3. Trees.

a)

$X_1 > 2.5$

Y / \ N

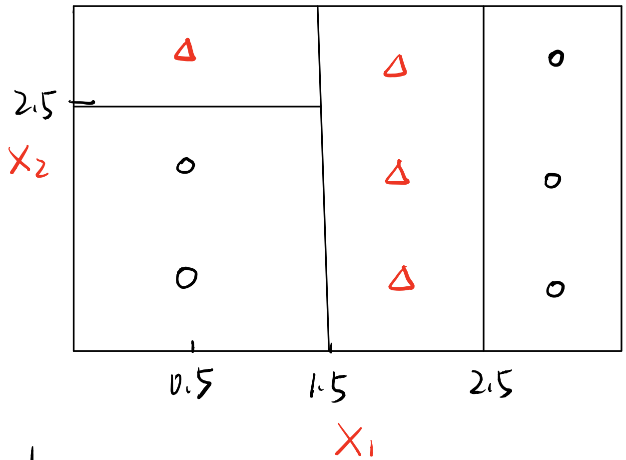
1 $\quad X_1 > 1.5$

Y / \ N

0 $\quad X_2 > 2.5$

Y / \ N

0 $\quad$ 1.



b) $L_1 = -\frac{3}{4}\cdot\log\frac{3}{4} - \frac{1}{4}\cdot\log\frac{1}{4} = 0.2442$

$L_2 = -\frac{3}{8}\cdot\log\frac{3}{8} - \frac{5}{8}\cdot\log\frac{5}{8} = 0.2873$

$L_3 = -\frac{1}{6}\cdot\log\frac{1}{6} - \frac{5}{6}\cdot\log\frac{5}{6} = 0.1957$

$L_4 = -\frac{9}{13}\cdot\log\frac{9}{13} - \frac{4}{13}\cdot\log\frac{4}{13} = 0.2681$

$L_5 = -\frac{2}{7}\log\frac{2}{7} - \frac{5}{7}\cdot\log\frac{5}{7} = 0.2598.$

$$\left(E(\tau) = \frac{|\tau_L|}{|\tau|} E(\tau_L) + \frac{|\tau_R|}{|\tau|} E(\tau_R) \right)$$

$E_{L_2+L_3} = \frac{40}{70} \times 0.2873 + \frac{30}{70} \times 0.1957 = 0.2480$

$E_{L_1} = \frac{40}{110} \times 0.2442 + \frac{70}{110} \times 0.2480 = 0.2466$

$E_{L_4+L_5} = \frac{65}{100} \times 0.2681 + \frac{35}{100} \times 0.2598 = 0.2652.$

$\text{Entropy} = \frac{110}{210} \times 0.2466 + \frac{100}{210} \times 0.2652 = 0.2555.$

c) i) $X = 2$, $Y = L_2 = 2.3$

ii) L_1's mean $= 13.46$, L_2's mean $= 2.3$.

- In-Sample SSE:

$L_1: 0.06^2 + 1.36^2 + 1.84^2 + 1.34^2 + 1.76^2 = 10.132$

L_2: $1.2^2 + 0.6^2 + 0.9^2 + 1.5^2 = 4.86.$

SSE in sample $= 14.992.$

- Out-of-sample SSE:

$$\bar{Y} = \frac{13.4+12.1+15.3+14.8+11.7+2.3+3.5+1.7+3.2+0.8}{10}$$

$$= 7.88$$

$$SSE_{(out)} = 326.36.$$