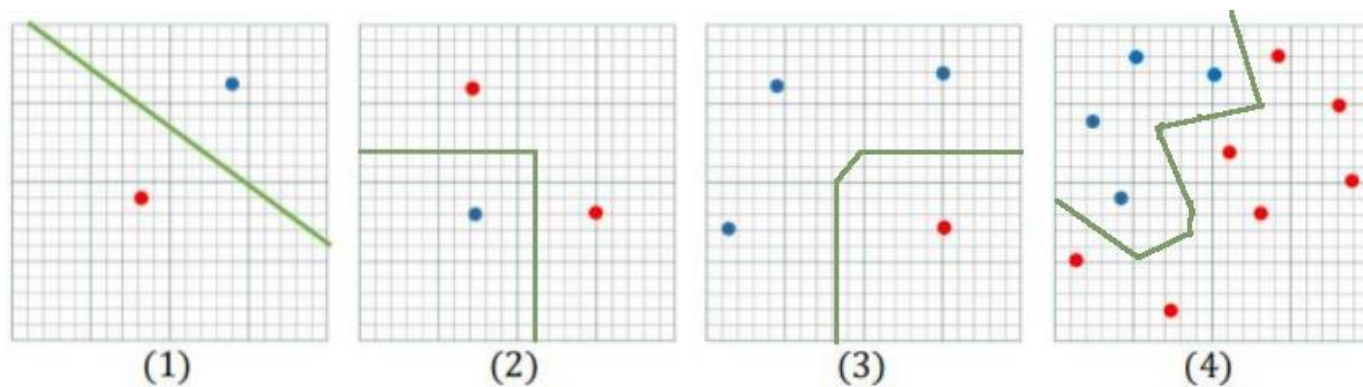
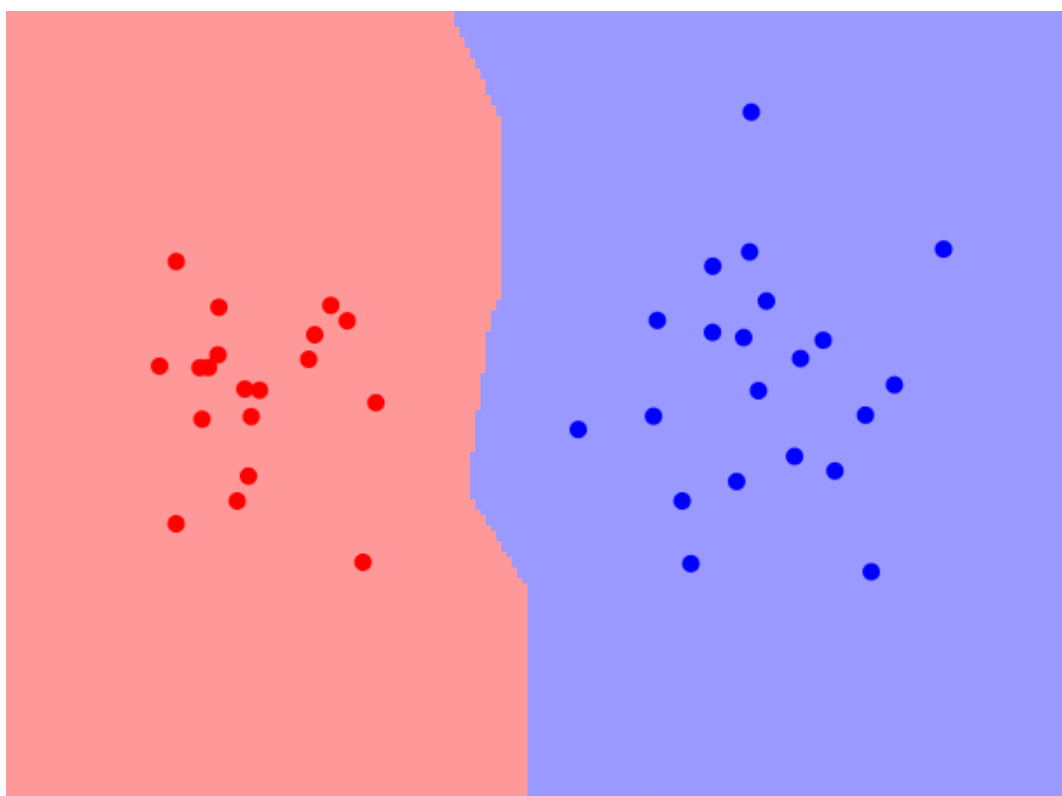


1



بله، در صورت ذخیره فاصله به جای همه اطلاعات، در مصرف حافظه صرفه جویی می‌شود، به دلیل اینکه فقط قسمتی از اطلاعات مرز دسته بندی هارا مشخص می‌کنند، ذخیره دیتاهایی که در مرز قرار ندارند بیهوده است (مانند شکل زیر)



میبینیم که بخش زیادی از اطلاعات در تعیین مرز نقشی ندارند، در نتیجه میتوان گفت ذخیره مرز شیوه بهتری برای ذخیره یک مدل NN-1 است

• خیر، میتوان داده آموزشی به knn اضافه کرد.

به دلیل اینکه در هر بار predict کردن در یک مدل knn، فاصله ها محاسبه شده و سورت می‌شوند، یک مدل pre-trained وجود ندارد و اضافه کردن اطلاعات جدید مشکلی در مدل ایجاد نمی‌کند.

2

درخت‌های تصمیم در صورت رشد کامل، همیشه overfit می‌شوند، هرس کردن یک درخت تصمیم کمک می‌کند که درخت کمتر به دیتای آموزش حساس باشد و بهتر بر روی دیتاهای تست عمل کند.

در پیش هرس، قبل از اینکه درخت به رشد کامل برسد الگوریتم متوقف می‌شود. در پس هرس، به الگوریتم اجازه ساخت درخت کامل داده می‌شود، سپس زیر درخت‌هایی از آن جایگزین می‌شود.

3

$X = (\text{headache}=1, \text{cough}=0, \text{fever}=1)$

Total yes = 5

$P(\text{headache} = 1 \mid \text{yes}) = 2/5$

$P(\text{cough} = 0 \mid \text{yes}) = 1/5$

$P(\text{fever} = 1 \mid \text{yes}) = 3/5$

$P(X \mid Y = \text{yes}) = 2/5 * 1/5 * 3/5 = 6/125$

Total no = 5

$P(\text{headache} = 1 \mid \text{no}) = 3/5$

$P(\text{cough} = 0 \mid \text{no}) = 3/5$

$P(\text{fever} = 1 \mid \text{no}) = 2/5$

$P(X \mid Y = \text{no}) = 2/5 * 3/5 * 3/5 = 18/125$

$Y = \text{no}$

Record	Age	Marital	Income	Risk
1	0.000000	0	0.880246	0
2	0.250000	0.5	0.011912	0
3	0.136364	1	0.193700	0
4	0.659091	1	0.000000	0
5	0.068182	0	0.924691	0
6	0.386364	0	0.399532	1
7	0.727273	0	0.190900	1
8	0.750000	0.5	1.000000	1
9	0.636364	0.5	0.902757	1
10	1.000000	0.5	0.483542	1

$X = (30, \text{Single}, \$30000)$ after normalization : $X = (0.181818, 0, 0.241631)$

Distances:

1	2	3	4	5	6	7	8	9	10
0.663993	0.554454	1.002179	1.134096	0.692447	0.258402	0.547809	1.071426	0.945356	0.988909

3 minimum distances:

2	6	7
0.554454	0.258402	0.547809

Using maximum voting, prediction is 1 (Good Risk)