1. Consider the following training set, in which each example has two tertiary attributes (0, 1, or 2) and one of two possible classes (*X* or *Y*).

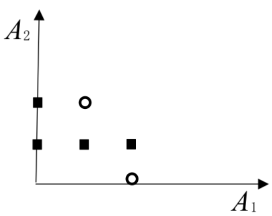
Example	A_1	A_2	Class
1	0	1	X
2	2	1	X
3	1	1	X
4	0	2	X
5	1	2	Y
6	2	0	Y

1) What feature would be chosen for the split at the root of a decision tree using the information gain criterion? Show the details. (Note: we split attributes at each value of the attributes, for example, $A_1=0, A_1=1, A_1=2$)

2) What would the Naïve Bayes algorithm predict for the class of the following new example? Show the details of the solution.

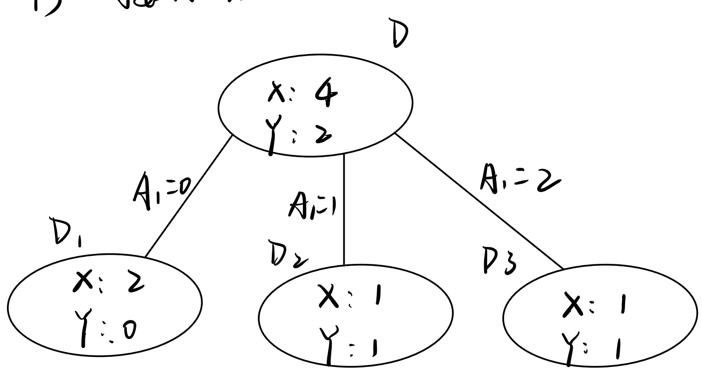
Example	A_1	A_2	Class
7	2	2	?

3) Draw the decision boundaries for the nearest neighbor algorithm assuming that we are using standard Euclidean distance to compute the nearest neighbors.



4) Which of these classifiers will be the least likely to classify the following data points correctly? Please explain the reason.
   a. ID3.
   b. Naïve Bayes
   c. Logistic Regression
   d. KNN

1) 按A_1分:



$H(D) = -\frac{4}{6}\log\frac{4}{6} - \frac{2}{6}\log\frac{2}{6}$
$= 0.918$

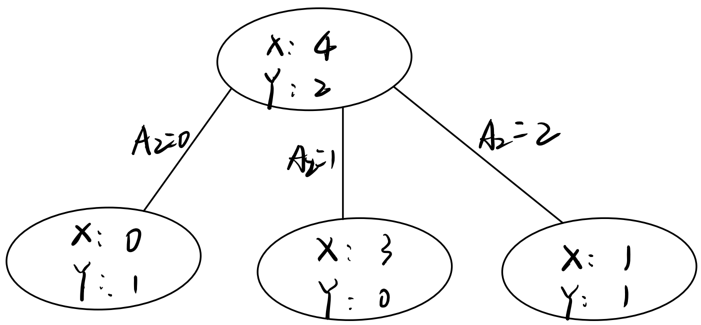
$H(D|A_1=0) = 0$

$H(D|A_1=1) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2}$
$= 1$

$H(D|A_1=2) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2}$
$= 1$

$\Rightarrow IG(D|A_1) = 0.918 -$
$(\frac{2}{6}\times 0 + \frac{2}{6}\times 1 + \frac{2}{6}\times 1) = 0.251$

2. 按A_2分.



$H(D)=0.918$

$H(D|A_2=0)=0$

$H(D|A_2=1)=0$

$H(D|A_2=2)=-\frac{1}{2}\log\frac{1}{2}-\frac{1}{2}\log\frac{1}{2}$
$=1$

$\Rightarrow IG(D|A_2)=0.918-(\frac{1}{6}\times 0+\frac{3}{6}\times 0+\frac{2}{6}\times 1)=0.585$

$\because 0.585>0.251$

$\therefore$ 应当选择 A_2.

2) $C_{NB}=\arg\max\limits_{c\in C} P(c)\prod_{i=1}^{d} P(x_i|c)$

$P(X)=\frac{2}{3}$ $\quad P(A_1=2|X)=\frac{1}{4}$ $\quad P(A_1=2|Y)=\frac{1}{2}$

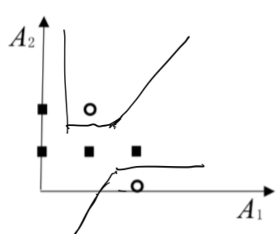
$P(Y)=\frac{1}{3}$ $\quad P(A_2=2|X)=\frac{1}{4}$ $\quad P(A_2=2|Y)=\frac{1}{2}$

$P(X)P(A_1=2|X)P(A_2=2|X)=\frac{2}{3}\times\frac{1}{4}\times\frac{1}{4}=\frac{1}{24}$

$P(Y)P(A_1=2|Y)P(A_2=2|Y)=\frac{1}{3}\times\frac{1}{2}\times\frac{1}{2}=\frac{1}{12}$

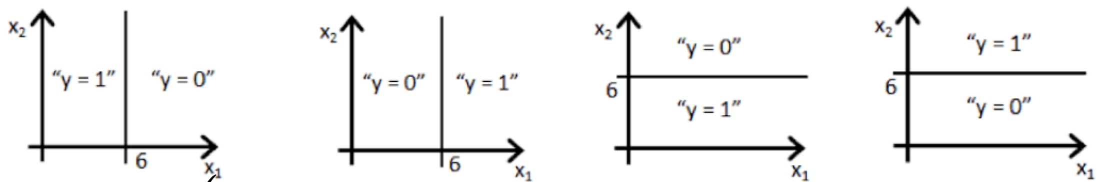
$\Rightarrow$ 朴素贝叶斯预测结果为Y

3.



4. C. 此数据非线性可分的数据且数据量过少，使用逻辑回归易过拟合.

2. You have trained a logistic classifier y=sigmoid($w_0+w_1x_1+w_2x_2$). Suppose w_0=6, w_1=-1, and w_2=0. Which of the following figures represents the decision boundary found by your classifier?



A B C D

sigmoid函数: $f(z) = \frac{1}{1+e^{-z}}$

$y = \text{sigmoid}(6-x_1) = \frac{1}{1+e^{x_1-6}} \Rightarrow A$

3. Suppose we are given a dataset $D=\{(x^{(1)},r^{(1)}),\ldots,(x^{(N)},r^{(N)})\}$ and aim to learn some patterns using the following algorithms. Match the update rule for each algorithm.

Algorithms:	**Update Rules:**
A: SGD for Logistic Regression $y = \text{sigmoid}(\boldsymbol{w}^T x)$	***1.*** $\boldsymbol{w}_t \leftarrow \boldsymbol{w}_t + \eta\,(r^{(l)} - \boldsymbol{w}_t^T x^{(l)})$
B: Least Mean Squares for Linear Regression $y = \boldsymbol{w}^T x$	***2.*** $\boldsymbol{w}_t \leftarrow \boldsymbol{w}_t + \frac{1}{1+\exp \eta(r^{(l)} - y^{(l)})}$
C: Perceptron $y = \text{sign}(\boldsymbol{w}^T x)$ (where sign(a)=1 if a>0 else -1)	***3.*** $\boldsymbol{w}_t \leftarrow \boldsymbol{w}_t + \eta(r^{(l)} - y^{(l)})x_i^{(l)}$