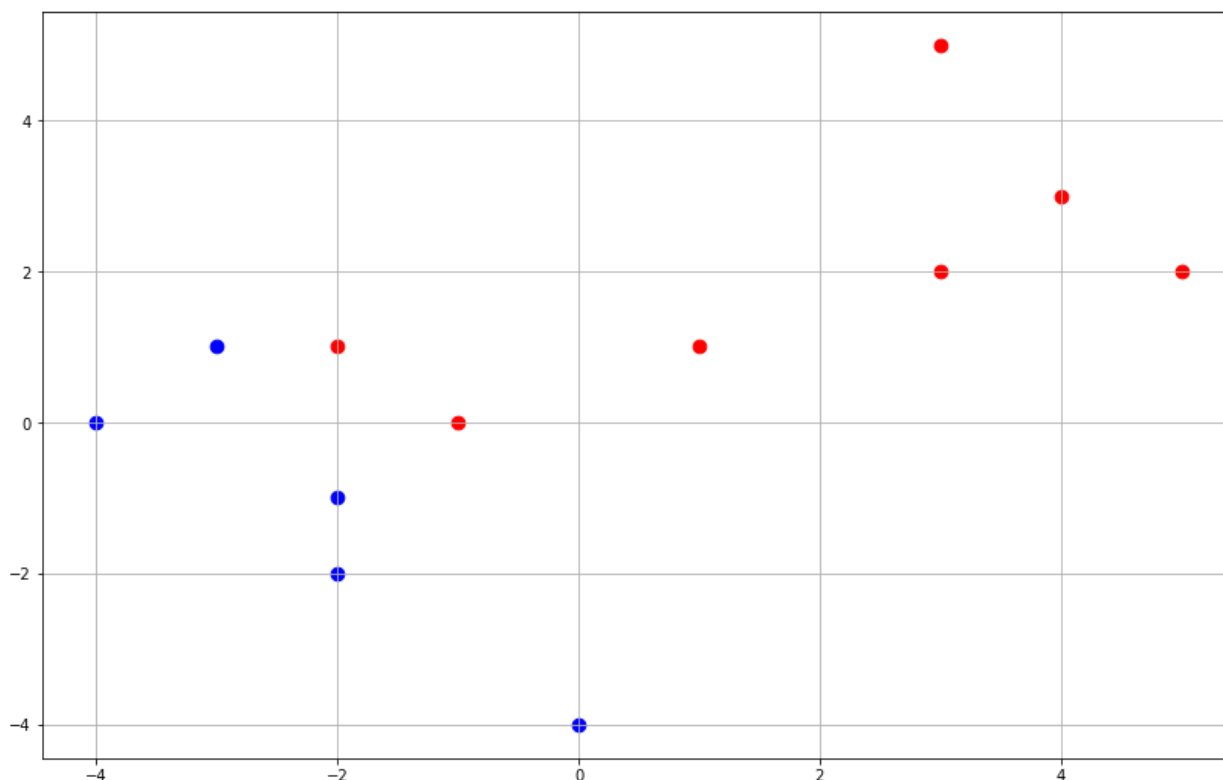




سلام بر تمام دانشجویان عزیز، چند نکته مهم:

1. حجم گزارش به هیچ عنوان معیار نمره‌دهی نیست، در حد نیاز توضیح دهید.
2. نکته‌ی مهم در گزارش نویسی روشن بودن پاسخها می‌باشد، اگر فرضی برای حل سوال استفاده می‌کنید حتماً آن را ذکر کنید، اگر جواب نهایی عددی است به صورت واضح آن را بیان کنید.
3. برای سوالات شبیه سازی، فقط از دیتاست داده شده استفاده از کنید. شکل ها به طور واضح و در فرمت درست گزارش شوند.
4. از بین سوالات **شبیه سازی** حتماً به هر دو مورد پاسخ داده شود. حداکثر تا نمره ۱۱۰ (۱۰ نمره امتیازی) لحاظ خواهد شد.
5. هرگونه شباهت در گزارش و کد مربوط به شبیه سازی، به منزله **تقلب** می باشد و کل نمره تمرین **صفر** می‌شود.
6. در صورت داشتن سوال، از طریق ایمیل farbodmoosavi@ut.ac.ir سوال خود را مطرح کنید.

1. فرض کنید دو کلاس داده A و B به صورت شکل زیر تقسیم شده اند. داده‌های کلاس A با رنگ قرمز و داده‌های دسته B با رنگ آبی مشخص شده اند. با فرض توزیع گاوسی برای هر کلاس، معادله مرز تصمیم مابین این دو کلاس را محاسبه کنید. تمامی مقادیر لازم (میانگین، احتمال پیشین و ...) را از مشاهدات براساس نمودار بدست بیاورید. همچنین مرز بدست آمده از محاسبات را بر روی شکل رسم کنید و درستی جواب خود را بررسی کنید. (15 نمره)



2. فرض کنید چگالی احتمال برای داده های دو کلاس مختلف به صورت زیر باشد:

$$p(x|\omega_i) = \frac{1}{\pi b} \cdot \frac{1}{1 + \left(\frac{x-a_i}{b}\right)^2}, \quad i = 1, 2.$$

فرض کنید که احتمال پیشین دو کلاس با هم برابر می باشد. (20 نمره)

الف) نشان دهید که حداقل احتمال خطا به صورت رابطه زیر بدست می آید:

$$P(\text{error}) = \frac{1}{2} - \frac{1}{\pi} \tan^{-1} \left| \frac{a_2 - a_1}{2b} \right|.$$

ب) تابع بدست آمده را به صورت تابعی از $\frac{|a_2 - a_1|}{b}$ رسم کنید.

ج) بیشترین مقدار خطا $P(\text{error})$ چه مقدار می باشد و تحت چه ظرایفی این اتفاق رخ می دهد؟ توضیح دهید.

3. در این سوال می خواهیم برخی تخمینگر های MLE را بررسی کنیم. (20 نمره)

الف) یک تاس m وجهی را n بار پرتاب میکنیم. در صورتی که هر وجه X_i بار ظاهر شود، تخمینگر MLE از p_i ها که احتمال آمدن هر وجه می باشد را بیابید.

یادآوری: X_i ها دارای توزیع چندجمله ای به شکل زیر می باشند:

$$f(x_1, \dots, x_m | p_1, \dots, p_m) = \frac{n!}{\prod x_i!} \prod p_i^{x_i}$$

در بیشینه سازی تابع بالا به شرط $\sum P_i = 1$ دقت کنید.

ب) متغیر تصادفی X دارای توزیع $\text{Unif}([0, \theta])$ می باشد. با فرض مشاهده n نمونه iid از این توزیع، تخمینگر MLE از θ را بیابید.

ج) توزیع تخمینگر قسمت ب را بیابید. بررسی کنید که این تخمینگر unbiased است یا نه.

د) n نمونه iid از یک توزیع پواسون با پارامتر λ داریم. تخمینگر MLE این پارامتر را بدست آورید.

4. فرض کنید تعدادی داده از یک توزیع گاوسی با کوواریانس Σ که مشخص است و میانگین μ که اطلاعی از آن نداریم در

دست داریم. حال فرض کنید که این میانگین خود یک متغیر تصادفی باشد و از یک توزیع گاوسی با میانگین m_0 و کوواریانس Σ_0 باشد. (20 نمره)

الف) تخمین MAP برای μ را بیابید.

ب) فرض کنید که مختصات فضا را توسط یک تبدیل خطی به صورت $X' = AX$ ، که در آن A یک ماتریس غیرتکین است. بررسی کنید که آیا تخمین بدست آمده در قسمت الف، تخمین مناسبی از μ' نیز می باشد یا نه.

5. (شبیه سازی) در این سوال، طبقه بندی طراحی کنید که بتوانیم، که ۲ کلاس متفاوت (دو تیم فوتبال منچستر یونایتد و چلسی) با استفاده از دیتاست داده شده، را تشخیص دهیم. جهت طبقه بندی، میتوانید میانگین رنگ در هر عکس را محاسبه نمایید، سپس بر اساس مقدار به دست آمده، با مقدار رنگ آبی و قرمز مقایسه نمایید. برای دیتاست داده شده، این طبقه بند را تست کنید. ماتریس Confusion را گزارش دهید. مقادیر accuracy, precision, recall را محاسبه کنید، و نتایج هر کدام را توضیح دهید. (15 نمره)

6. (شبیه سازی) در این سوال قصد داریم دیتاست Iris را بررسی کنیم. در این دیتاست سه ویژگی برای هر کلاس داده شده است و در کل داده ها به سه کلاس تقسیم می شوند. (20 نمره)

الف) ابتدا توضیح مختصری راجع به طبقه بندهای naïve bayes و optimal bayes داده و آن ها را با فرض گاوسی بودن داده های ورودی پیاده سازی کنید. دقت کنید که الگوریتم شما نباید هیچ پیش فرضی بر از داده های ورودی داشته باشد و باید برای هر داده ای کار کند. (در این قسمت مجاز به استفاده از کتابخانه های آماده نیستید و باید الگوریتم را خودتان پیاده سازی کنید).

ب) الگوریتم های پیاده سازی شده را برای مجموعه داده Iris تست کنید و نتایج به دست آمده (Confusion matrix, دقت طبقه بند) را برای هر روش گزارش کنید و در مورد نتایج بحث کنید.

ج) در این قسمت با استفاده از پکیج های آماده پایتون فقط الگوریتم naïve bayes را با فرض گاوسی بودن اجرا کنید و نتیجه را گزارش کنید.
