

6 Decision Trees and Random Forest

6.1 Decision Tree for classification

i.

$$\text{Parent Node :} \quad p_0 = 6/12 \quad p_1 = 6/12 \quad \implies G_P = 0.500$$

(a) $x_0 \leq 0.25$

$$\text{Left Node :} \quad p_0 = 2/4 \quad p_1 = 2/4 \quad \implies G_L = 0.500$$

$$\text{Right Node :} \quad p_0 = 4/8 \quad p_1 = 4/8 \quad \implies G_R = 0.500$$

$$\text{Goodness-of-split :} \quad N_L = 4/12 \quad N_R = 8/12 \quad \implies \Delta G = 0.000$$

(b) $x_1 \leq 0.5$

$$\text{Left Node :} \quad p_0 = 3/6 \quad p_1 = 3/6 \quad \implies G_L = 0.500$$

$$\text{Right Node :} \quad p_0 = 3/6 \quad p_1 = 3/6 \quad \implies G_R = 0.500$$

$$\text{Goodness-of-split :} \quad N_L = 6/12 \quad N_R = 6/12 \quad \implies \Delta G = 0.000$$

(c) $x_0 \leq 0.85$

$$\text{Left Node :} \quad p_0 = 4/9 \quad p_1 = 5/9 \quad \implies G_L = 0.494$$

$$\text{Right Node :} \quad p_0 = 2/3 \quad p_1 = 1/3 \quad \implies G_R = 0.444$$

$$\text{Goodness-of-split :} \quad N_L = 9/12 \quad N_R = 3/12 \quad \implies \Delta G = 0.019$$

- ii. In order to grow the tree, at each node, we calculate the goodness-of-fit for each split function and select the split function with the maximum goodness of split. The calculation is similar to the previous task 6.1(i).

In the steps below we directly show the calculation of goodness-of-split. The split function with the maximum goodness-of-split at each node is shown in bold font.

Root node: $G_P = 0.5$

$$x_0 \leq 0.150 \implies \Delta G = 0.500 - ((2/12) \times 0.500 + (10/12) \times 0.500) = 0.000$$

$$x_0 \leq 0.250 \implies \Delta G = 0.500 - ((4/12) \times 0.500 + (8/12) \times 0.500) = 0.000$$

$$x_0 \leq 0.500 \implies \Delta G = 0.500 - ((6/12) \times 0.500 + (6/12) \times 0.500) = 0.000$$

$$x_0 \leq 0.750 \implies \Delta G = 0.500 - ((8/12) \times 0.500 + (4/12) \times 0.500) = 0.000$$

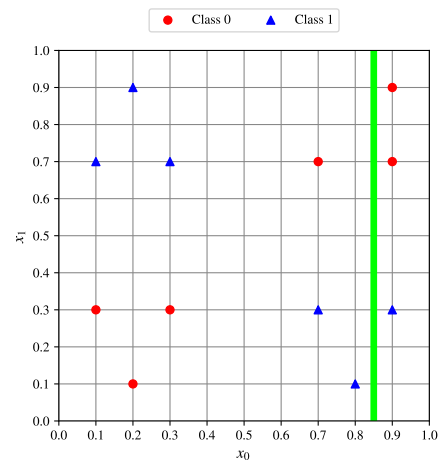
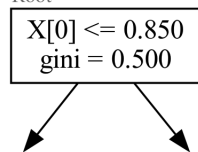
$$x_0 \leq \mathbf{0.850} \implies \mathbf{\Delta G = 0.500 - ((9/12) \times 0.494 + (3/12) \times 0.444) = 0.019}$$

$$x_1 \leq 0.200 \implies \Delta G = 0.500 - ((2/12) \times 0.500 + (10/12) \times 0.500) = 0.000$$

$$x_1 \leq 0.500 \implies \Delta G = 0.500 - ((6/12) \times 0.500 + (6/12) \times 0.500) = 0.000$$

$$x_1 \leq 0.800 \implies \Delta G = 0.500 - ((10/12) \times 0.500 + (2/12) \times 0.500) = 0.000$$

Root



Node L: $G_P = 0.494$

$$x_0 \leq 0.150 \implies \Delta G = 0.494 - ((2/9) \times 0.500 + (7/9) \times 0.490) = 0.002$$

$$x_0 \leq 0.250 \implies \Delta G = 0.494 - ((4/9) \times 0.500 + (5/9) \times 0.480) = 0.005$$

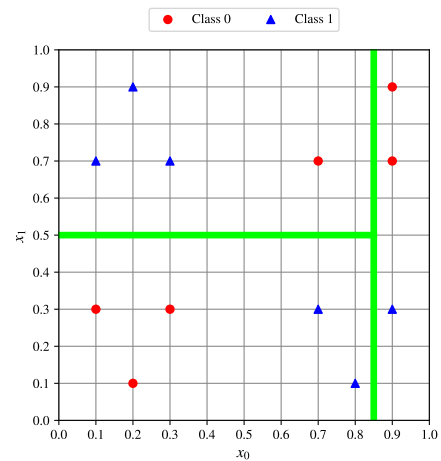
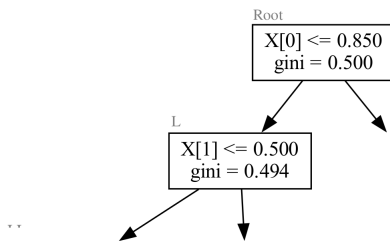
$$x_0 \leq 0.500 \implies \Delta G = 0.494 - ((6/9) \times 0.500 + (3/9) \times 0.444) = 0.012$$

$$x_0 \leq 0.750 \implies \Delta G = 0.494 - ((8/9) \times 0.500 + (1/9) \times 0.000) = 0.049$$

$$x_1 \leq 0.200 \implies \Delta G = 0.494 - ((2/9) \times 0.500 + (7/9) \times 0.490) = 0.002$$

$$\mathbf{x_1 \leq 0.500 \implies \Delta G = 0.494 - ((5/9) \times 0.480 + (4/9) \times 0.375) = 0.060}$$

$$x_1 \leq 0.800 \implies \Delta G = 0.494 - ((8/9) \times 0.500 + (1/9) \times 0.000) = 0.049$$



Node LL: $G_P = 0.48$

$$x_0 \leq 0.150 \implies \Delta G = 0.480 - ((1/5) \times 0.000 + (4/5) \times 0.500) = 0.080$$

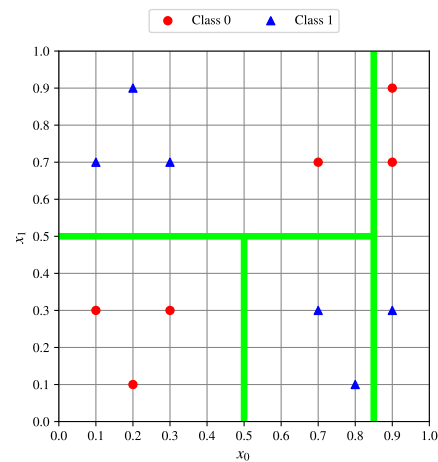
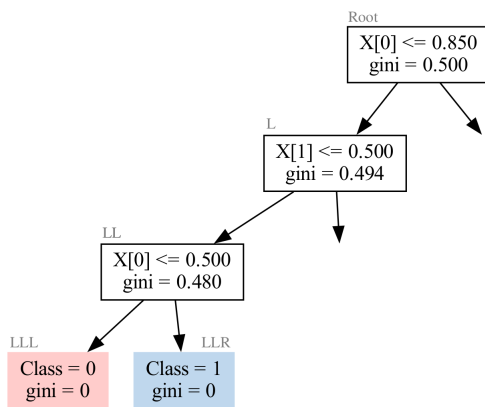
$$x_0 \leq 0.250 \implies \Delta G = 0.480 - ((2/5) \times 0.000 + (3/5) \times 0.444) = 0.213$$

$$\mathbf{x_0 \leq 0.500 \implies \Delta G = 0.480 - ((3/5) \times 0.000 + (2/5) \times 0.000) = \mathbf{0.480}$$

$$x_0 \leq 0.750 \implies \Delta G = 0.480 - ((4/5) \times 0.375 + (1/5) \times 0.000) = 0.180$$

$$x_1 \leq 0.200 \implies \Delta G = 0.480 - ((2/5) \times 0.500 + (3/5) \times 0.444) = 0.013$$

Since both the left and right nodes of $x_0 \leq 0.500$ have Gini score = 0, we can set these two nodes as leaf nodes.



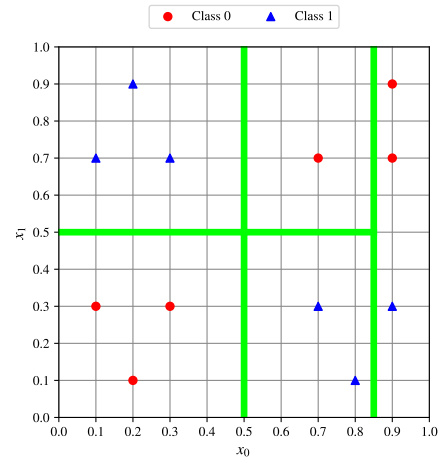
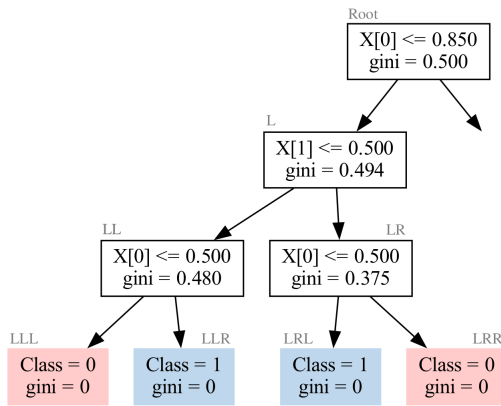
Node LR: $G_P = 0.375$

$$x_0 \leq 0.150 \implies \Delta G = 0.375 - ((1/4) \times 0.000 + (3/4) \times 0.444) = 0.042$$

$$x_0 \leq 0.250 \implies \Delta G = 0.375 - ((2/4) \times 0.000 + (2/4) \times 0.500) = 0.125$$

$$x_0 \leq \mathbf{0.500} \implies \mathbf{\Delta G = 0.375} - ((3/4) \times \mathbf{0.000} + (1/4) \times \mathbf{0.000}) = \mathbf{0.375}$$

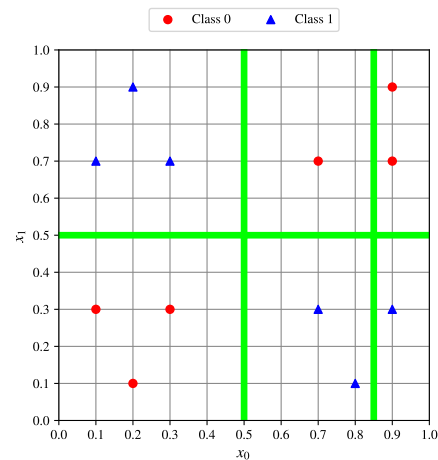
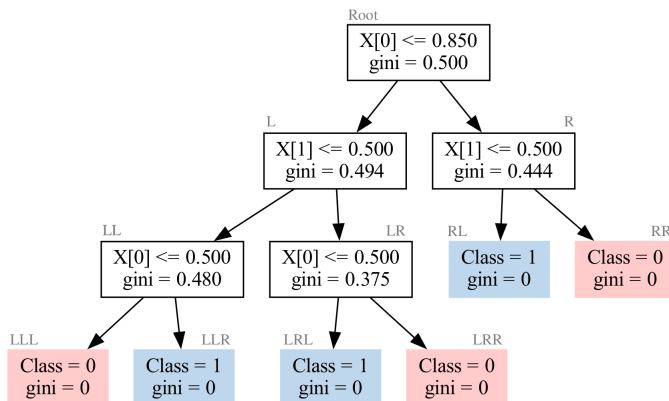
$$x_1 \leq 0.800 \implies \Delta G = 0.375 - ((3/4) \times 0.444 + (1/4) \times 0.000) = 0.042$$



Node R: $G_P = 0.444$

$$x_1 \leq \mathbf{0.500} \implies \mathbf{\Delta G = 0.444} - ((1/3) \times \mathbf{0.000} + (2/3) \times \mathbf{0.000}) = \mathbf{0.444}$$

$$x_1 \leq 0.800 \implies \Delta G = 0.444 - ((2/3) \times 0.500 + (1/3) \times 0.000) = 0.111$$



iii. Traversing through the graph, it is clear that

(a) $\mathbf{x} = [0.7 \ 0.9]^\top$ will be class 0 (Leaf node LRR)

(b) $\mathbf{x} = [0.5 \ 0.5]^\top$ will be class 0 (Leaf node LLL)