# Problem 2

Red = 20 and Blue = 29

* Entropy(Color) = -20\*log(20/49)/49 – 29\*log(29/49)/49 = 0.975
* Compare x1=3 and x1=4 for the x1 optimal threshold
  + When x1=3
    - Entropy(x1<3) = Entropy(x1<3|Red) + Entropy(x1<3|Blue)

=-17\*log(17/26)/26 – 9\*log(9/26)/26 = 0.957

* + - Entropy(x1>=3) = Entropy(x1>=3|Red) + Entropy(x1>=3|Blue)

= -3\*log(3/23)/23 – 20\*log(20/23)/23 = 0.56

* + - Gain(x1 = 3) = Entropy(Color) - Entropy(x1>=3) - Entropy(x1<3)

= 0.975 – 26\*0.957/49 – 23\*0.56/49 = 0.204

* + When x1=4
    - Entropy(x1<4) = Entropy(x1<4|Red) + Entropy(x1<4|Blue)

=-20\*log(20/29)/29 – 9\*log(9/29)/29 = 0.895

* + - Entropy(x1>=4) = Entropy(x1>=4|Red) + Entropy(x1>=4|Blue)

= -0\*log(0/20)/20 – 20\*log(20/20)/20 = infinity

* + - Gain(x1 = 4) = Entropy(Color) - Entropy(x1>=4) - Entropy(x1<4)

= 0.975 – 29\*0.895/49 – 20\*infinity/49 = -infinity

* + Compare two Gains(x1=3) and Gain(x1=4), x1=4 is the optimal threshold
* Compare x2=3 and x2=5 for the x2 optimal threshold
  + When x2=3
    - Entropy(x2<3) = Entropy(x2<3|Red) + Entropy(x2<3|Blue)

=-10\*log(10/20)/20 – 10\*log(10/20)/20 = 1

* + - Entropy(x2>=3) = Entropy(x2>=3|Red) + Entropy(x2>=3|Blue)

= -10\*log(10/29)/29 – 19\*log(19/29)/29 = 0.929

* + - Gain(x2 = 3) = Entropy(Color) - Entropy(x2>=3) - Entropy(x2<3)

= 0.975 – 20\*1/49 – 29\*0.929/49 = 0.017

* + When x2=5
    - Entropy(x2<5) = Entropy(x2<5|Red) + Entropy(x2<5|Blue)

=-20\*log(20/38)/38 – 18\*log(18/38)/38 = 0.998

* + - Entropy(x2>=5) = Entropy(x2>=5|Red) + Entropy(x2>=5|Blue)

= -0\*log(0/11)/11 – 11\*log(11/11)/11 = infinity

* + - Gain(x2 = 5) = Entropy(Color) - Entropy(x2>=5) - Entropy(x2<5)

= 0.975 – 39\*0.998/49 – 11\*infinity/49 = -infinity

* + Compare two Gains(x2=3) and Gain(x2=5), x2=3 is the optimal threshold

Construct the decision tree; since Gain(x1=3) > Gain(x2=3) (0.204 > 0.017), the first attribute split is at x1=3 then x2=3.