نه لزوماً دقیقی) از متریک‌های زیر پیشنهاد کنید:  
TPR, TNR, FPR, FNR, Precision, NPV, Accuracy, MCC
- به‌طوری‌که مدل‌های تقلبی مثل سناریوی بخش (الف) امتیاز پایینی بگیرند،
- اما مدل‌های واقعاً متعادل (از نظر رفتار روی هر دو کلاس) امتیاز خوبی بگیرند

کار شما:

1. پیشنهاد دهید کدام متریک‌ها باید حتماً در این تابع حضور پررنگ داشته باشند و کدام می‌توانند وزن کم‌تری داشته باشند؛

2. با استدلال مفهومی (بدون عدد) توضیح دهید که این تابع چگونه مدل‌های با:

- FNR بالا

- یا FPR بسیار بالا

- یا MCC پایین

را به‌شدت جریمه می‌کند، حتی اگر Accuracy و Precision ظاهراً خوب باشند؟!   
 \_\_\_\_\_   
 \_\_\_\_\_   
 \_\_\_\_\_

**سوال 4 -** (سوال جایزه (از کتاب روحانی رانکوهی + همون جلسه "دیتا چیست؟") ← جایزه‌ش اینه که باید جواب بدید 😊

فایلی با مشخصات زیر داریم:

طول رکورد  $R = 150$  بایت، اندازهٔ بلاک  $B = 700$  بایت، شمارهٔ اولین بلاک فایل  $RBA_{BOF} = 8$ .  
برنامه کاربر با دستور SEEK به رکورد شماره ۳۷ (شماره‌گذاری از صفر است). مراجعه می‌کند و سپس READ می‌نماید.

الف) ضریب بلاک‌بندی ( $B_f$ ) و RBA بلاک حاوی رکورد ۳۷ را محاسبه کنید.  
ب) اگر رکورد ۳۷ دقیقاً از یک بلاک به بلاک بعدی کشیده شود (یعنی بخشی از آن در بلاک اول و بخشی در بلاک دوم قرار گیرد، سیستم فایل برای بازگرداندن کامل این رکورد چند عمل I/O (خواندن بلاک) نیاز دارد؟ چرا؟

ج) با استفاده از جدول عملیات کاربر، توضیح دهید که برنامه‌نویس برای خواندن رکورد بعدی (رکورد ۳۸) چه عملیاتی را باید صادر کند و سیستم فایل منطقی چه مراحل را (تا پیش از I/O) طی می‌کند.

د) به عنوان حسن ختام، بفرمایید بگید که دیتا چی بود؟ (کلی بگید، نمره سوال می‌پره!) 4 ویژگی دیتا را شرح دهید؟ اسم فایل دیتاس؟ خود فایل چی؟ به عبارت RE (regular expression) می‌تواند دیتا شود؟ نمی‌تواند؟!!! چرا؟

---

---

---

---

موفق و موید: )