Recitation 11 Solutions

1.

18.6 Note that to compute each split, we need to compute $Remainder(A_i)$ for each attribute A_i, and select the attribute the provides the minimal remaining information, since the existing information prior to the split is the same for all attributes we may choose to split on.

Computations for first split: remainders for $A1$, $A2$, and $A3$ are

$$(4/5)(-2/4\log(2/4) - 2/4\log(2/4)) + (1/5)(-0 - 1/1\log(1/1)) = 0.800$$
$$(3/5)(-2/3\log(2/3) - 1/3\log(1/3)) + (2/5)(-0 - 2/2\log(2/2)) \approx 0.551$$
$$(2/5)(-1/2\log(1/2) - 1/2\log(1/2)) + (3/5)(-1/3\log(1/3) - 2/3\log(2/3)) \approx$$

Choose A_2 for first split since it minimizes the remaining information needed to classify all examples. Note that all examples with $A_2 = 0$, are correctly classified as $B = 0$. So we only need to consider the three remaining examples (x_3, x_4, x_5) for which $A_2 = 1$.

After splitting on A_2, we compute the remaining information for the other two attributes on the three remaining examples (x_3, x_4, x_5) that have $A_2 = 1$. The remainders for $A1$ and $A3$ are

$$(2/3)(-2/2\log(2/2) - 0) + (1/3)(-0 - 1/1\log(1/1)) = 0$$
$$(1/3)(-1/1\log(1/1) - 0) + (2/3)(-1/2\log(1/2) - 1/2\log(1/2)) \approx 0.667.$$

So, we select attribute A1 to split on, which correctly classifies all remaining examples.

2.

$R_1(1,1) = P(A_1(\mathbf{x}_i) = 1 | y_i = 1) = \frac{2}{2} = 1$
$R_1(1,0) = P(A_1(\mathbf{x}_i) = 1 | y_i = 0) = \frac{2}{3}$
$R_2(1,1) = P(A_2(\mathbf{x}_i) = 1 | y_i = 1) = \frac{2}{2} = 1$
$R_2(1,0) = P(A_2(\mathbf{x}_i) = 1 | y_i = 0) = \frac{1}{3}$
$R_3(1,1) = P(A_3(\mathbf{x}_i) = 1 | y_i = 1) = \frac{1}{2}$
$R_3(1,0) = P(A_3(\mathbf{x}_i) = 1 | y_i = 0) = \frac{1}{3}$

$\mathbf{x}_6$:

$P(y_6 = 1 | \mathbf{x}_6 = \langle 0,0,0 \rangle)$
$\propto P(y_6 = 1) \cdot P(A_1(\mathbf{x}_6) = 0 | y_6 = 1) \cdot P(A_2(\mathbf{x}_6) = 0 | y_6 = 1) \cdot P(A_3(\mathbf{x}_6) = 0 | y_6 = 1)$
$= \frac{2}{5} \cdot \frac{0}{2} \cdot \frac{0}{2} \cdot \frac{1}{2} = 0.$
$P(y_6 = 0 | \mathbf{x}_6 = \langle 0,0,0 \rangle)$
$\propto P(y_6 = 0) \cdot P(A_1(\mathbf{x}_6) = 0 | y_6 = 0) \cdot P(A_2(\mathbf{x}_6) = 0 | y_6 = 0) \cdot P(A_3(\mathbf{x}_6) = 0 | y_6 = 0)$
$= \frac{3}{5} \cdot \frac{1}{3} \cdot \frac{2}{3} \cdot \frac{2}{3} = \frac{4}{45} > 0.$

So we predict $y_6 = 0$.

$\mathbf{x}_7$:

$P(y_7 = 1 | \mathbf{x}_7 = \langle 0,1,1 \rangle)$
$\propto P(y_7 = 1) \cdot P(A_1(\mathbf{x}_7) = 0 | y_7 = 1) \cdot P(A_2(\mathbf{x}_7) = 1 | y_7 = 1) \cdot P(A_3(\mathbf{x}_7) = 1 | y_7 = 1)$
$= \frac{2}{5} \cdot \frac{0}{2} \cdot \frac{2}{2} \cdot \frac{1}{2} = 0.$
$P(y_7 = 0 | \mathbf{x}_7 = \langle 0,1,1 \rangle)$
$\propto P(y_7 = 0) \cdot P(A_1(\mathbf{x}_7) = 0 | y_7 = 0) \cdot P(A_2(\mathbf{x}_7) = 1 | y_7 = 0) \cdot P(A_3(\mathbf{x}_7) = 1 | y_7 = 0)$
$= \frac{3}{5} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{45} > 0.$

So we predict $y_7 = 0$.

With Laplace correction:

$R_1(1,1) = P(A_1(\mathbf{x}_i) = 1 | y_i = 1) = \frac{3}{4}$
$R_1(1,0) = P(A_1(\mathbf{x}_i) = 1 | y_i = 0) = \frac{3}{5}$
$R_2(1,1) = P(A_2(\mathbf{x}_i) = 1 | y_i = 1) = \frac{3}{4}$
$R_2(1,0) = P(A_2(\mathbf{x}_i) = 1 | y_i = 0) = \frac{2}{5}$
$R_3(1,1) = P(A_3(\mathbf{x}_i) = 1 | y_i = 1) = \frac{2}{3}$
$R_3(1,0) = P(A_3(\mathbf{x}_i) = 1 | y_i = 0) = \frac{2}{4}$

$\mathbf{x}_6$:

$P(y_6 = 1 | \mathbf{x}_6 = \langle 0,0,0 \rangle) \propto \frac{2}{5} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{4} = \frac{1}{45}.$
$P(y_6 = 0 | \mathbf{x}_6 = \langle 0,0,0 \rangle) \propto \frac{3}{5} \cdot \frac{2}{5} \cdot \frac{3}{5} \cdot \frac{3}{5} = \frac{54}{625} > \frac{1}{45}.$

So we predict $y_6 = 0$.

$\mathbf{x}_7$:

$P(y_7 = 1 | \mathbf{x}_7 = \langle 0,1,1 \rangle) \propto \frac{2}{5} \cdot \frac{1}{3} \cdot \frac{3}{4} \cdot \frac{2}{4} = \frac{1}{20}.$
$P(y_7 = 0 | \mathbf{x}_7 = \langle 0,1,1 \rangle) \propto \frac{3}{5} \cdot \frac{2}{5} \cdot \frac{2}{5} \cdot \frac{2}{5} = \frac{24}{625} < \frac{1}{20}.$

So we predict $y_7 = 1$.

The decision tree predicts $y_6 = 0$ and $y_7 = 0$.