Tutorial - 3

Sahil Bondre - 618co021

Q1　$S = \{13, 15, 16, 16, 19, X, 20, 21, 22, 22, 25,$
　　　$25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 36, Y$
　　　$45, 46, 52, 70\}$

a　KNN based imputation involves finding
　　nearest cluster using KNN & using
　　the nearest value of the cluster

　　Let $k = 6$
i　KNN for $x \mid k = 6 = \{19, 20, 21, 22, 22, 16\}$
　　average $= 20$
　　$\therefore x = 20$

ii　KNN for $y \mid k = 6 = \{35, 35, 35, 36, 45, 46\}$
　　averag $= 38.67$
　　$\therefore y = 38.67$

b　Single value imputation involves replacing
　　the value by a single value like
　　mean, median, mode

　　Median of $S = 25$
　　$\therefore$　$x = 25$
　　　　　$y = 25$

Q2 $S = \{R, G, B, Y, R, G, G, G, B, R, Y, W\}$

[y = Black]

a In most frequent value replacement, the missing values are replaced with the mode of the data

Example : Green (G) is the most frequent occuring data [mode]

$\therefore$ X = Green

Y = Green

b In global constant replacement, the missing value is replaced by a global constant defined by user

Red

Example: Let ~~Black~~ be the global constant

$\therefore$ X = ~~Black~~ Red

Y = ~~Black~~ Red