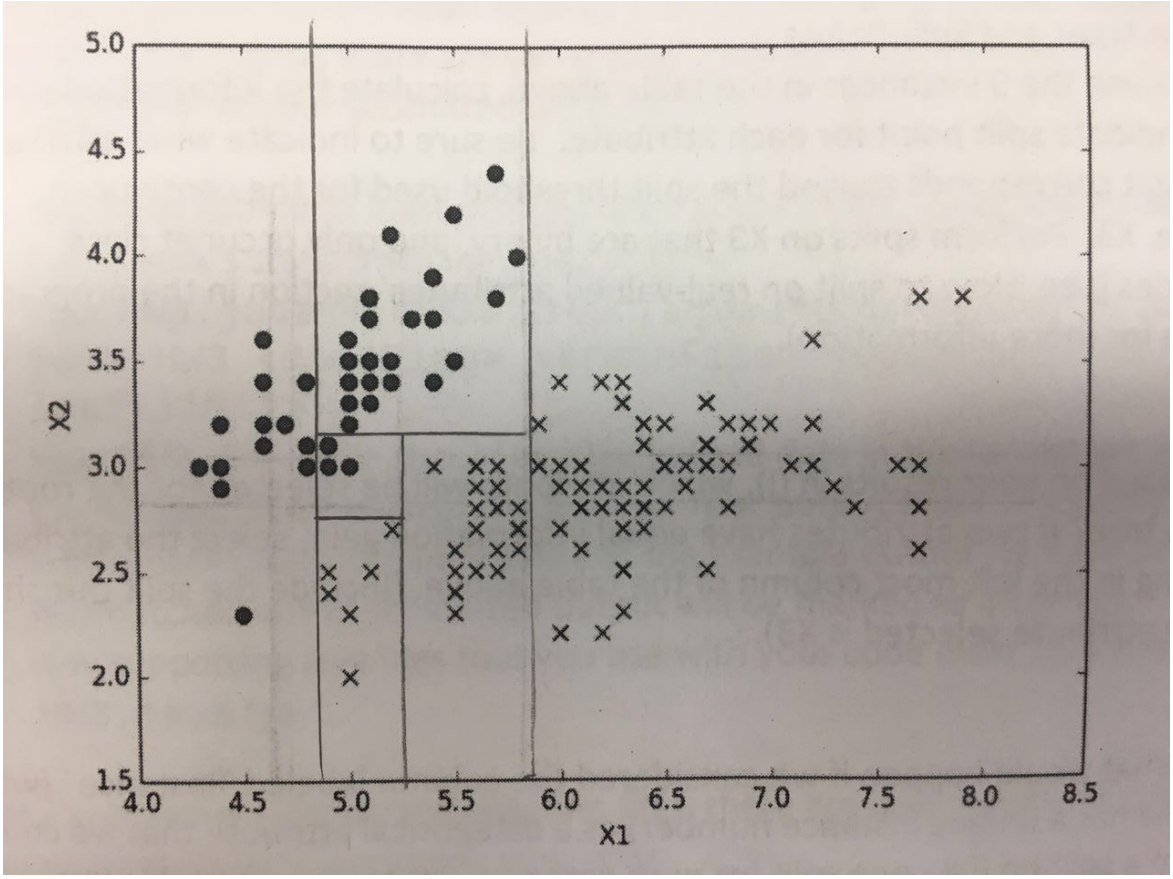
Problem 2 Decision Trees

(a) Manually Construct a Decision Tree



(b) Decision Trees and Split Points

(i)

$$H(Class) = H\left(\frac{3}{9}, \frac{6}{9}\right) = 0.9183$$

Split on X1:

T(4) : 2Y 2N F(5) : 4Y 1N

$$H\left(\frac{Class}{X1}\right) = \frac{4}{9} * H\left(\frac{2}{4}, \frac{2}{4}\right) + \frac{5}{9} * H\left(\frac{4}{5}, \frac{1}{5}\right) = \frac{4}{9} * 1 + \frac{5}{9} * 0.7219 = 0.8455$$

Information gain: 0.9183 – 0.8455 = 0.0728

Split on X2:

T(5): 4Y 1N F(4): 2Y 2N

$$H\left(\frac{Class}{X2}\right) = \frac{4}{9} * H\left(\frac{2}{4}, \frac{2}{4}\right) + \frac{5}{9} * H\left(\frac{4}{5}, \frac{1}{5}\right) = \frac{4}{9} * 1 + \frac{5}{9} * 0.7219 = 0.8455$$

Information gain: 0.9183 – 0.8455 = 0.0728

Split on X3:

After sorting the attribute values:

1	3	4	5	5.1	6	7	7.5	8
Y	N	Y	Y	Y	Y	Y	N	N

Attribute values to split: (1+3)/2 = 2; (7+7.5)/2 = 7.25;

Split on 2: X3<= 2 (1) : 1Y X3> 2 (8) : 5Y 3N

$$H\left(\frac{Class}{X3}\right) = \frac{1}{9} * H\left(\frac{1}{1}, \frac{0}{1}\right) + \frac{8}{9} * H\left(\frac{5}{8}, \frac{3}{8}\right) = \frac{8}{9} * 0.9544 = 0.8484$$

Information gain: 0.9183 – 0.8484 = 0.0699

Split on 3.5: X3 <= 3.5 (2): 1Y1N X3>3.5 (7) : 5Y2N

$$H\left(\frac{Class}{X3}\right) = \frac{2}{9} * H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{7}{9} * H\left(\frac{5}{7}, \frac{2}{7}\right) = \frac{2}{9} * 1 + \frac{7}{9} * 0.8631 = 0.8935$$

Information gain: 0.9183 – 0.8935 = 0.0248

Split on 7.25: X3 <= 7.25 (7) : N1 Y6 X3 > 7.25 (2) : N2

$$H\left(\frac{Class}{X3}\right) = \frac{7}{9} * H\left(\frac{1}{7}, \frac{6}{7}\right) + \frac{2}{9} * H\left(\frac{2}{2}, \frac{0}{2}\right) = \frac{7}{9} * 0.5916 = 0.4601$$

Information gain: 0.9183 – 0.4601 = 0.4582

(ii)

Based on results in (i), X3 will be selected for the root of the decision tree. The threshold is 7.25.

(iii)

No, the “Instance” shouldn’t be used as a categorical attribute. Let’s say we find an optimal threshold for the training data T, with 1 < T < 9. However, as we used this decision tree to predict new data points, whose instance attribute increases from 10, such as 10, 11, 12, 13, 14, etc. All new data points fall into the subtree where the Instance attribute > T. Therefore, the “Instance” is meaningless if it is treated as a categorical attribute. Instance is not a proper attribute for the points.