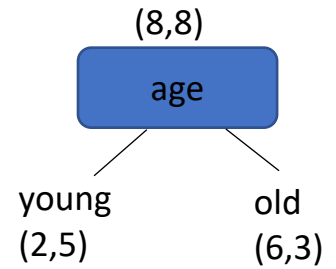


Q2)

Information Gain

root node:

Example: Age →



$$IG(S, \text{Age}) = - \overbrace{\frac{8}{16} \log_2 \left(\frac{8}{16} \right) - \frac{8}{16} \log_2 \left(\frac{8}{16} \right)}^{\text{Entropy(parent)}} - \frac{7}{16} \left(\overbrace{-\frac{2}{7} \log_2 \left(\frac{2}{7} \right) - \frac{5}{7} \log_2 \left(\frac{5}{7} \right)}^{\text{Entropy(child1:young)}} \right) - \frac{9}{16} \left(\overbrace{-\frac{6}{9} \log_2 \left(\frac{6}{9} \right) - \frac{3}{9} \log_2 \left(\frac{3}{9} \right)}^{\text{Entropy(child2:old)}} \right) = 0.422$$

Similarly we calculate the IG for other possible decision nodes:

$$IG(S, \text{Gender}) = 0.366$$

$$IG(S, \text{Smoker}) = 0.334$$

$$IG(S, \text{Chest Pain}) = 0.616$$

Chest Pain is the best option for “root” node based on IG.

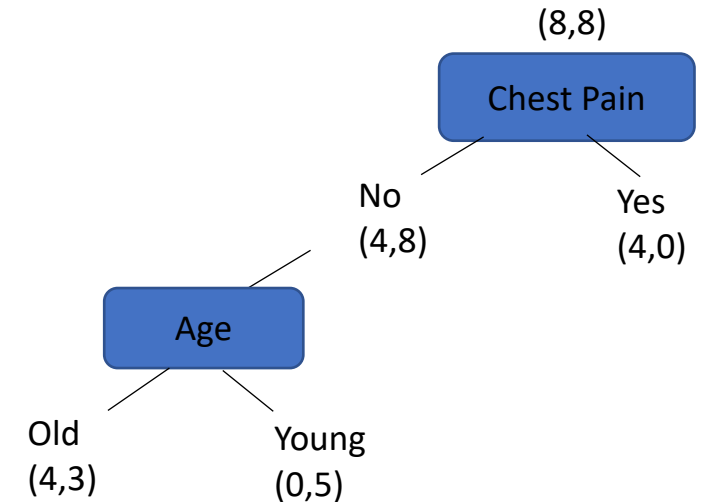
For “No” branch:

$$IG(\text{Chest-pain-NO}, \text{Gender}) = 0.477$$

$$IG(\text{Chest-pain-NO}, \text{Age}) = 0.708$$

$$IG(\text{Chest-pain-NO}, \text{Smoker}) = 0.401$$

Smoker is the best option for this node based on IG.

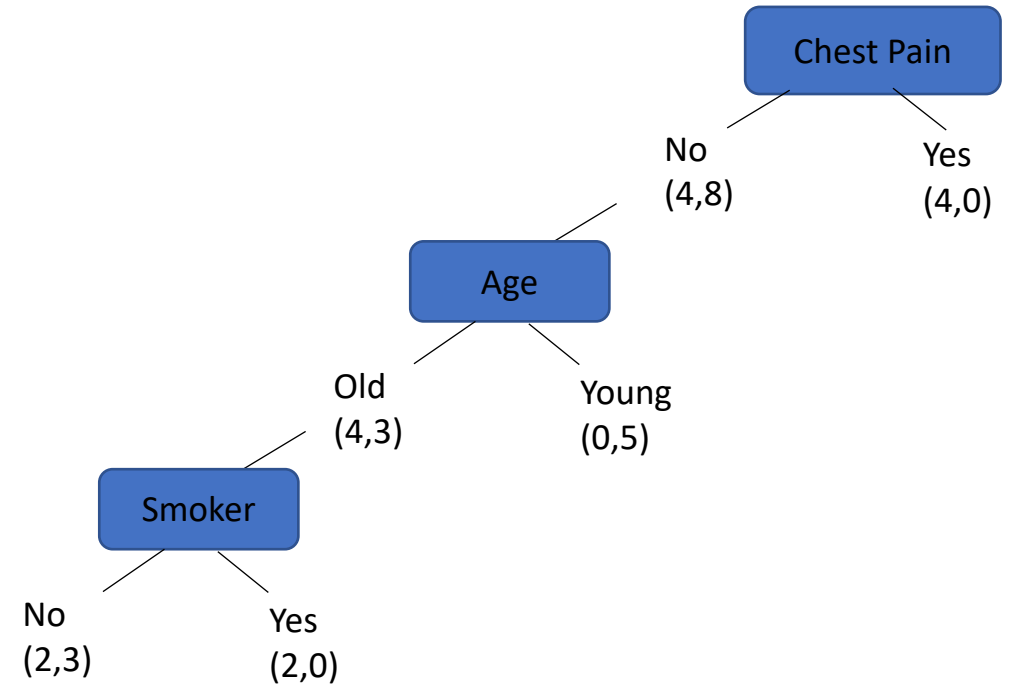


For “old” branch:

$$IG(\text{Old, Gender}) = 0.864$$

$$IG(\text{Old, Smoker}) = 0.991$$

Smoker is the best option for this node based on IG.



$$\text{Accuracy}_{\text{Training}} = 14/16 = 87.5\%$$

Gini Index

Gini index = $1 - \sum (p_i)^2$ (We would prefer choosing the attribute/feature with the least Gini index)
root node:

Gini Index(child1:No) Gini Index(child2:yes)

$$\text{Gini}(S, \text{Smoker}) = \frac{9}{16} \left(1 - \left(\left(\frac{4}{9} \right)^2 + \left(\frac{5}{9} \right)^2 \right) \right) + \frac{7}{16} \left(1 - \left(\left(\frac{4}{7} \right)^2 + \left(\frac{3}{7} \right)^2 \right) \right) = 0.492$$

$$\text{Gini}(S, \text{Gender}) = 0.469$$

$$\text{Gini}(S, \text{Age}) = 0.428$$

$$\text{Gini}(S, \text{Chest Pain}) = 0.333$$

Chest pain is the best option for “root” node based on Gini Index.

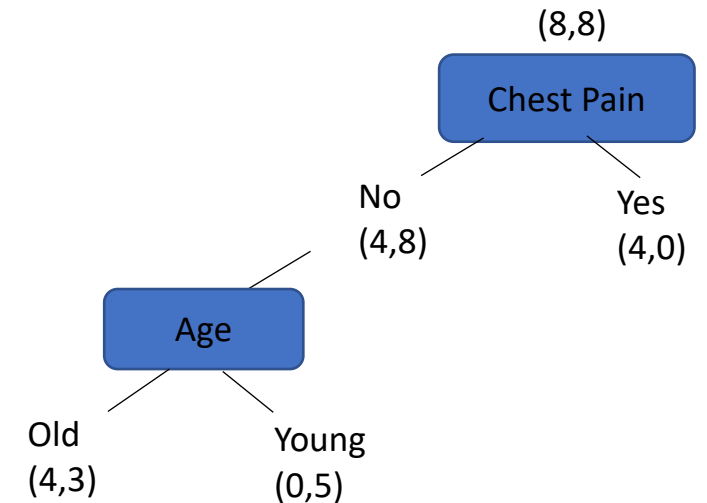
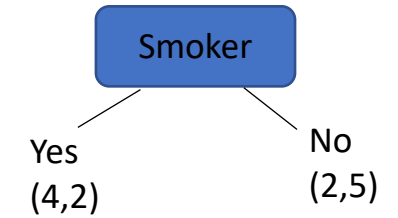
For “No” branch:

$$\text{Gini}(\text{Chest-pain-NO}, \text{Gender}) = 0.389$$

$$\text{Gini}(\text{Chest-pain-NO}, \text{Age}) = 0.286$$

$$\text{Gini}(\text{Chest-pain-NO}, \text{Smoker}) = 0.438$$

Age is the best option for this node based on Gini Index.

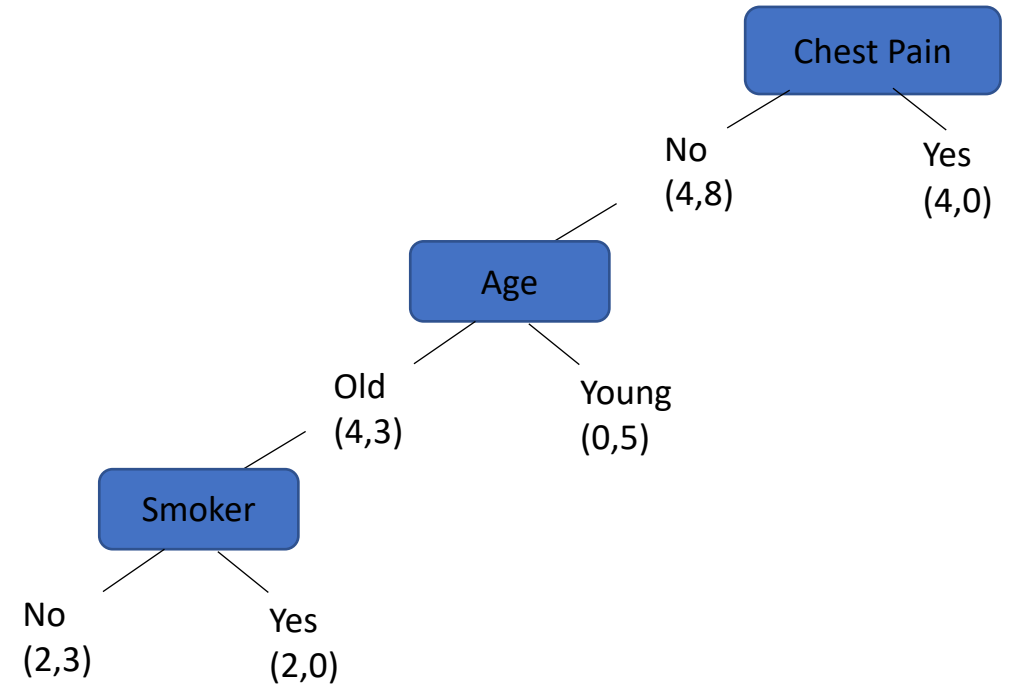


For “old” node:

$$IG(\text{old, Gender}) = 0.405$$

$$IG(\text{old, Smoker}) = 0.343$$

Smoker is the best option for this node based on Gini Index.



$$\text{Accuracy}_{\text{Training}} = 14/16 = 87.5\%$$