Decision Trees – Classification

Intro

How do decision trees know which features to split from? How are the root node and subsequent nodes chosen?

Decision trees is a greedy algorithm which optimizes how nodes are split through maximization of information gain. This is done through searching for the feature with smallest entropy.

Definitions

Entropy: a measure of randomness – a measure to minimize

$$\mathrm{E(X)} = \sum_{\mathrm{i=1}}^{\mathrm{n}} -p_i \log_2 p_i$$

where n is the number of classes or labels

Information Gain: the reduction in entropy – a measure to maximize

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

Example

Supposed we have the following training data:

Completed Chores	Completed Homework	Played Video Games
Yes	Yes	Yes
Partially	Yes	Yes
Yes	No	Yes
No	Yes	Yes
No	No	No
Partially	No	No
Yes	No	Yes
No	Yes	Yes

Contingency Table

Completed Chores	Video Games	No Video Games	Total
Yes	3	0	3
No	2	1	3
Partially	1	1	2
Total	6	2	8

Entropy (Played Video Games) = $-\frac{6}{8}\log_2\frac{6}{8} - \frac{2}{8}\log_2\frac{2}{8} = .81$

Entropy (Played Video Games | Completed Chores = Yes) = $-\frac{3}{3}\log_2\frac{3}{3} - \frac{0}{3}\log_2\frac{0}{3} = 0$

Entropy (Played Video Games | Completed Chores = No) = $-\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} = .92$

Entropy (Played Video Games | Completed Chores = Partially) = $-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$

Weighted Average

Entropy (Played Video Games | Completed Chores) = $\frac{3}{8} \times 0 + \frac{3}{8} \times .92 + \frac{2}{8} \times 1 = .595$

Information Gain

$$= \text{E(Video Games)} - \text{E(Video Games | Chores)}$$

$$= .81 - .595 = .215$$

Contingency Table

Completed Homework	Video Games	No Video Games	Total
Yes	3	1	4
No	1	3	4
Total	4	4	8

Entropy (Played Video Games | Completed Homework = Yes) = $-\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4} = .81$

Entropy (Played Video Games | Completed Homework = No) = $-\frac{1}{4}\log_2\frac{1}{4} - \frac{3}{4}\log_2\frac{3}{4} = .81$

Weighted Average

Entropy (Played Video Games | Completed Homework) = $\frac{4}{8} \times .81 + \frac{4}{8} \times .81 = .81$

Information Gain

$$= \text{E(Video Games)} - \text{E(Video Games | Homework)}$$

$$= .81 - .81 = 0$$

<u>Results</u>

Since the information gain for completing chores is larger than the information gain for completing homework ($.215 > 0$), the feature chosen for the root node will be completing chores.