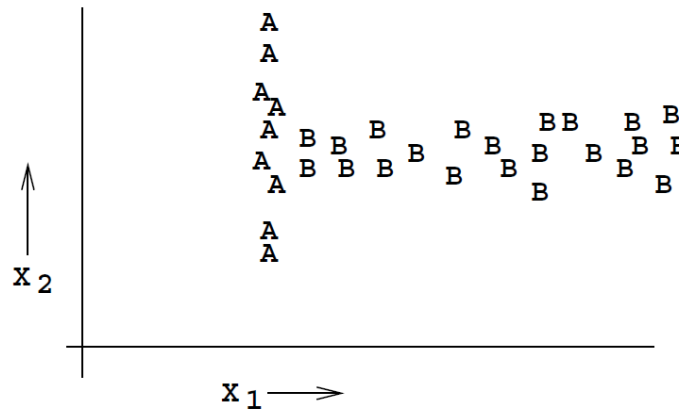


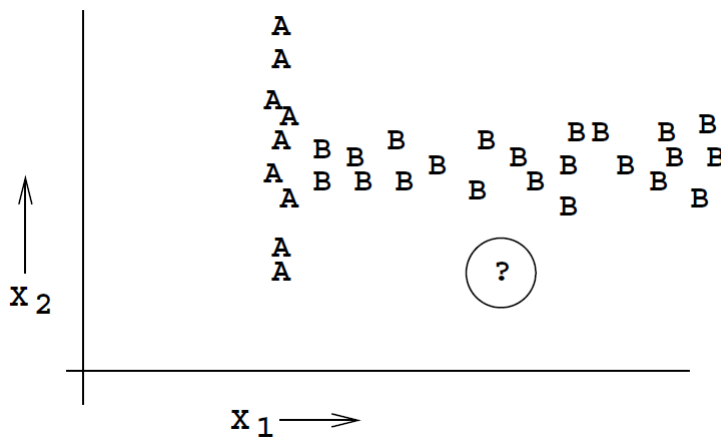
سوال ۱

با استفاده از داده‌های زیر (که دارای دو ویژگی با مقادیر حقیقی X_1 و X_2 هستند) یک دسته‌بند بیز ساده گاوسی را آموزش داده‌ایم.

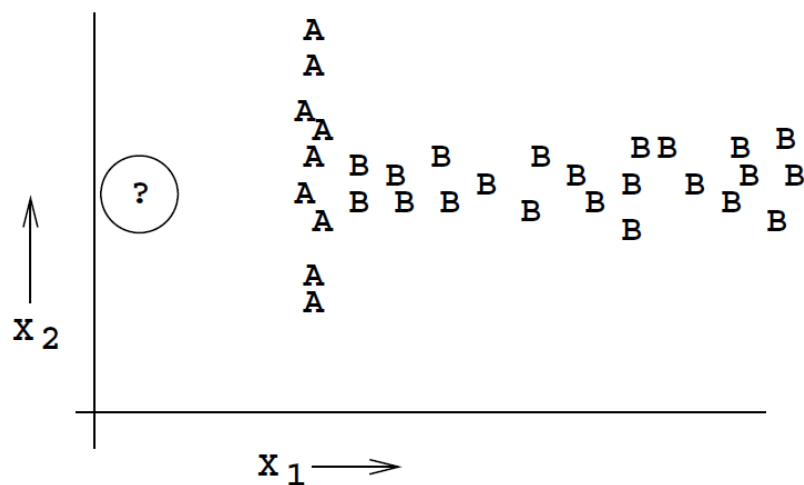


با ذکر دلیل برای هر یک از حالات زیر تعیین کنید که برچسب داده‌ی تست (که با علامت سوال مشخص شده است) چه خواهد بود.

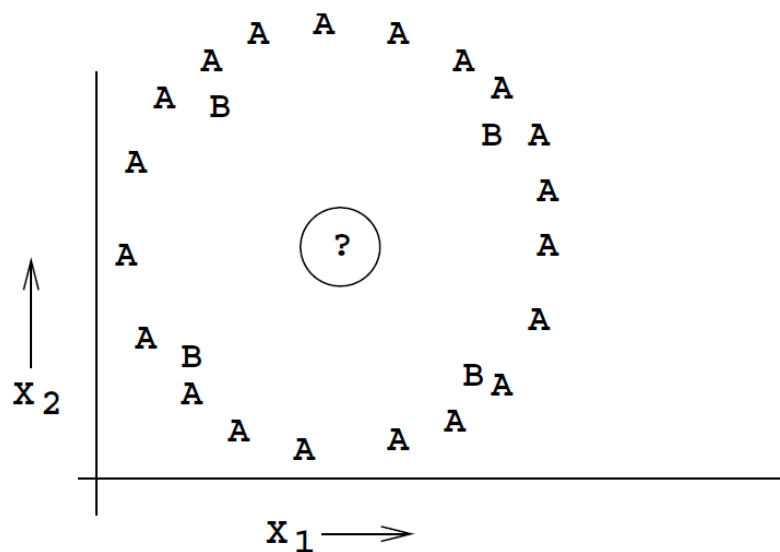
(الف)



(ب)



ج) فرض کنید با مجموعه‌ی داده‌ی دیگری که در ذیل آمده است یک دسته‌بند جدید را آموزش داده باشیم. برچسب داده‌ی تست در این حالت چه خواهد بود و چرا؟



سوال ۲

با استفاده از مدل‌های generative می‌خواهیم یک مسئله دسته‌بندی را حل نماییم. در نظر بگیرید که در این مسئله داده‌ها دارای n بعد بوده و تعداد دسته‌ها برابر m می‌باشد. اگر ویژگی‌ها باینری باشند، تعداد پارامترهای مورد نیاز برای آموزش یک دسته‌بند را در حالات زیر بدست آورده و با یکدیگر مقایسه نمایید.

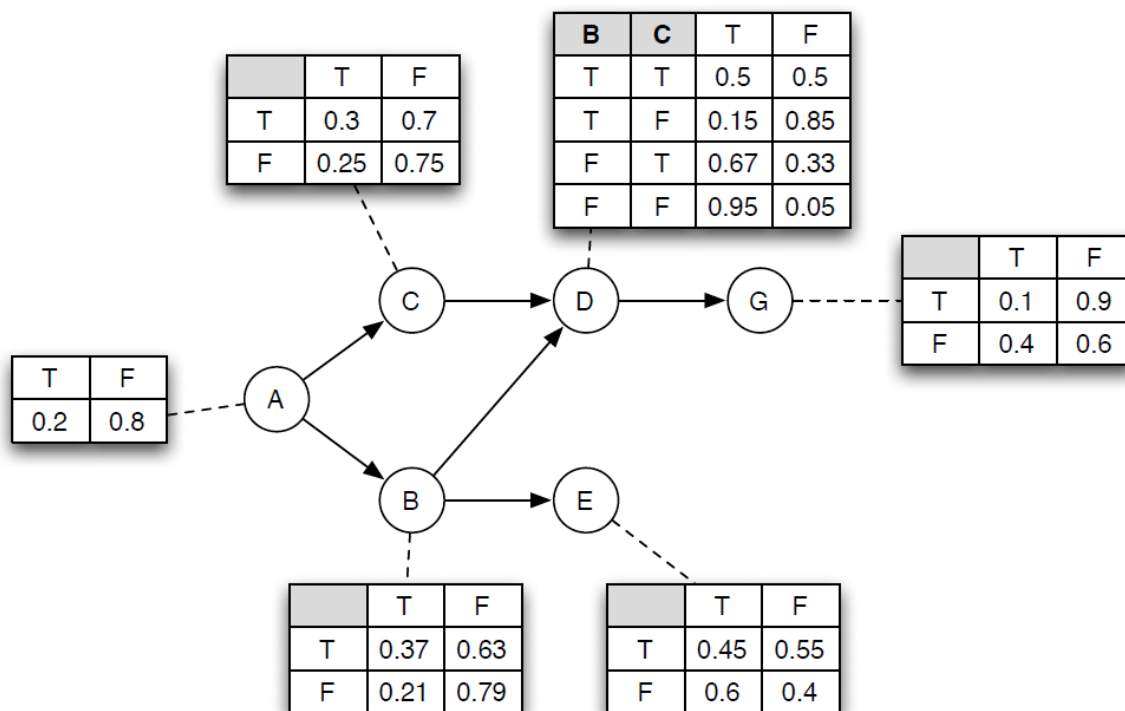
الف) هیچ استقلالی روی ویژگی‌ها وجود نداشته باشد.

ب) ویژگی‌ها با دانستن دسته، از یکدیگر مستقل باشند.

ج) ویژگی‌ها باینری نبوده و دارای مقادیر حقیقی هستند اما نسبت به یکدیگر مستقل باشند.

سوال ۳

احتمال $P(B|D=T)$ در شبکه‌ی بیزین زیر محاسبه نمایید.



سوال ۴

تفاوت مدل‌های Generative و Discriminative در چیست؟

سوال ۵

با استفاده از مراجع [۱] و [۲] و سایر منابع، دسته بند بیز ساده و Logistic Regression را با یکدیگر مقایسه نمایید. (حداکثر یک صفحه)

پیاده سازی

تعریف مسئله: فرض کنید یک شبکه‌ی اجتماعی داریم که در آن افراد می‌توانند به محتوای موجود روی این شبکه (نظیر اخبار) رای دهند. تعداد ۱۵۹۳ خبر مختلف از روی این شبکه به منظور بررسی جهت‌گیری مختلف افراد جمع آوری شده است. هر یک از افراد داخل این شبکه می‌تواند برای هر خبر ارائه شده با یک عدد گسسته امتیازدهی انجام دهد. در صورتی که فرد خبر را مشاهده نکرده باشد یا مشاهده کرده ولی امتیازی نداده باشد، مقدار صفر به عنوان امتیاز فرد به آن خبر در نظر گرفته می‌شود. کلیه افراد شبکه به دو دسته قابل تفکیک هستند. شرکت X برای پخش محتوای خود روی شبکه قصد دارد تا بنابر دسته‌ای که یک عضو از شبکه در آن قرار دارد اخبار متناسب با علایق و نظرات وی به او ارائه نماید. برای این منظور این شرکت با استفاده از اطلاعات و امتیازدهی‌های موجود به دنبال شناسایی دسته‌ی متناظر با یک فرد می‌باشد.

برای حل این مساله می‌خواهیم از دسته‌بندهای موجود استفاده کرده و عملکرد آن‌ها را روی این مجموعه داده بررسی نماییم تا به یک مدل مناسب دست پیدا کنیم. (به دلیل کوچک شدن پارامترها می‌توانید از لگاریتم احتمالات استفاده نمایید. همچنین فایل داده‌های آموزش و تست این بخش ضمیمه فایل تمرین شده است)

Logistic Regression

الف) دسته‌بند Logistic Regression را برای این مسئله پیاده سازی نمایید.

ب) تاثیر جمله‌ی $\frac{1}{2}||w||_2^2$ Regularization را روی این دسته بند بررسی نمایید. برای بدست آوردن λ مناسب از k -fold cross validation با مقدار $k=10$ استفاده نمایید ($\lambda \in \{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2\}$).

ج) برای هر یک از موارد بالا خطای آموزش و تست دسته‌بند را گزارش کرده و نمودار ROC مدل‌های بدست آمده را رسم نمایید.

Naïve Bayes

الف) می‌خواهیم از مدل بیز ساده برای مسئله بالا استفاده کنیم. برای این منظور بایستی فرض نماییم که امتیاز یک فرد به یک خبر مستقل از امتیاز وی به سایر اخبار می‌باشد. این دسته‌بند را پیاده سازی کرده و خطای آموزش و تست را گزارش نمایید.

ب) نمودار ROC مدل بدست آمده را رسم نمایید.

Bayes Network

مجموعه داده‌ی Cars را از لینک زیر دریافت کنید.

<http://archive.ics.uci.edu/ml/datasets/Car+Evaluation>

هدف این مسئله آموزش مدلی برای پیش‌بینی مقبولیت خودرو با توجه به ویژگی‌های موجود است. برای این منظور در نظر داریم از شبکه‌ی بیز استفاده کنیم. در این قسمت مجاز به استفاده از کتابخانه‌های موجود می‌باشید (به عنوان مثال [pgmpy](#) برای پایتون). برای گزارش دقت مدل از 10-fold Cross Validation استفاده کنید.

الف) ابتدا مدل بیز ساده را بر روی مجموعه داده آموزش داده و دقت آن را گزارش کنید. برای حداقل دو زیرمجموعه‌ی مختلف و دلخواه از ویژگی‌ها این فرآیند را تکرار کنید.

ب) شبکه‌ی بیزی با ساختارهای مختلف (حداقل ۳ ساختار) را آموزش داده و دقت و گراف مدل‌ها را گزارش کنید. دلایل ارائه‌ی هر ساختار را نیز به طور خلاصه توضیح دهید.

توضیحات تمرین:

- شما باید سورس کد خود به همراه گزارش (پاسخ سؤالات و تحلیل نتایج پیاده‌سازی‌ها) را در قالب یک فایل Zip با نام فایل `xxxxxx_HW3` که `xxxxxx` شماره دانشجویی شما است، تا تاریخ **۱۷ آذر** در سایت درس بارگذاری کنید.
- در کلیه موارد بخش پیاده‌سازی بایستی نتایج خود را تفسیر نمایید. درصد زیادی از امتیاز این بخش مربوط به تفسیر و توضیح نتایج است.
- پیاده‌سازی با متلب یا پایتون می‌تواند انجام شود.
- به غیر از موارد ذکر شده، مجاز به استفاده از هیچ کتابخانه آماده‌ای نیستید.
- در صورت نیاز می‌توانید سؤالات خود را از طریق ایمیل `ceitml17@gmail.com` مطرح نمایید.

[1] Jordan, A. (2002). On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. *Advances in neural information processing systems*, 14, 841.

[2] <https://www.cs.cmu.edu/~tom/mlbook/NBayesLogReg.pdf>