1. (a)

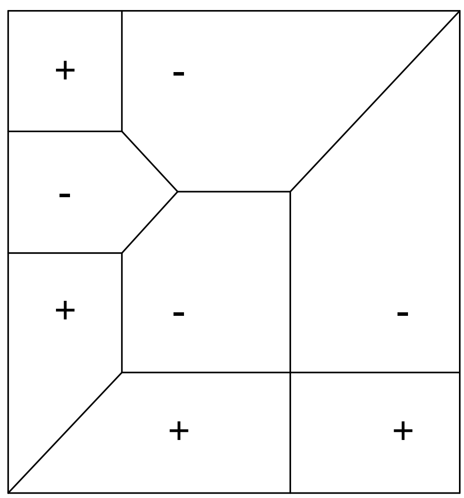
$$\begin{aligned} P(E|A,C,D) &= \frac{P(A,C,D,E)}{P(A,C,D)} \\ &= \frac{P(A,B,C,D,E)+P(A,\neg B,C,D,E)}{P(A,B,C,D,E)+P(A,\neg B,C,D,E)+P(A,B,C,D,\neg E)+P(A,\neg B,C,D,\neg E)} \end{aligned}$$

(b)

$$P(E|ACD) = P(E|CD)$$

2. The expected utility of "gamble" is $U(g) = 10 \times 0.1 = 1$, which is smaller than the expected utility of "not gamble", 2. So I would prefer the choice "not gamble".(lack spirit of adventuring)

3. (a)



(b)The area domained by positive examples is $\frac{6}{16}$ of the whole area, and the area domained by negative examples is $\frac{10}{16}$. Since test examples are distributed uniformly, they are more likely to be in the negative region, so negative class will be predicted most often.

4. Forward selection will be faster. The forward selection algorithm adds features from an empty set step by step; the backward elimination algorithm removes features from a complete set step by step. So in the process, the forward selection algorithm calculate the distances in a lower dimension space, but the backward elimination algorithm always does it in a higher dimension space.

5. 1. choose the best feature from X_1,X_2, and X_3, shown in figure 1: So, at the first step, choose X_1 as the best feature.

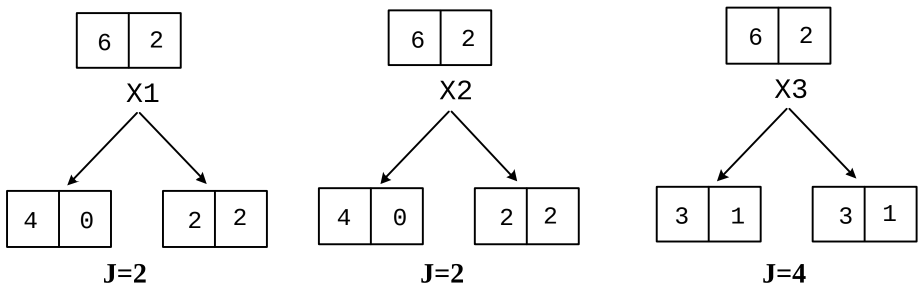


Figure 1: step 1

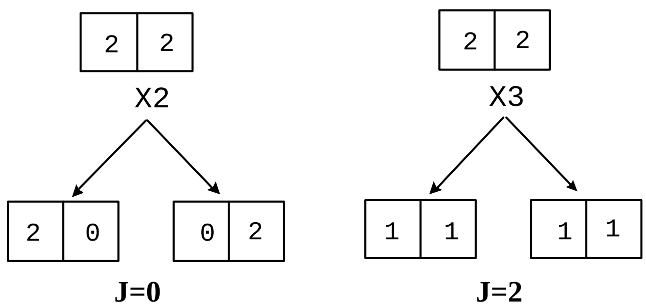


Figure 2: step 2

2. choose the best feature between X_2 and X_3 to separate the subset when $X_1 = 0$, shown in figure 2 : X_2 is the best feature to split the subset at this step. And the work is done.

6. Initialization(ignoring missing values):

$$P(A) = 0.25 \quad P(B|A) = 0 \quad P(B|\neg A) = \frac{1}{3} \quad P(C|A) = P(C\neg A) = 1$$

E-step:

$$P(C|\neg A, B) = P(C|\neg A) = 1) \quad P(C|\neg A, \neg B) = P(C|\neg A) = 1$$

So the missing value is estimated as $< 0, 1, 1 >$, and $< 0, 0, 1 >$.

M-step: exactly the same as the initial parameters. So the process converges.

7. If we assume boolean values of 1(true) and -1(false), we can build the following neural network represent the boolean function:

$$Y = (X_1\ AND\ X_2)\ OR\ (X_2\ AND\ \neg X_3)\ OR\ (\neg X_2\ AND\ X_3)$$

The shadow nodes are perceptrons, which means they take input and output the sign value(1 or -1) of the weighted combination of inputs and the bias. The network structure is shown in figure 3, the weights in the network defined as:

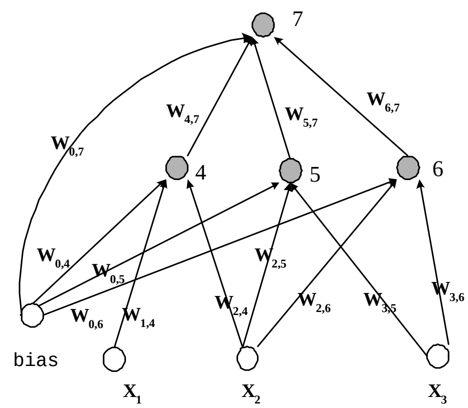


Figure 3: nn

$w_{0,4} = w_{0,5} = w_{0,6} = -0.5$, $w_(0,7) = 0.8$, $w_{1,4} = w_{2,4} = w_{2,5} = w_{3,6} = 0.5$, $w_{2,6} = w_{3,5} = -0.5$, and $w_{4,7} = w_{5,7} = w_{6,7} = \frac{1}{3}$