Decision Tree Example

Say we want to create a decision tree using the ID3 algorithm to decide if our friend will want to play tennis given some features relating to the weather. From previous experience we have the following data:

Outlook	Humidity	Play Tennis?
Sunny	High	No
Sunny	High	No
Overcast	High	Yes
Rain	High	Yes
Rain	Normal	Yes
Rain	Normal	No
Overcast	Normal	Yes
Sunny	High	No
Sunny	Normal	Yes
Rain	Normal	Yes
Sunny	Normal	Yes
Overcast	High	Yes
Overcast	Normal	Yes
Rain	High	No

So our features are Outlook and Humidity and our label is Play Tennis.

To construct a Decision Tree we are going to need to start by figuring out which node we should put at the root of our tree. Recall to do this we want to see which split on which feature will maximize information gain:

$$I(Y;X) = H(Y) - H(Y|X)$$

Decision Tree Example

Say we want to create a decision tree using the ID3 algorithm to decide if our friend will want to play tennis given some features relating to the weather. From previous experience we have the following data:

Outlook	Humidity	Play Tennis?
Sunny	High	No
Sunny	High	No
Overcast	High	Yes
Rain	High	Yes
Rain	Normal	Yes
Rain	Normal	No
Overcast	Normal	Yes
Sunny	High	No
Sunny	Normal	Yes
Rain	Normal	Yes
Sunny	Normal	Yes
Overcast	High	Yes
Overcast	Normal	Yes
Rain	High	No

So our features are Outlook and Humidity and our label is Play Tennis.

To construct a Decision Tree we are going to need to start by figuring out which node we should put at the root of our tree. Recall to do this we want to see which split on which feature will maximize information gain:

$$I(Y;X) = H(Y) - H(Y|X)$$

$What\ is\ the\ equation\ for\ H(Y)?$

Decision Tree Example

Say we want to create a decision tree using the ID3 algorithm to decide if our friend will want to play tennis given some features relating to the weather. From previous experience we have the following data:

Outlook	Humidity	Play Tennis?
Sunny	High	No
Sunny	High	No
Overcast	High	Yes
Rain	High	Yes
Rain	Normal	Yes
Rain	Normal	No
Overcast	Normal	Yes
Sunny	High	No
Sunny	Normal	Yes
Rain	Normal	Yes
Sunny	Normal	Yes
Overcast	High	Yes
Overcast	Normal	Yes
Rain	High	No

So our features are Outlook and Humidity and our label is Play Tennis.

To construct a Decision Tree we are going to need to start by figuring out which node we should put at the root of our tree. Recall to do this we want to see which split on which feature will maximize information gain:

$$I(Y;X) = H(Y) - H(Y|X)$$

$$H(Y) = -\sum_{y} P(Y = y)log_2(P(Y = y))$$

Decision Tree Example

Say we want to create a decision tree using the ID3 algorithm to decide if our friend will want to play tennis given some features relating to the weather. From previous experience we have the following data:

Outlook	Humidity	Play Tennis?
Sunny	High	No
Sunny	High	No
Overcast	High	Yes
Rain	High	Yes
Rain	Normal	Yes
Rain	Normal	No
Overcast	Normal	Yes
Sunny	High	No
Sunny	Normal	Yes
Rain	Normal	Yes
Sunny	Normal	Yes
Overcast	High	Yes
Overcast	Normal	Yes
Rain	High	No

So our features are Outlook and Humidity and our label is Play Tennis.

To construct a Decision Tree we are going to need to start by figuring out which node we should put at the root of our tree. Recall to do this we want to see which split on which feature will maximize information gain:

$$I(Y;X) = H(Y) - H(Y|X)$$

$$H(Y) = -\sum_{y} P(Y = y) log_2(P(Y = y))$$

$What\ is\ the\ equation\ for\ H(Y|X)?$

Decision Tree Example

Say we want to create a decision tree using the ID3 algorithm to decide if our friend will want to play tennis given some features relating to the weather. From previous experience we have the following data:

Outlook	Humidity	Play Tennis?
Sunny	High	No
Sunny	High	No
Overcast	High	Yes
Rain	High	Yes
Rain	Normal	Yes
Rain	Normal	No
Overcast	Normal	Yes
Sunny	High	No
Sunny	Normal	Yes
Rain	Normal	Yes
Sunny	Normal	Yes
Overcast	High	Yes
Overcast	Normal	Yes
Rain	High	No

So our features are Outlook and Humidity and our label is Play Tennis.

To construct a Decision Tree we are going to need to start by figuring out which node we should put at the root of our tree. Recall to do this we want to see which split on which feature will maximize information gain:

$$I(Y;X) = H(Y) - H(Y|X)$$

$$H(Y) = -\sum_{y} P(Y = y)log_2(P(Y = y))$$

$$H(Y|X) = -\sum_{x} P(X = x) \sum_{y} P(Y = y|X = x)\, log_2(P(Y = y|X = x))$$

Decision Tree Example

Let's start by looking at the information gained if we choose to split on Outlook:

Outlook	Play Tennis?
Sunny	No
Sunny	No
Overcast	Yes
Rain	Yes
Rain	Yes
Rain	No
Overcast	Yes
Sunny	No
Sunny	Yes
Rain	Yes
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rain	No

First let's calculate $H(Y)$:

Decision Tree Example

Let's start by looking at the information gained if we choose to split on Outlook:

Outlook	Play Tennis?
Sunny	No
Sunny	No
Overcast	Yes
Rain	Yes
Rain	Yes
Rain	No
Overcast	Yes
Sunny	No
Sunny	Yes
Rain	Yes
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rain	No

First let's calculate *H(Y)*:

$$H(Y) = -\sum_{y} P(Y = y) log_2\big(P(Y = y)\big)$$

Decision Tree Example

Let's start by looking at the information gained if we choose to split on Outlook:

Outlook	Play Tennis?
Sunny	No
Sunny	No
Overcast	Yes
Rain	Yes
Rain	Yes
Rain	No
Overcast	Yes
Sunny	No
Sunny	Yes
Rain	Yes
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rain	No

First let's calculate *H(Y)*:

$$H(Y) = -\sum_{y} P(Y = y) log_2\big(P(Y = y)\big)$$

$$= -P(Y = \text{"}No\text{"}) log_2\big(P(Y = \text{"}No\text{"})\big) - P(Y = \text{"}Yes\text{"}) log_2(P(Y = \text{"}Yes\text{"}))$$

Decision Tree Example

Let's start by looking at the information gained if we choose to split on Outlook:

Outlook	Play Tennis?
Sunny	No
Sunny	No
Overcast	Yes
Rain	Yes
Rain	Yes
Rain	No
Overcast	Yes
Sunny	No
Sunny	Yes
Rain	Yes
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rain	No

First let's calculate *H(Y)*:

$$H(Y) = -\sum_{y} P(Y = y) log_2(P(Y = y))$$

$$= -P(Y = "No")log_2(P(Y = "No")) - P(Y = "Yes")log_2(P(Y = "Yes"))$$

$$= -\frac{5}{14} log_2\left(\frac{5}{14}\right) - \frac{9}{14} log_2\left(\frac{9}{14}\right) = 0.9403$$

Since we have 5 'No' values in our 'Play Tennis?' Column

Since we have 9 'Yes' values in our 'Play Tennis?' Column

Decision Tree Example

Let’s start by looking at the information gained if we choose to split on Outlook:

Outlook	Play Tennis?
Sunny	No
Sunny	No
Overcast	Yes
Rain	Yes
Rain	Yes
Rain	No
Overcast	Yes
Sunny	No
Sunny	Yes
Rain	Yes
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rain	No

Next let’s calculate $H(Y|X)$:

Decision Tree Example

Let's start by looking at the information gained if we choose to split on Outlook:

Outlook	Play Tennis?
Sunny	No
Sunny	No
Overcast	Yes
Rain	Yes
Rain	Yes
Rain	No
Overcast	Yes
Sunny	No
Sunny	Yes
Rain	Yes
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rain	No

Next let's calculate *H(Y|X)*:

$$H(Y|X) = -\sum_x P(X = x) \sum_y P(Y = y|X = x)\, log_2(P(Y = y|X = x))$$

Decision Tree Example

Let's start by looking at the information gained if we choose to split on Outlook:

Outlook	Play Tennis?
Sunny	No
Sunny	No
Overcast	Yes
Rain	Yes
Rain	Yes
Rain	No
Overcast	Yes
Sunny	No
Sunny	Yes
Rain	Yes
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rain	No

Next let's calculate *H(Y|X)*:

$$H(Y|X) = -\sum_x P(X=x) \sum_y P(Y=y|X=x)\, log_2(P(Y=y|X=x))$$

$$= -P(X = \text{"}Sunny\text{"})[P(Y = \text{"}No\text{"}|X = \text{"}Sunny\text{"})log_2\big(P(Y = \text{"}No\text{"}|X = \text{"}Sunny\text{"})\big)$$
$$+P(Y = \text{"}Yes\text{"}|X = \text{"}Sunny\text{"})log_2(P(Y = \text{"}Yes\text{"}|X = \text{"}Sunny\text{"})]$$

$$-P(X = \text{"}Overcast\text{"})[P(Y = \text{"}No\text{"}|X = \text{"}Overcast\text{"})log_2\big(P(Y = \text{"}No\text{"}|X = \text{"}Overcast\text{"})\big)$$
$$+P(Y = \text{"}Yes\text{"}|X = \text{"}Overcast\text{"})log_2(P(Y = \text{"}Yes\text{"}|X = \text{"}Overcast\text{"})]$$

$$-P(X = \text{"}Rain\text{"})[P(Y = \text{"}No\text{"}|X = \text{"}Rain\text{"})log_2\big(P(Y = \text{"}No\text{"}|X = \text{"}Rain\text{"})\big)$$
$$+P(Y = \text{"}Yes\text{"}|X = \text{"}Rain\text{"})log_2(P(Y = \text{"}Yes\text{"}|X = \text{"}Rain\text{"})]$$

Decision Tree Example

Let's start by looking at the information gained if we choose to split on Outlook:

Outlook	Play Tennis?
Sunny	No
Sunny	No
Overcast	Yes
Rain	Yes
Rain	Yes
Rain	No
Overcast	Yes
Sunny	No
Sunny	Yes
Rain	Yes
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rain	No

Next let's calculate *H(Y|X)*:

$$H(Y|X) = -\sum_x P(X=x) \sum_y P(Y=y|X=x)\, log_2(P(Y=y|X=x))$$

$$= -\frac{5}{14}[\frac{3}{5} log_2\left(\frac{3}{5}\right) + \frac{2}{5} log_2\left(\frac{2}{5}\right)]$$

$$-\frac{4}{14}[\frac{0}{4} log_2\left(\frac{0}{4}\right) + \frac{4}{4} log_2\left(\frac{4}{4}\right)]$$

$$-\frac{5}{14}[\frac{2}{5} log_2\left(\frac{2}{5}\right) + \frac{3}{5} log_2\left(\frac{3}{5}\right)]$$

=0.6935

Decision Tree Example

Let's start by looking at the information gained if we choose to split on Outlook:

Outlook	Play Tennis?
Sunny	No
Sunny	No
Overcast	Yes
Rain	Yes
Rain	Yes
Rain	No
Overcast	Yes
Sunny	No
Sunny	Yes
Rain	Yes
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rain	No

Finally let's calculate *I(Y;X)*:

Decision Tree Example

Let's start by looking at the information gained if we choose to split on Outlook:

Outlook	Play Tennis?
Sunny	No
Sunny	No
Overcast	Yes
Rain	Yes
Rain	Yes
Rain	No
Overcast	Yes
Sunny	No
Sunny	Yes
Rain	Yes
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rain	No

Finally let's calculate *I(Y;X)*:

$$I(Y;X) = H(Y) - H(Y|X)$$

$$= 0.9403 - 0.6935 = 0.2468$$

What do we do next?

Decision Tree Example

Let's start by looking at the information gained if we choose to split on Outlook:

Outlook	Play Tennis?
Sunny	No
Sunny	No
Overcast	Yes
Rain	Yes
Rain	Yes
Rain	No
Overcast	Yes
Sunny	No
Sunny	Yes
Rain	Yes
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rain	No

Finally let's calculate *I(Y;X)*:

$$I(X;Y) = H(Y) - H(Y|X)$$

$$= 0.9403 - 0.6935 = 0.2468$$

What do we do next?

We simply repeat for the other attribute.

Decision Tree Example

Calculating information gain from splitting on Humidity:

Humidity	Play Tennis?
High	No
High	No
High	Yes
High	Yes
Normal	Yes
Normal	No
Normal	Yes
High	No
Normal	Yes
Normal	Yes
Normal	Yes
High	Yes
Normal	Yes
High	No

First we can notice that *H(Y)*:

$$H(Y) = -\sum_{y} P(Y = y) log_2\big(P(Y = y)\big)$$

$$= -P(Y = "No") log_2\big(P(Y = "No")\big) - P(Y = "Yes") log_2(P(Y = "Yes"))$$

$$= -\frac{5}{14} log_2\left(\frac{5}{14}\right) - \frac{9}{14} log_2\left(\frac{9}{14}\right) = 0.9403$$

Is exactly the same

Decision Tree Example

Calculating information gain from splitting on Humidity:

Humidity	Play Tennis?
High	No
High	No
High	Yes
High	Yes
Normal	Yes
Normal	No
Normal	Yes
High	No
Normal	Yes
Normal	Yes
Normal	Yes
High	Yes
Normal	Yes
High	No

Next let's calculate *H(Y|X)*:

$$H(Y|X) = -\sum_x P(X=x) \sum_y P(Y=y|X=x)\, log_2(P(Y=y|X=x))$$

$$= -P(X = "High")[P(Y = "No"|X = "High")log_2\big(P(Y = "No"|X = "High")\big)$$
$$+P(Y = "Yes"|X = "High")log_2(P(Y = "Yes"|X = "High")]$$

$$-P(X = "Normal")[P(Y = "No"|X = "Normal")log_2\big(P(Y = "No"|X = "Normal")\big)$$
$$+P(Y = "Yes"|X = "Normal")log_2(P(Y = "Yes"|X = "Normal")]$$

Decision Tree Example

Calculating information gain from splitting on Humidity:

Humidity	Play Tennis?
High	No
High	No
High	Yes
High	Yes
Normal	Yes
Normal	No
Normal	Yes
High	No
Normal	Yes
Normal	Yes
Normal	Yes
High	Yes
Normal	Yes
High	No

Next let's calculate *H(Y|X)*:

$$H(Y|X) = -\sum_x P(X=x) \sum_y P(Y=y|X=x)\, log_2(P(Y=y|X=x))$$

$$= -\frac{7}{14}[\frac{4}{7} log_2\left(\frac{4}{7}\right) + \frac{3}{7} log_2\left(\frac{3}{7}\right)]$$

$$-\frac{7}{14}[\frac{1}{7} log_2\left(\frac{1}{7}\right) + \frac{6}{7} log_2\left(\frac{6}{7}\right)]$$

=0.7885

Decision Tree Example

Calculating information gain from splitting on Humidity:

Humidity	Play Tennis?
High	No
High	No
High	Yes
High	Yes
Normal	Yes
Normal	No
Normal	Yes
High	No
Normal	Yes
Normal	Yes
Normal	Yes
High	Yes
Normal	Yes
High	No

Finally let's calculate *I(Y;X)*:

$$I(Y;X) = H(Y) - H(Y|X)$$

$$= 0.9403 - 0.7885 = 0.1518$$

Decision Tree Example

So which should we pick?

Outlook	Humidity	Play Tennis?
Sunny	High	No
Sunny	High	No
Overcast	High	Yes
Rain	High	Yes
Rain	Normal	Yes
Rain	Normal	No
Overcast	Normal	Yes
Sunny	High	No
Sunny	Normal	Yes
Rain	Normal	Yes
Sunny	Normal	Yes
Overcast	High	Yes
Overcast	Normal	Yes
Rain	High	No

$I(Y;\text{"Outlook"}) = 0.2468$

$\mathrm{I(Y;}\text{"}Humidity\text{"})=0.1518$

Decision Tree Example

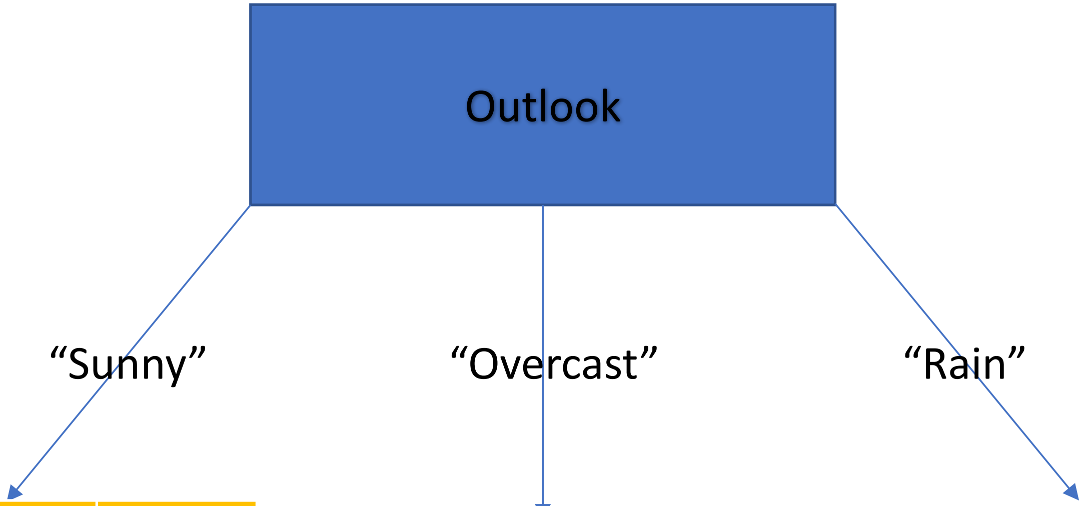
So which should we pick? Well we want to maximize the information gain with this split. So we should split on Outlook since:

Outlook	Humidity	Play Tennis?
Sunny	High	No
Sunny	High	No
Overcast	High	Yes
Rain	High	Yes
Rain	Normal	Yes
Rain	Normal	No
Overcast	Normal	Yes
Sunny	High	No
Sunny	Normal	Yes
Rain	Normal	Yes
Sunny	Normal	Yes
Overcast	High	Yes
Overcast	Normal	Yes
Rain	High	No

$$I(Y;\text{"Outlook"}) = 0.2468 > 0.1518 = \mathrm{I}(\mathrm{Y};\text{"}Humidity\text{"})$$

Decision Tree Example

So we have that Outlook is our root node, what next? Well we notice that the data gets split as follows:



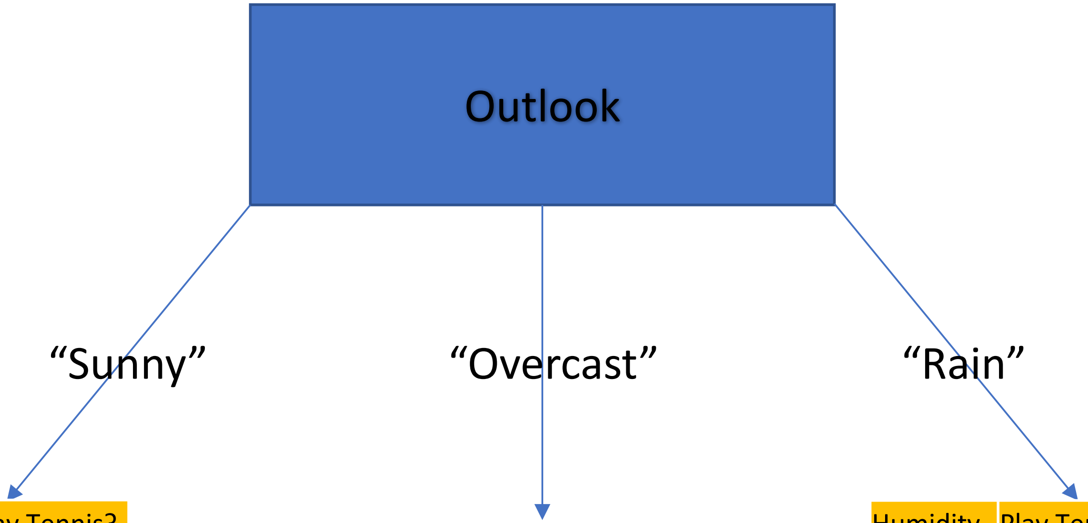
Outlook	Humidity	Play Tennis?
Sunny	High	No
Sunny	High	No
Sunny	High	No
Sunny	Normal	Yes
Sunny	Normal	Yes

Outlook	Humidity	Play Tennis?
Overcast	High	Yes
Overcast	Normal	Yes
Overcast	High	Yes
Overcast	Normal	Yes

Outlook	Humidity	Play Tennis?
Rain	High	Yes
Rain	Normal	Yes
Rain	Normal	No
Rain	Normal	Yes
Rain	High	No

Decision Tree Example

We can now ignore the outlook column since it is constant in each of the splits



Humidity	Play Tennis?
High	No
High	No
High	No
Normal	Yes
Normal	Yes

Humidity	Play Tennis?
High	Yes
Normal	Yes
High	Yes
Normal	Yes

Humidity	Play Tennis?
High	Yes
Normal	Yes
Normal	No
Normal	Yes
High	No

Decision Tree Example

Then we just repeat the process over again to decide which feature we put next in the tree. For the 3 sub datasets that we have:

Humidity	Play Tennis?
High	No
High	No
High	No
Normal	Yes
Normal	Yes

Humidity	Play Tennis?
High	Yes
Normal	Yes
High	Yes
Normal	Yes

Humidity	Play Tennis?
High	Yes
Normal	Yes
Normal	No
Normal	Yes
High	No

Return to page 2 of this slide and replace the original dataset with one of these subsets and follow the guide again.