

## Given Dataset

Feature 1	Feature 2	Target (Y)
5	A	10
7	A	20
6	B	15
8	B	25
9	C	30
10	C	35

## Step-by-Step Calculation

### 1. Initial Variance Calculation

Overall Mean and Variance of Target Values:

Mean (Overall Target):

$$\bar{Y} = \frac{10 + 20 + 15 + 25 + 30 + 35}{6} = 22.5$$

Variance (Overall):

$$= \frac{156.25 + 6.25 + 56.25 + 6.25 + 56.25 + 156.25}{6}$$

$$\sigma^2 = \frac{437.5}{6} = 72.9167$$

### 2. Split by Feature 1

We choose a threshold of 7 for **Feature 1**.

Group 1 (Feature 1 ≤ 7):

- Instances: 5, 7, 6
- Target Values: 10, 20, 15

**Variance for Group 1:**

$$= \frac{25 + 25 + 0}{3}$$
$$= \frac{50}{3} \approx 16.67$$

**Group 2 (Feature 1>7):**

- Instances: 8, 9, 10
- Target Values: 25, 30, 35

**Variance for Group 2:**

$$= \frac{25 + 0 + 25}{3}$$
$$= \frac{50}{3} \approx 16.67$$

**Weighted Variance for Feature 1:**

$$\text{Weighted Variance} = \frac{3}{6} \times 16.67 + \frac{3}{6} \times 16.67$$

$$\text{Weighted Variance} = 16.67$$

**Variance Reduction for Feature 1:**

$$\text{Variance Reduction}(\text{Feature 1}) = 72.92 - 16.67 = 56.25$$

**Evaluate Feature 2**

**Group A:**

- Instances: 5, 7
- Target Values: 10, 20

**Variance for Group A:**

Variance(Group A)= 25

**Group B:**

- Instances: 6, 8
- Target Values: 15, 25

**Variance for Group B:**

Variance(Group B)=25

**Group C:**

- Instances: 9, 10
- Target Values: 30, 35

**Variance for Group C:**

Variance(Group C)=26.25

**Weighted Variance for Feature 2:**

Weighted Variance(Split)= $2/6 \times 25 + 2/6 \times 25 + 2/6 \times 6.25 = 18.75$

**Variance Reduction for Feature 2:**

Variance Reduction(Feature 2)= $72.92 - 18.75 = 54.17$

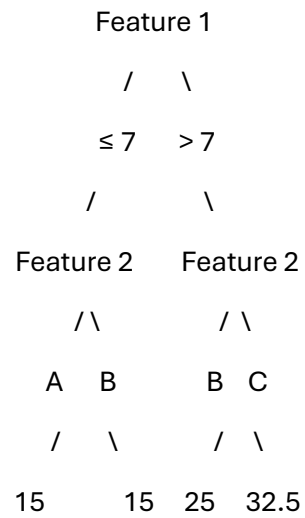
**4. Decision Tree Construction**

Since the variance reduction for **Feature 1** is higher than that for **Feature 2**, Feature 1 will be used as the root. We will then use **Feature 2** to further split the groups.

**Final Decision Tree Structure****Root Node: Feature 1  $\leq 7$** 

- **Left Branch** (Feature 1  $\leq 7$ ):
  - **Split by Feature 2:**
    - **Group A:** Target = 15
    - **Group B:** Target = 15
- **Right Branch** (Feature 1  $> 7$ ):
  - **Split by Feature 2:**
    - **Group B:** Target = 25
    - **Group C:** Target = 32.5

## Tree Diagram:



## Summary

### 1. Root Node: Feature 1 $\leq 7$

#### ○ Left Subtree:

- Feature 2 = A  $\rightarrow$  Target = 15
- Feature 2 = B  $\rightarrow$  Target = 15

#### ○ Right Subtree:

- Feature 2 = B  $\rightarrow$  Target = 25
- Feature 2 = C  $\rightarrow$  Target = 32.5