SECTION - 1 (Question - 3)

Q1 → Huffman Coding ⇒

↳ This technique was developed by David Huffman.

↳ The codes are generated using the Huffman technique or procedures is called Huffman codes. These codes are prefix codes and optimum for a given model (set of probabilities).

↳ Assigning codewords to the modeled data is known as coding.

↳ It is easy to see that the first observation is correct.

↳ If symbols that occur more often had codewords that were longer than the codewords for symbols that occurred less often, the average number of bits per symbol would be longer than if the conditions were reversed.

↳ Therefore, a code that assigns longer codewords to symbols that occur more frequently cannot be optimum.

↳ The Huffman procedure is obtained by adding a simple requirement to these two observations.

↳ That is, if γ and δ are the two least probable symbols in the alphabet, if the codeword for γ was $m*0$, the codeword for δ would be $m*1$.

↳ Here, m is string of 1s and 0s, and * denotes concatination.

* Properties of Huffman coding ⇒

1. Unique prefix code ⇒ No ~~fuff~~ Huffman Code is a prefix of any other Huffman code. It prevents any ambiguity in decoding.

2. Optimality ⇒

Minimum redundancy code proved optimal for a given data model i.e. a given, accurate, probability distribution:

i) The two least frequency symbols will have the same length for their Huffman codes, differing only at the last bit.

ii) Symbol that occur more frequently will have

shorter Huffman Codes than symbols that occur less frequently.

c) The average code length for an information source S is strictly less than $H(S)+1$

$$\bar{l} < H(S)+1$$

where $\bar{l}$ is the average code length.

* Huffman coding algorithm ⇒

Huffman coding algorithm is a bottom up approach. Steps of the algorithm are :-

i) Initialization : Put all symbols on a list sorted according to their frequency count.

ii) Repeat until the list has only one symbol left :

a) From the list, pick two symbols with the lowest frequency counts form a Huffman subtree that has these two symbols as child nodes and create a parent node.

b) Assign the sum of the children's frequency counts to the parent and insert it into

the list such that the order is maintained.

c) Delete the children from the list.

iii) Assign a codeword for each leaf, based on the path from the root.