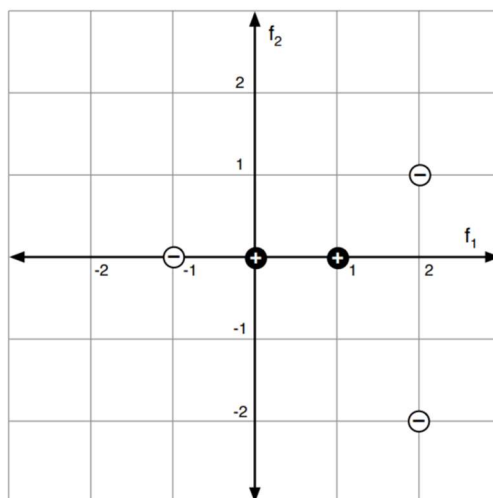


۱. [درخت تصمیم] نقاط داده عبارت‌اند از:

منفی:  $\begin{bmatrix} 2 \\ -2 \end{bmatrix}$ ,  $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ ,  $\begin{bmatrix} -1 \\ 0 \end{bmatrix}$

مثبت:  $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ ,  $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$



الف) با استفاده از الگوریتم C4.5، یک درخت تصمیم برای داده‌های بالا بسازید.

ب) مرزهای تصمیم را روی نمودار بالای صفحه رسم کنید.

ج) درخت تصمیم برای نقطه جدید  $\begin{bmatrix} 1 \\ -1.01 \end{bmatrix}$  چه کلاسی را پیش‌بینی می‌کند؟

می‌توانید از جدول زیر برای تسریع محاسبات استفاده کنید:

x	y	$-(x/y) * \lg(x/y)$	x	y	$-(x/y) * \lg(x/y)$
1	2	0.50	1	5	0.46
1	3	0.53	2	5	0.53
2	3	0.39	3	5	0.44
1	4	0.50	4	5	0.26
3	4	0.31			

۲. [نزدیک‌ترین همسایه] مرزهای تصمیم برای ۱-نزدیک‌ترین همسایه (1-Nearest Neighbors) را روی نمودار بالا (همان داده‌ها) رسم کنید. رسم شما باید به اندازه‌ای دقیق باشد که بتوان تشخیص داد نقاط با مختصات عدد صحیح روی مرز قرار دارند یا خیر، و اگر نه، در کدام ناحیه هستند.

الف) ۱-نزدیک‌ترین همسایه برای برای نقطه جدید  $\begin{bmatrix} 1 \\ -1.01 \end{bmatrix}$  چه کلاسی را پیش‌بینی می‌کند؟

ب) ۳-نزدیک‌ترین همسایه برای برای نقطه جدید  $\begin{bmatrix} 1 \\ -1.01 \end{bmatrix}$  چه کلاسی را پیش‌بینی می‌کند؟

۳. [بیز ساده لوحانه] یک مسئله Naive Bayes با سه ویژگی  $x_1$  تا  $x_3$  را در نظر بگیرید. فرض کنید که ما در مجموع ۱۲ نمونه آموزشی مشاهده کرده‌ایم: ۶ نمونه مثبت و ۶ نمونه منفی. در اینجا نقاط واقعی آورده شده‌اند:

الف) احتمالات زیر را تقریب بزنید و از تصحیح لاپلاسی (Laplacian Correction) استفاده کنید:

$x_1$	$x_2$	$x_3$	$y$
0	1	1	0
1	0	0	0
0	1	1	0
1	1	0	0
0	0	1	0
1	0	0	0
1	0	1	1
0	1	0	1
1	1	1	1
0	0	0	1
0	1	0	1
1	0	1	1

•  $\Pr(x_1 = 1 | y = 0)$

•  $\Pr(x_2 = 1 | y = 1)$

•  $\Pr(x_3 = 0 | y = 0)$

ب) برای داده (1,1,1) چه برجستگی پیش بینی می کنید؟

ج) الگوریتم naïve bayes برای این دادگان عملکرد خوبی ندارد. چرا؟

۴. [SVM] داده های کلاس های مثبت و منفی به ترتیب زیر هستند:

Negative: (-1, -1) (2, 1) (2, -1)

Positive: (-2, 1) (-1, 1)

الف) معادله یک جداکننده خطی را که دارای بیشترین حاشیه هندسی برای داده های بالا است، ارائه دهید.

ب) بردارهای پشتیبان (support vectors) این جداکننده کدامند؟