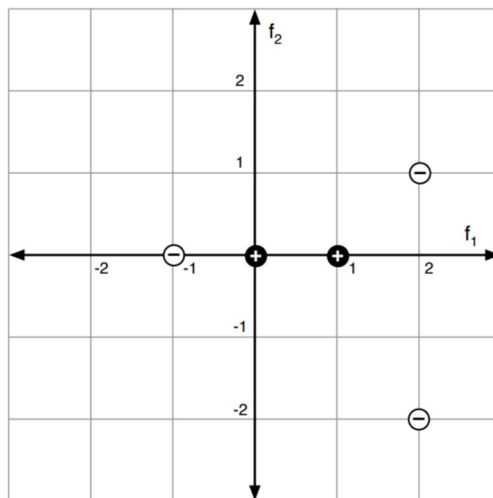
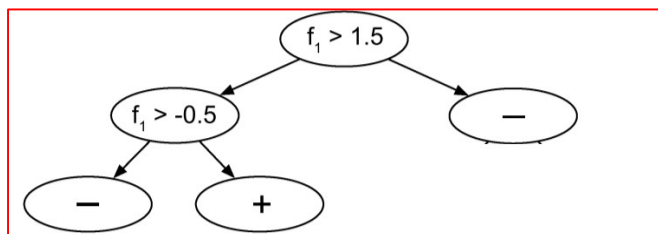


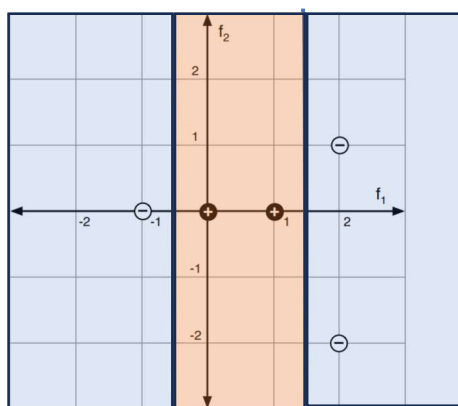
۱. [درخت تصمیم] نقاط داده عبارت‌اند از:

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

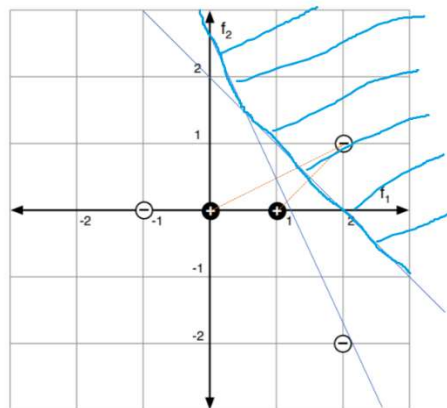
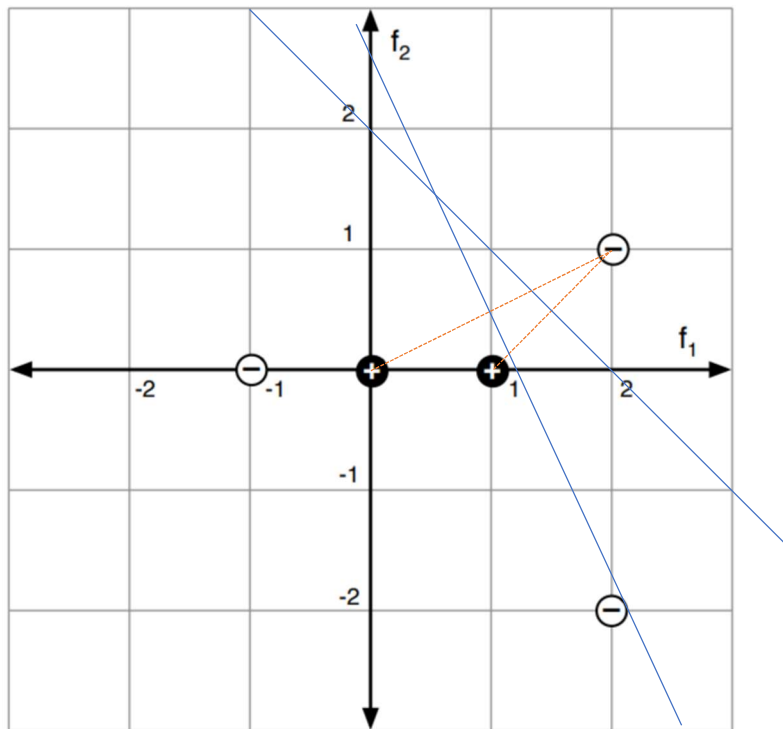
الف) با استفاده از الگوریتم C4.5، یک درخت تصمیم برای داده‌های بالا بسازید.

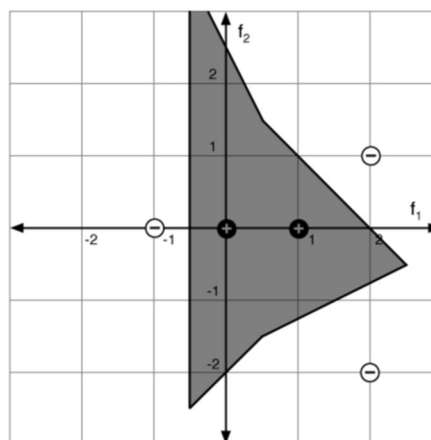
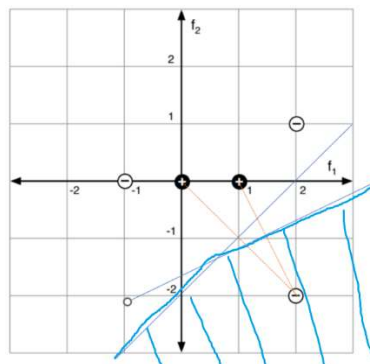
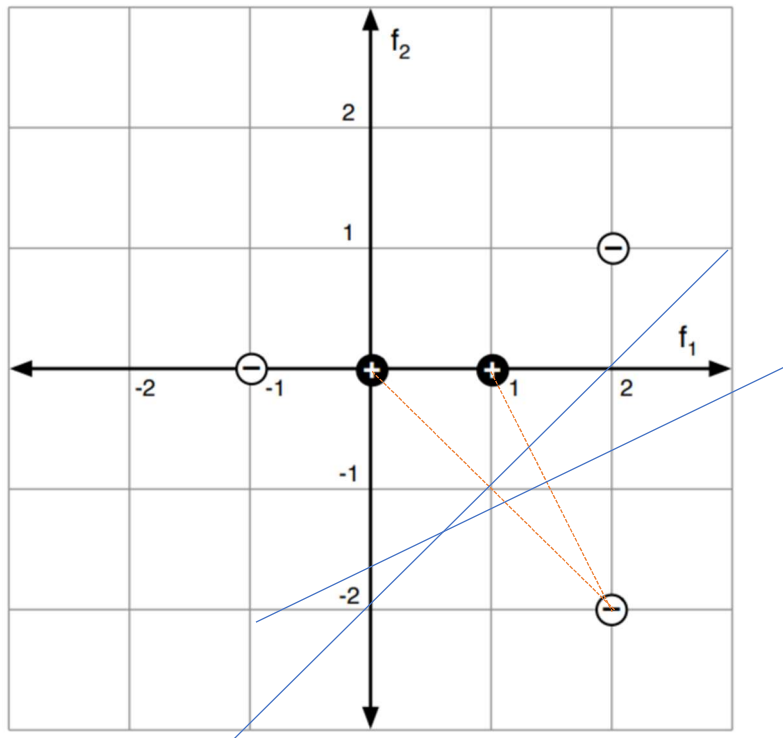


ب) مرزهای تصمیم را روی نمودار بالای صفحه رسم کنید.

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

۲. [نزدیک ترین همسایه] مرزهای تصمیم برای ۱-نزدیکترین همسایه (1-Nearest Neighbors) را روی نمودار بالا (همان داده ها) 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- نزدیک ترین همسایه برای برای نقطه جدید $\begin{bmatrix} 1 \\ -1.01 \end{bmatrix}$ چه کلاسی را پیش بینی می کند؟ مثبت

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:

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

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

$$\Pr(x_1 = 1 | y = 0) = \frac{3+1}{6+2} \quad \bullet$$

$$\Pr(x_2 = 1 | y = 1) = \frac{3+1}{6+2} \quad \bullet$$

$$\Pr(x_3 = 0 | y = 0) = \frac{3+1}{6+2} \quad \bullet$$

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

دلخواه چون

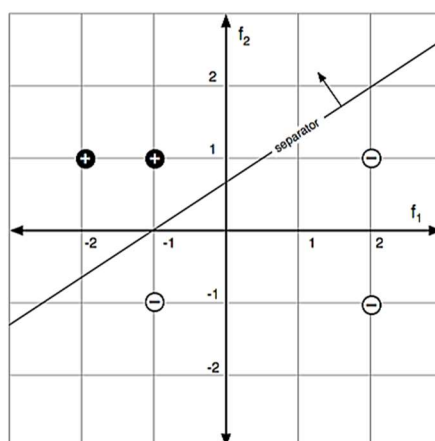
$$\Pr(x_1 = 1, x_2 = 1, x_3 = 1 | y = 1) = \Pr(x_1 = 1, x_2 = 1, x_3 = 1 | y = 0) \\ \propto \frac{4}{8} * \frac{4}{8} * \frac{4}{8}$$

ج) الگوریتم naïve bayes برای این دادگان عملکرد خوبی ندارد. چرا؟ فرض استقلال شرطی برقرار نیست

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 داده‌های بالا است، ارائه دهید.

$$-w_1 - w_2 + b = -1$$

$$2w_1 + w_2 + b = -1$$

$$-w_1 + w_2 + b = 1$$

از حل این دستگاه خواهیم داشت: $w_1 = -2/3$ و $w_2 = 1$ و $b = -2/3$

خط جداکننده: $-2x_1 + 3x_2 - 2 = 0$

ب) بردارهای پشتیبان (support vectors) این جداکننده کدامند؟ $(-1,-1), (-1,1), (2,1)$