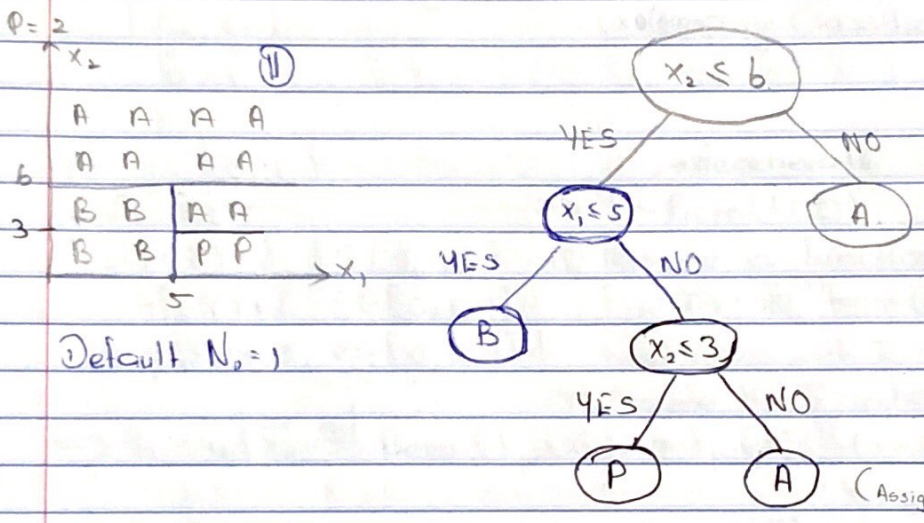
Lecture -19 04.17.2020

CART (1984) [Classification and Regression Tree]
Categorical (Norminal)

$y = [C_1, C_2, \ldots, C_K]$ K levels.
if $K = 2 \Rightarrow$ binary classification
$C_1 = 0, C_2 = 1$



Classification Algorithm

Same is the regression tree algorithm except

(1) The objective function which evaluates each split is not weighted avg of SSE_l, SSE_r. Instead we use the weighted avg. of $Gini_l$, $Gini_r$.

$$Gini_w = \frac{n_l\, Gini_l + n_r\, Gini_r}{n_l + n_r}$$

$$Gini_l = \sum_{k=1}^{K} \hat{p}_k (1 - \hat{p}_k) \quad \text{where } \hat{p}_k = \sum_{i=1}^{n_l} \frac{\mathbb{1}_{y_i = C_k}}{n_l}$$

(n_l: # of observation in node.)

(D) Leaf Assignment $\hat{y}$ = mode [$\hat{y}$'s]