



دکتر طاهائی
دانشکده ریاضی و علوم کامپیوتر

مقدمه ای بر داده کاوی
نیمسال اول ۱۴۰۳-۱۴۰۴
گردآورندگان: نرگس قنبری- سارا پورخلیل

تمرین اول: طبقه بندی بیز ساده

سوال ۱

در یک سیستم تشخیص بیماری، سه ویژگی مستقل برای تعیین اینکه آیا فرد به بیماری "X" مبتلا است یا نه استفاده می شود:

• وجود تب (Fever)

• وجود سرفه (Cough)

• وجود سردرد (Headache)

احتمالات زیر از داده های جمع آوری شده به دست آمده است:

$$P(X \text{ بیماری}) = 0.3,$$

$$P(X \text{ عدم بیماری}) = 0.7,$$

$$P(X \text{ بیماری} | \text{تب}) = 0.8,$$

$$P(X \text{ عدم بیماری} | \text{تب}) = 0.2,$$

$$P(X \text{ بیماری} | \text{سرفه}) = 0.6,$$

$$P(X \text{ عدم بیماری} | \text{سرفه}) = 0.4,$$

$$P(X \text{ بیماری} | \text{سردرد}) = 0.7,$$

$$P(X \text{ عدم بیماری} | \text{سردرد}) = 0.3.$$

فرض کنید فردی به مطب پزشک مراجعه می کند و این سه علامت را دارد: تب، سرفه، و سردرد.

۱. با استفاده از طبقه بندی بیز ساده، احتمال اینکه فرد مبتلا به بیماری X باشد را محاسبه کنید.

۲. با توجه به محاسبات، تعیین کنید که آیا فرد مبتلا به بیماری X است یا خیر.

۳. توضیح دهید چرا فرض بیز ساده مبنی بر استقلال شرطی ممکن است در تشخیص بیماری ها در دنیای واقعی صحیح نباشد.

۴. اگر فرض کنیم ویژگی ها (تب، سرفه، سردرد) همبستگی دارند (مثلاً تب و سرفه معمولاً با هم اتفاق می افتند)، چگونه می توانیم مدل بیز ساده را اصلاح کنیم تا این همبستگی را در نظر بگیریم؟ آن را به اختصار توضیح دهید.



دکتر طاهائی
دانشکده ریاضی و علوم کامپیوتر

مقدمه ای بر داده کاوی
نیمسال اول ۱۴۰۳-۱۴۰۴
گردآورندگان: نرگس قنبری- سارا پورخلیل

Name	GPA	Effort	Hirable
Sarah	poor	lots	Yes
Dana	average	some	No
Alex	average	some	No
Annie	average	lots	Yes
Emily	excellent	lots	Yes
Pete	excellent	lots	No
John	excellent	lots	No
Kathy	poor	some	No

سوال ۲

داده‌های فرضی بالا را در مورد ویژگی‌های دانش‌آموز و اینکه آیا هر دانشجو باید استخدام شود یا خیر، در نظر بگیرید.

- از *bayes naive* استفاده کنید تا تعیین کنید که آیا فردی با حضور عالی، معدل ضعیف و تلاش زیاد باید استخدام شود یا خیر.

سوال ۳

دسته‌بندی متن کاربردی است که در آن برای یک متن ارائه شده (یک document)، دسته‌ای خاص از بین دسته‌های از پیش تعیین شده مشخص می‌شود. مدل بیز ساده انگارانه (Bayes Naive) ابزاری مناسب برای این کاربرد است و نوعاً از آن برای دسته‌بندی متن استفاده می‌شود.

- به طور دقیق تشریح کنید که مدل ساده‌انگارانه برای این مسئله چگونه ساخته می‌شود. فرض کنید مجموعه‌ای از متن‌های مختلف در اختیار داریم که دسته هر متن نیز مشخص شده است (مجموعه آموزشی).
- به طور دقیق تشریح کنید که برای یک متن جدید، چگونه می‌توان دسته آن را تشخیص داد.
- آیا فرض استقلال شرطی که مدل بیز ساده انگارانه بر آن استوار است، فرض منطقی‌ای در این مسئله است؟ توضیح دهید.