

# Decision Tree Induction

# Algorithm

- (The algorithm in the project description or slides)

# Example on Golf Data

## Attributes

## Target

Outlook	Temp	Humidity	Windy	Play
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

- First, determine the root node by calculating weighted averaged Gini index for each attribute

$$\text{Gini}(t) = 1 - \sum_{i=0}^{c-1} [p(i|t)]^2,$$

# Gini of Outlook

$$\text{Gini}(t) = 1 - \sum_{i=0}^{c-1} [p(i|t)]^2,$$

**Attributes**      **Target**

Outlook	Temp	Humidity	Windy	Play
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

$$\text{Gini}(\text{Rainy}) = 1 - (2/5)^2 - (3/5)^2$$

# Gini of Outlook

$$\text{Gini}(t) = 1 - \sum_{i=0}^{c-1} [p(i|t)]^2,$$

**Attributes**      **Target**

Outlook	Temp	Humidity	Windy	Play
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

$$\text{Gini}(\text{Rainy}) = 1 - (2/5)^2 - (3/5)^2$$

$$\text{Gini}(\text{Overcast}) = 1 - (4/4)^2 - (0/4)^2$$

# Gini of Outlook

$$\text{Gini}(t) = 1 - \sum_{i=0}^{c-1} [p(i|t)]^2,$$

**Attributes**      **Target**

Outlook	Temp	Humidity	Windy	Play
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

$$\text{Gini}(\text{Rainy}) = 1 - (2/5)^2 - (3/5)^2$$

$$\text{Gini}(\text{Overcast}) = 1 - (4/4)^2 - (0/4)^2$$

$$\text{Gini}(\text{Sunny}) = 1 - (3/5)^2 - (2/5)^2$$

# Gini of Outlook

**Attributes**      **Target**

Outlook	Temp	Humidity	Windy	Play
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

$$\text{Gini(Rainy)} = 1 - (2/5)^2 - (3/5)^2$$

$$\text{Gini(Overcast)} = 1 - (4/4)^2 - (0/4)^2$$

$$\text{Gini(Sunny)} = 1 - (3/5)^2 - (2/5)^2$$

$$\text{Gini(Outlook)} = (5/14) * \text{Gini(Rainy)} + (4/14) * \text{Gini(Overcast)} + (5/14) * \text{Gini(Sunny)}$$

# Gini of Temp

**Attributes**      **Target**

Outlook	Temp	Humidity	Windy	Play
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

$$\text{Gini(Hot)} = 1 - (2/4)^2 - (2/4)^2$$

$$\text{Gini(Mild)} = 1 - (2/6)^2 - (4/6)^2$$

$$\text{Gini(Cool)} = 1 - (3/4)^2 - (1/4)^2$$

$$\begin{aligned} \text{Gini(Temp)} &= (4/14)^* \text{Gini(Hot)} + (6/14)^* \\ &\text{Gini(Mild)} + (4/14)^* \text{Gini(Cool)} \end{aligned}$$



# Gini of Humidity

**Attributes**      **Target**

Outlook	Temp	Humidity	Windy	Play
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

$$\text{Gini(High)} = 1 - (4/7)^2 - (3/7)^2$$

$$\text{Gini(Normal)} = 1 - (1/7)^2 - (6/7)^2$$

$$\text{Gini(Humidity)} = (7/14)^* \text{Gini(High)} + (7/14)^* \text{Gini(Normal)}$$

# Gini of Windy

**Attributes**      **Target**

Outlook	Temp	Humidity	Windy	Play
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

$$\text{Gini(False)} = 1 - (6/8)^2 - (2/8)^2$$

$$\text{Gini(True)} = 1 - (3/6)^2 - (3/6)^2$$

$$\text{Gini(Humidity)} = (8/14) * \text{Gini(False)} + (6/14) * \text{Gini(True)}$$

# Example on Golf Data

## Attributes

## Target

Outlook	Temp	Humidity	Windy	Play
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

- Determine the root node by calculating weighted averaged Gini index for each attribute

$$\text{Gini(Outlook)}=0.3428$$

$$\text{Gini(Temp)}=0.4405$$

$$\text{Gini(Humidity)}=0.3673$$

$$\text{Gini(Windy)}=0.4286$$

# Example on Golf Data

## Attributes

## Target

Outlook	Temp	Humidity	Windy	Play
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

- Determine the root node by calculating weighted averaged Gini index for each attribute

$$\text{Gini(Outlook)}=0.3428$$

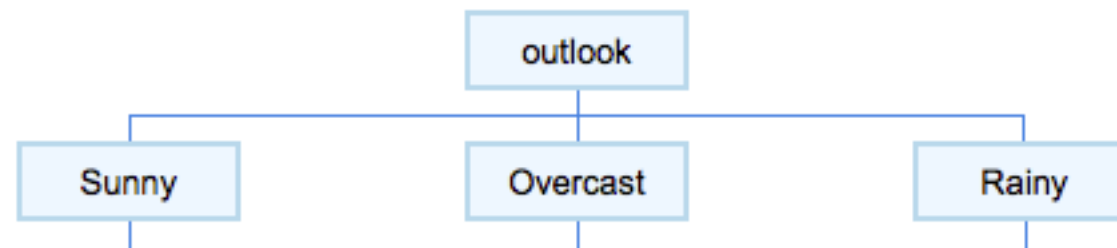
**Choose the minimum**

$$\text{Gini(Temp)}=0.4405$$

$$\text{Gini(Humidity)}=0.3673$$

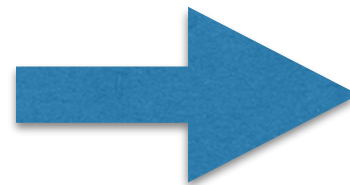
$$\text{Gini(Windy)}=0.4286$$

# Determine the Root Node



# Determine the Split Attribute when Outlook=Sunny

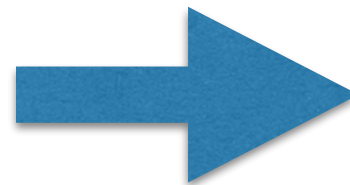
Outlook	Temp	Humidity	Windy	Play
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No



Outlook	Temp	Humidity	Windy	Play
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Sunny	Mild	Normal	False	Yes
Sunny	Mild	High	True	No

# Determine the Split Attribute when Outlook=Sunny

Outlook	Temp	Humidity	Windy	Play
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No



Outlook	Temp	Humidity	Windy	Play
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Sunny	Mild	Normal	False	Yes
Sunny	Mild	High	True	No

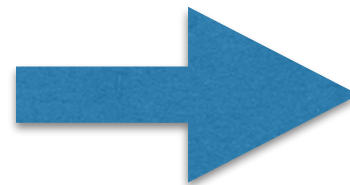
Gini(Temp)=?

Gini(Humidity)=?

Gini(Windy)=?

# Determine the Split Attribute when Outlook=Sunny

Outlook	Temp	Humidity	Windy	Play
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No



Outlook	Temp	Humidity	Windy	Play
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Sunny	Mild	Normal	False	Yes
Sunny	Mild	High	True	No

Gini(Temp)=?

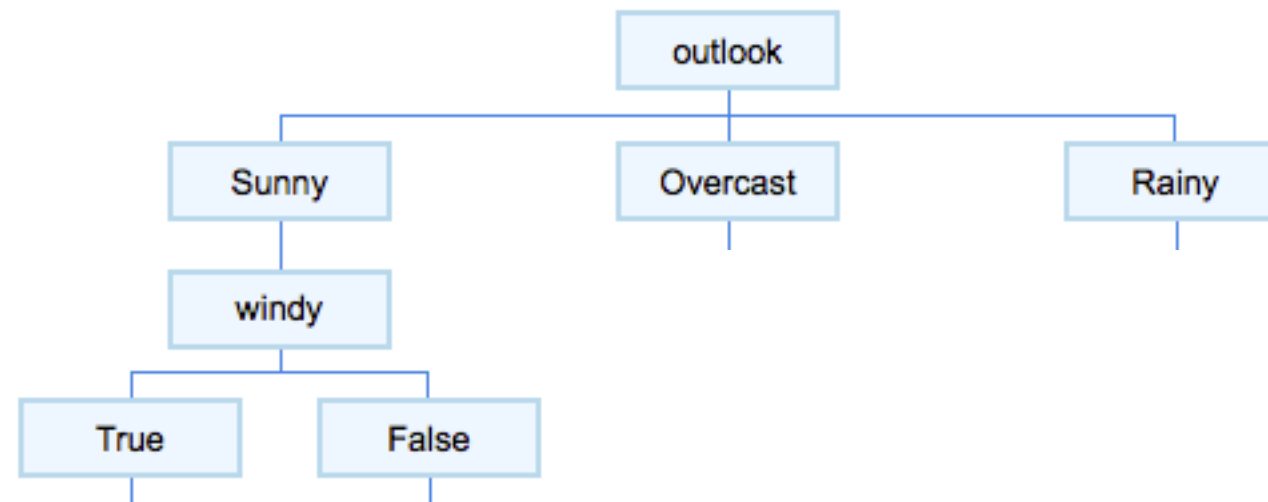
Gini(Humidity)=?

Gini(Windy)=?

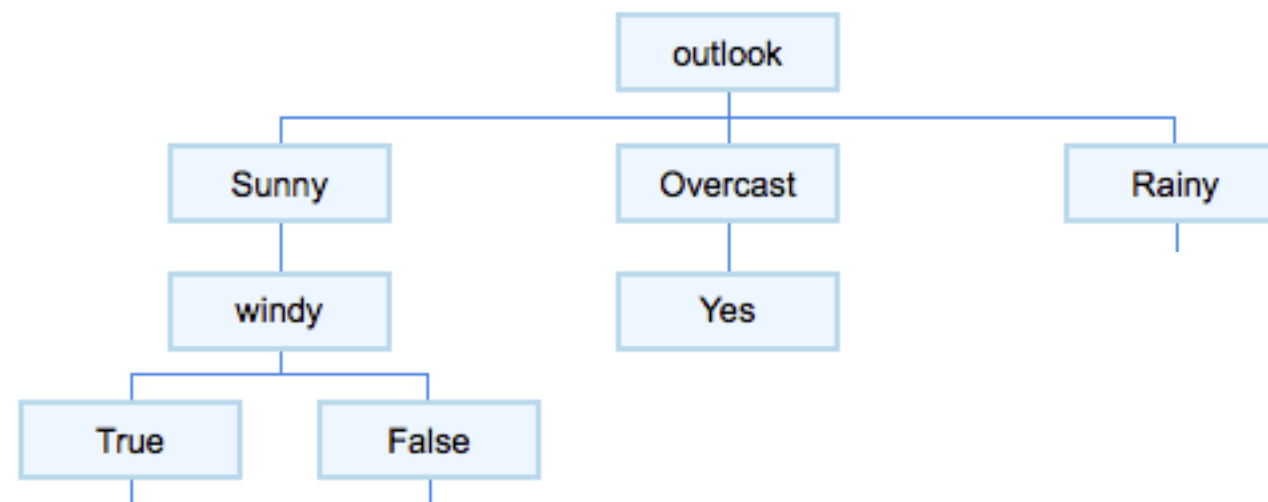
**Choose the minimum**



# Determine the Split Attribute when Outlook=Sunny



# Determine the Split Attribute when Outlook=Overcast

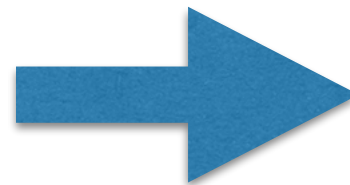


Outlook	Temp	Humidity	Windy	Play
Overcast	Hot	High	False	Yes
Overcast	Cool	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes

**This is a leaf node  
since all target labels  
are the same**

# Determine the Split Attribute when Outlook=Rainy

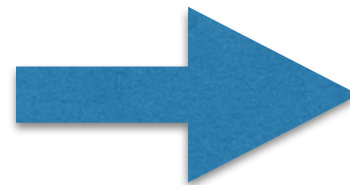
Outlook	Temp	Humidity	Windy	Play
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No



Outlook	Temp	Humidity	Windy	Play
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Rainy	Mild	Normal	True	Yes

# Determine the Split Attribute when Outlook=Rainy

Outlook	Temp	Humidity	Windy	Play
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No



Outlook	Temp	Humidity	Windy	Play
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Rainy	Mild	Normal	True	Yes

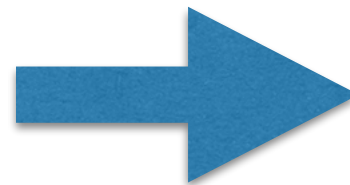
Gini(Temp)=?

Gini(Humidity)=?

Gini(Windy)=?

# Determine the Split Attribute when Outlook=Rainy

Outlook	Temp	Humidity	Windy	Play
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No



Outlook	Temp	Humidity	Windy	Play
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Rainy	Mild	Normal	True	Yes

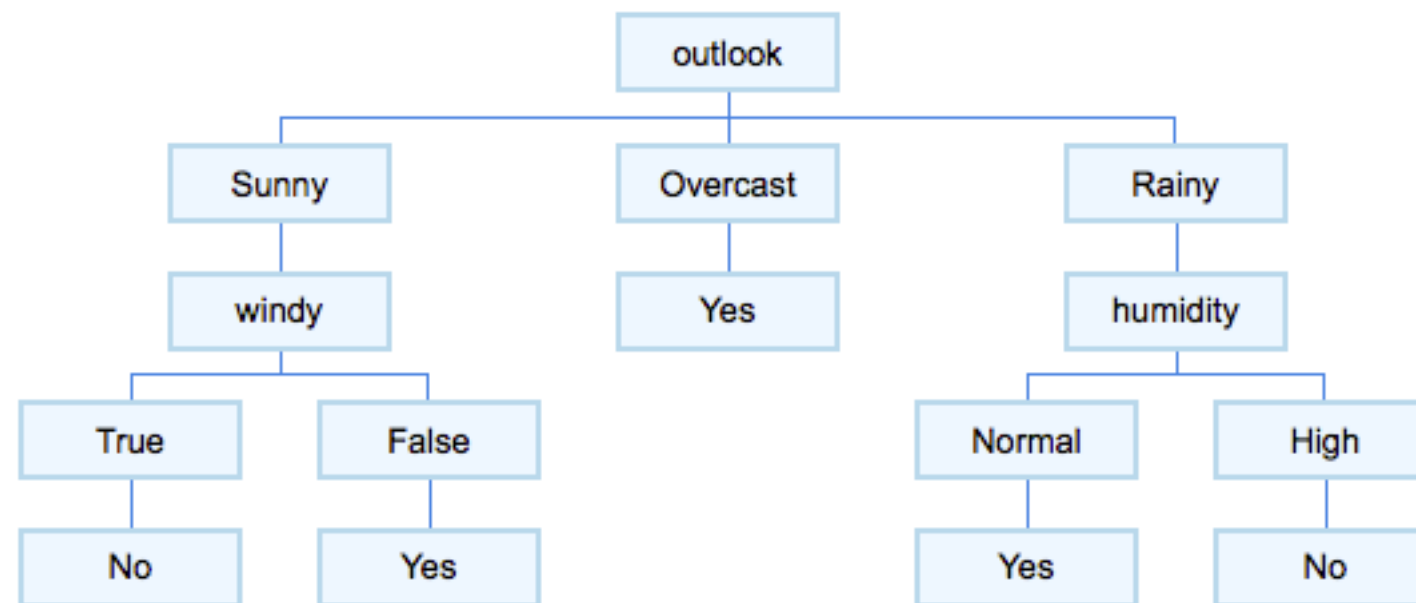
Gini(Temp)=?

Gini(Humidity)=?

Gini(Windy)=?

**Choose the minimum**

# Keep going until reach the stopping condition



Stopping condition:

1. All the records have the same class label

OR

2. There is no further attributes can be used to split, then take the **majority** as the class label