Week 2 - Tutorial 1

Decision Tree

Luke Chang

The University of Auckland

Feb, 2022

Decision Tree Overview

- A decision tree models different possible decision paths, where each decision node is a conditional.
- Decision nodes are created based on a **splitting criterion**.
- Most common splitting criterions are: **Entropy** and **Gini**.
- A tree is created recursively from the root node to the leaf nodes.
- Decision tree can be used for both classification and regression.

Entropy

Informally, entropy measures the amount of uncertainty in a system.

Shannon Entropy

$$H(X) := -\sum_x P(X = x) \log_2 P(X = x)$$

Where $0 \log_2(0) \equiv 0$, since $\lim_{x \to 0} x \log_2(x) = 0$.

- What is the possible maximum entropy?
- What does it mean?
- What is the possible minimum entropy?
- What does it mean?

Example: A six-sided Die

Suppose we have a standard six-sided die with each of the faces has a probability of $1/6$.

$$P(X=i)=\frac{1}{6}, i \in \{1,2,3,4,5,6\}$$

Entropy:

$$H(X)=-\sum \frac{1}{6}\log_2(\frac{1}{6})=-6[\frac{1}{6}\log_2(\frac{1}{6})]=\log_2(6)\approx 2.6$$

Suppose we have a loaded six-sided die. All 6 faces are printed “1”.

$$P(X=i)=\begin{cases}1 & \text{if } i=1,\\ 0 & otherwise.\end{cases}$$

Entropy:

$$H(X)=-log_2(1)=0$$

Information Gain

Information Gain

Information Gain = Parent Entropy - Current Conditional Entropy

$$\mathrm{IG}(Y,X) := H(Y) - H(Y|X)$$

Where $H(Y|X)$ is the conditional entropy of the target variable Y given attribute X (Weighted sum):

$$H(Y|X) := \sum_x P(X=x)H(Y|X=x)$$

And $H(Y|X=x)$ is the conditional entropy of the target variable Y given $X=x$:

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

Estimating Probabilities

Let D_n be the subset of the data at node n in the tree, we estimate the probability by counting the number of success:

$$P(Y = y) := \frac{|\{i \in D_n : i_Y = y\}|}{|D_n|}$$

Similarly:

$$P(X = x) := \frac{|\{i \in D_n : i_X = x\}|}{|D_n|}$$

Conditional probability:

$$P(Y = y|X = x) := \frac{|\{i \in D_n : i_Y = y \wedge i_X = x\}|}{|\{i \in D_n : i_X = x\}|}$$

Example 1: Boolean Functions

Give decision trees to represent the following boolean functions:

1. $A \wedge \neg B$

A	B	¬B	Y
0	0	1	0
1	0	1	1
0	1	0	0
1	1	0	0

$A \wedge \neg B$ Probabilities

A	B	$\neg$B	Y
0	0	1	0
1	0	1	1
0	1	0	0
1	1	0	0

$$P(Y=1) = \frac{1}{4}$$
$$P(Y=1|A=1) = \frac{1}{2}$$
$$P(Y=1|A=0) = 0$$
$$P(Y=1|B=1) = 0$$
$$P(Y=1|B=0) = \frac{1}{2}$$

Root Entropy:

$$\begin{aligned} H(Y) &= -P(Y=1)\log_2 P(Y=1) - P(Y=0)\log_2 P(Y=0) \\ &= -\frac{1}{4}\log_2(\frac{1}{4}) - \frac{3}{4}\log_2(\frac{3}{4}) \\ &\approx 0.5 + 0.31 = 0.81 \end{aligned}$$

Entropies Condition on A

$$
\begin{aligned}
H(Y|A=1) &= -P(Y=1|A=1)\log_2 P(Y=1|A=1) - P(Y=0|A=1)\log_2 P(Y=0|A=1) \\
&= -\frac{1}{2}\log_2(\frac{1}{2}) - \frac{1}{2}\log_2(\frac{1}{2}) \\
&= 1
\end{aligned}
$$

$$
\begin{aligned}
H(Y|A=0) &= -P(Y=1|A=0)\log_2 P(Y=1|A=0) - P(Y=0|A=0)\log_2 P(Y=0|A=0) \\
&= -\frac{0}{2}\log_2(\frac{0}{2}) - \frac{2}{2}\log_2(\frac{2}{2}) \\
&= 0 - 0 = 0
\end{aligned}
$$

$$
\begin{aligned}
IG(Y|A) &= H(Y) - P(A=1)H(Y|A=1) - P(A=0)H(Y|A=0) \\
&\approx 0.81 - \frac{2}{4} - 0 = 0.31
\end{aligned}
$$

Entropies Condition on B

By symmetry:

$$
\begin{aligned}
H(Y|B=1) &= -P(Y=1|B=1)\log_2 P(Y=1|B=1) - P(Y=0|B=1)\log_2 P(Y=0|B=1) \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
H(Y|B=0) &= -P(Y=1|B=0)\log_2 P(Y=1|B=0) - P(Y=0|B=0)\log_2 P(Y=0|B=0) \\
&= 1
\end{aligned}
$$

$$
\begin{aligned}
IG(Y|B) &= H(Y) - P(B=1)H(Y|B=1) - P(B=0)H(Y|B=0) \\
&\approx 0.81 - 0 - \frac{2}{4} = 0.31
\end{aligned}
$$

$$IG(Y|A) \approx 0.31$$
$$IG(Y|B) \approx 0.31$$

There is a tie on both conditions. Let's use A as root node.

Draw Decision Tree using `sklearn`