Class 18

Optimal code (PF) for representing a source $X \sim P_X$.

$$X \in \mathcal{X}, \quad |\mathcal{X}| = K$$

$p_1 \geq \dots \geq p_K$ (this is the P_X). $p_i = P_X(x_i)$

Lemma 1: $l_1 \leq \dots \leq l_K$ where l_i is the length of codeword given to symbol x_i

Lemma 2: In the tree of any optimal PF code, no leaf is unused ($\Leftrightarrow$ every leaf is a codeword)

$\Updownarrow$

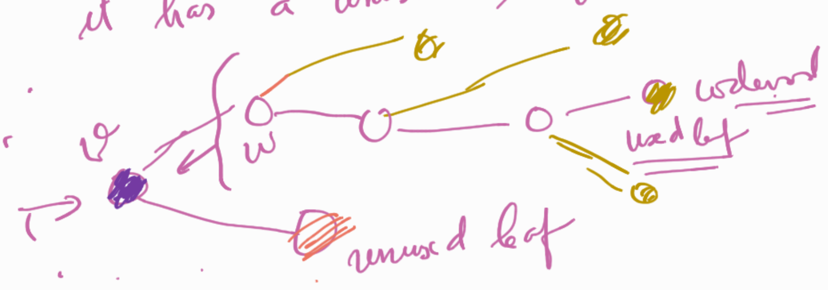
every node in the tree is either a codeword (or) it has two successors which are codewords.

Proof:

Let $\mathcal{C}$ be the optimal code.

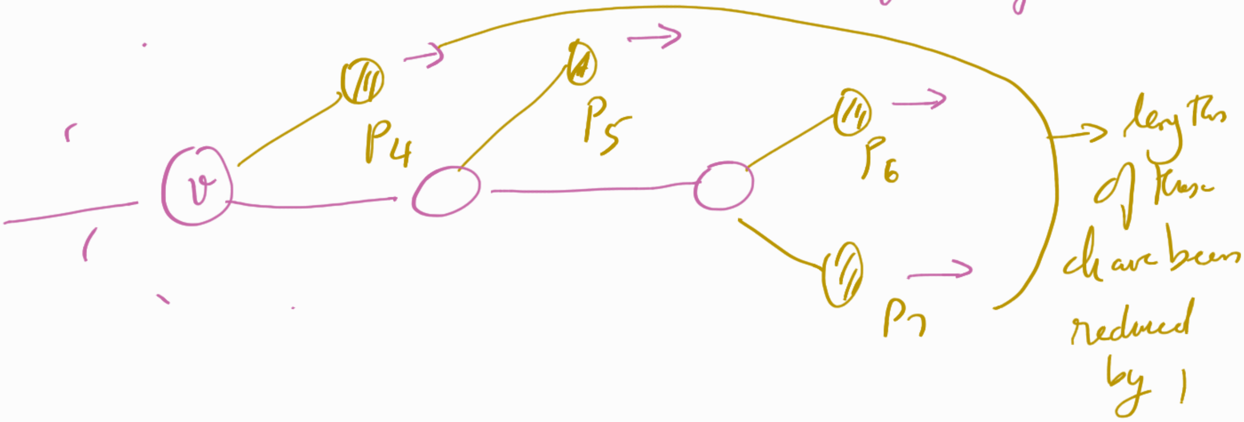
Let us consider the tree corresponding to this.

Suppose it has a unused leaf



Transform this

Get rid of the unused leaf & contract $v-\omega$ edge



Let the new code be $\mathcal{C}'$ ← with this new tree.

This code is also p-f!

Then clearly

$\bar{L}_{\mathcal{C}'} < \bar{L}_{\mathcal{C}}$

But this contradicts the optimality of $\mathcal{C}$..

— x —

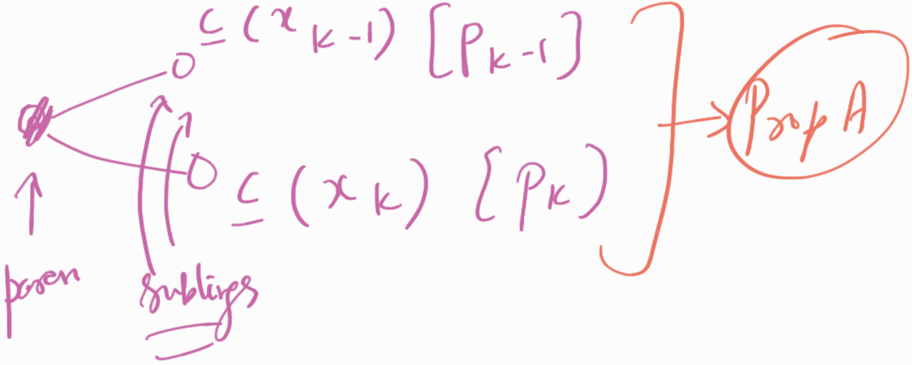
Thus the tree of an optimal code looks like

$L-2$

Lemma 3: There is an optimal code $\mathcal{C}$ for RV X such that the codewords associated to the 2 smallest probability symbols are siblings (share the same parent)

Property A

in $\mathcal{C}$,



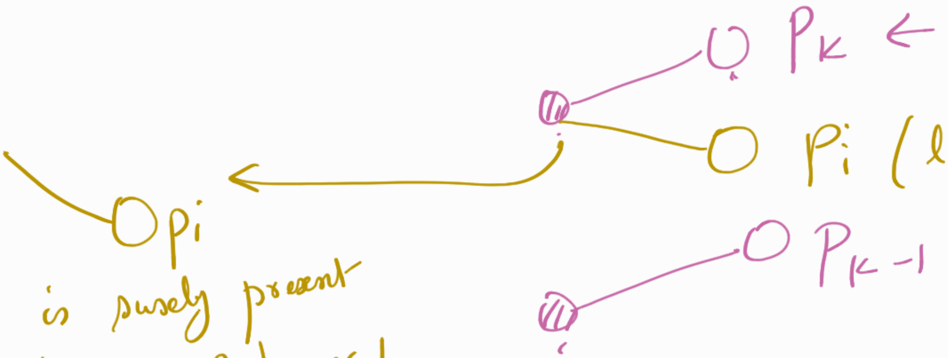
Proof:

Let $\tilde{\mathcal{C}}$ be arbitrary optimal code for RV.

Suppose $\tilde{\mathcal{C}}$ already satisfies Prop A. Then we are done. (we have nothing to prove)

Suppose $\tilde{\mathcal{C}}$ does not satisfy prop A.

Recall $l_{k-1} \le l_k$ as $p_{k-1} \geq p_k$.



p_k ← is at the deepest level because

$l_k \geq l_j$ $\forall j = 1 \ldots k-1$

$l_{k-1} \geq l_j$ $\forall j = 1 \ldots k-2$

p_i is surely present because of Lemma 1 & Lemma 2.

Note: $i \neq k-1$

Note: $l_i = l_k$

$$\Rightarrow \underbrace{l_i \le l_{k-1} \le l_k}_{\text{Lemma 1}} = l_i$$

by noticing that i & k are siblings

$$\Rightarrow l_i = l_{k-1} = l_k \quad //$$

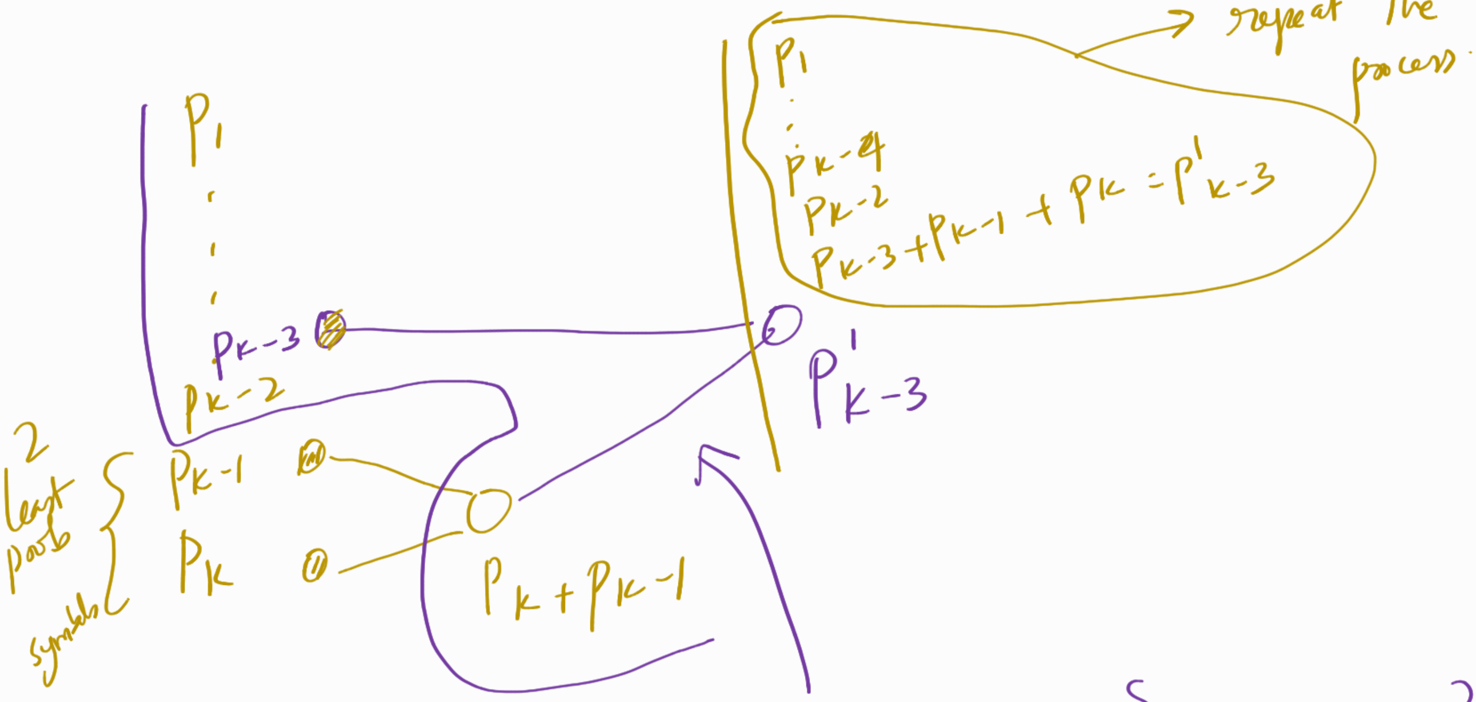
Interchange the codewords for the i^{th} & $(k-1)^{th}$ symbols & get a new code $\mathcal{C}$

Note that $\bar{L}_{\tilde{\mathcal{C}}} = \bar{L}_{\mathcal{C}} \Rightarrow \mathcal{C}$ is optimal.

Now: $\mathcal{C}$ satisfies Prop A $\rightarrow$ We are done.

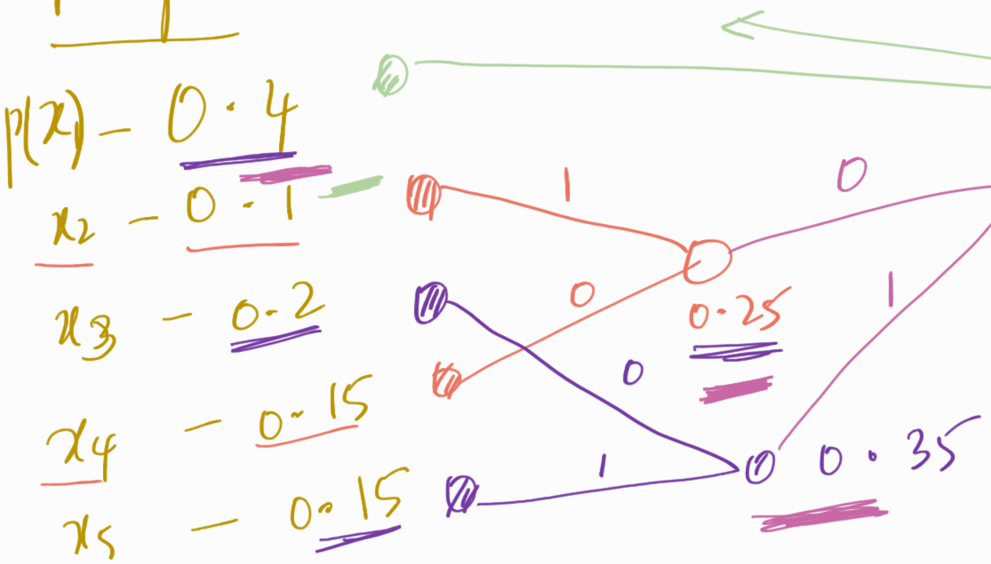
Huffman Coding:

P_X assume that $p_1 \geq \dots \geq p_k$



At this stage suppose $\{p_k+p_{k-1},\ p_{k-3}\}$ are the 2 least prob. symbols

Example:



$C(x_1) = 0$

$C(x_2) = 100$

$C(x_3) = 110$

$C(x_4) = 101$, $C(x_5) = 111$

Exercise: (1) Try S-F code for the above problem & check if the avg length of S-F > avg length of Huffman.

(2) Calculate H(X) & find out how much avg length of Huffman is more