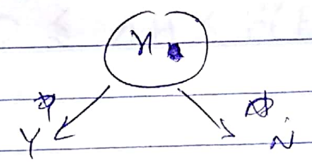
Q3. (a) Entropy $= -\sum_i P(x_i) \log_2 (P(x_i))$

when we split a feature F into Y and N.



weighted impurity $= P(Y)\,H(Y) + P(N)\,H(N)$

$= -P(Y) \sum_{i=1} P(x_i) \log_2 P(x_i)$

$\quad - P(N) \sum_{i=1} P(x_i) \log_2 P(x_i)$

$= -P(Y) \sum_{i=1} P(x_i|Y) \log_2 P(x_i|Y)$

$\quad - P(N) \sum_{i=1} P(x_i|N) \log_2 P(x_i|N)$

$= -\sum_{i=1} P(x_i, Y) \log_2 \left(\frac{P(x_i, Y)}{P(Y)} \right)$

$\quad - \sum_{i=1} P(x_i, N) \log_2 \left(\frac{P(x_i, N)}{P(N)} \right)$

$= -\sum_{i,F} P(x_i, F) \log_2 P(x_i, F)$

$\quad + P(Y) \log_2 P(Y) + P(N) \log_2 P(N)$

reduction $= H(x_i, F) - H(F)$

$\Delta H = H(x) - H(x, F) + H(F)$

$\because \quad H(x) \leq H(x, F) \leq H(x) + H(F)$

$$\Rightarrow \boxed{0 \leq \Delta H \leq H(F) \leq 1}$$

(b) when we deal with a multi branch case, we convert all categorial features into binary sets by making them questions with binary answers.

in that case all multiway branch get converted to a case where we have two possibility i.e. Yes or no

So for a B way branch, $B \geq 2$

$$0 \leq \Delta H(x) \leq \log_2(B).$$

Q5. Mutual Information is difference between entropy of unsplitted set and average of entropy of each split set, weighted by number of elements in subset.

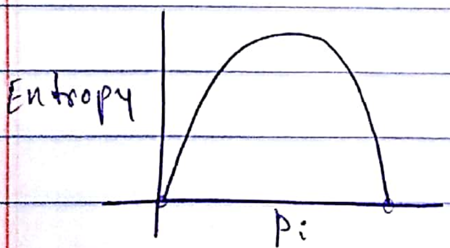
$$I.G(S, A) = E(S) - I(S, A)$$

$$= E(S) - \boxed{\sum_i \frac{|S_i|}{|S|} \cdot E(S_i)} \text{ let } X$$

and $\text{Entropy} = \sum_i \left(-p_i * \log_2(p_i)\right)$

p_i represents probability of i in Set.

So if ~~prob~~ p_i increases, log value reduces and same for low values of p_i



So if on split average Entropy reduces [represented by X] and Entropy of initial Set is Constant.

So we have a gain in Value of I.G.

Hence on low entropy we have higher gain as we are ~~have~~ sure that type of information we are getting is more pure.

Q5. Gini index is measure of how often would a randomly ~~labeled~~ chosen element from set would be incorrectly labelled.

$p_k \rightarrow$ probab. that k gets correct label

$(1-p_k) \rightarrow$ probab. that k gets wrong label.

$$\text{Gini index}(q) = \sum_{k=1}^{M} p_k \times \boxed{\sum_{\substack{k'=1 \\ k \neq k'}}^{M} p_{k'}} \Rightarrow \sum_{\substack{k=1 \\ k \neq k'}} p_{k'}$$

probability that k is not getting ~~correct~~ label (k')

$$= \sum_{k=1}^{M} p_k \times \sum_{k'=1} (1-p_k)$$

$$= \sum_{k=1} p_k \times (1-p_k)$$

Hence proved.

(*) it only works for $M > 2$, otherwise label for 2 elements is certain.