Decision Tree Regressor

Dataset

O/P ← (continuous)

Exp ↓	Career Gap	Salary
→ 2	Yes	40K
→ 2.5	Yes	42K
→ 3	No	52K
4	No	60K
4.5	Yes	56K

$\bar{y}$ = 50K [Mean]

DT Classifier

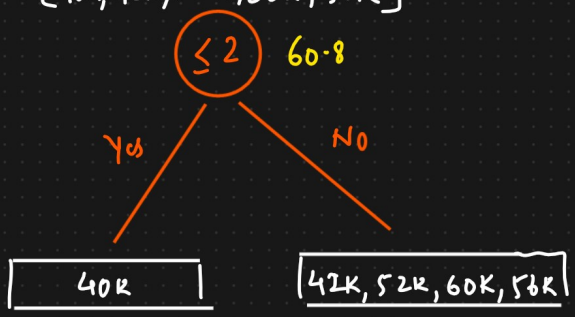
1. Entropy
2. G.I
3. Information Gain

DT Regressor

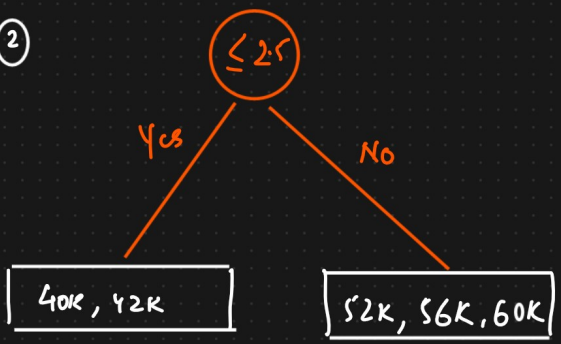
1. Variance Reduction
2. Variance

0.10

① [40K, 42K, 52K, 60K, 56K]



②



Final aim Variance Reduction

$$\text{Variance or Error} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y}_i)^2$$ [Mean Squared Error]

$$\text{Variance (Root)} = \frac{1}{5}\left[(40-50)^2+(42-50)^2+(52-50)^2 + (60-50)^2+(56-50)^2\right]$$

$$= \frac{1}{5}\left[100+64+4+100+36\right]$$

$$= 60.8$$

Var(Root) = 60.8

$$\text{Var(left)} = \frac{1}{2}\left[100+64\right]$$

$$= 82$$

$$\text{Var(Right)} = \frac{1}{3}\left[4+36+100\right]$$

$$= 46.66$$

Variance (Left) = $\frac{1}{1}\left[(40-50)^2\right]$

$= 100$ //.

Variance (Right) = $\frac{1}{4}\left[(64)+4+100+36\right]$.

$= 51$ //.

Variance Reduction : $Var(Root) - \sum w_i\, Var(child)$

$= 60.8 - \left[\frac{1}{5} * 100 + \frac{4}{5} * 51\right]$

Variance Reduction = 0

Variance Reduction =

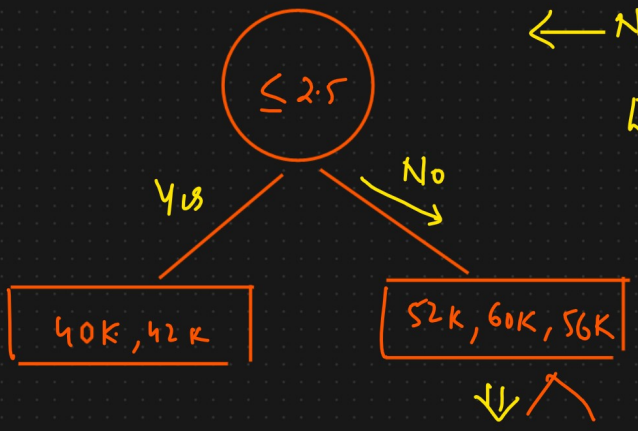
$60.8 - \left[\frac{2}{5} * 82 + \frac{3}{5} * 46.66\right]$

$= 0.004$

Variance Reduction < Variance Reduction

Select this split with feature.

Selected Split



← New Test Data

Exp = 3.5 years

$56K

$$\text{Average} = \frac{52 + 60 + 56}{3} = \boxed{56K}$$