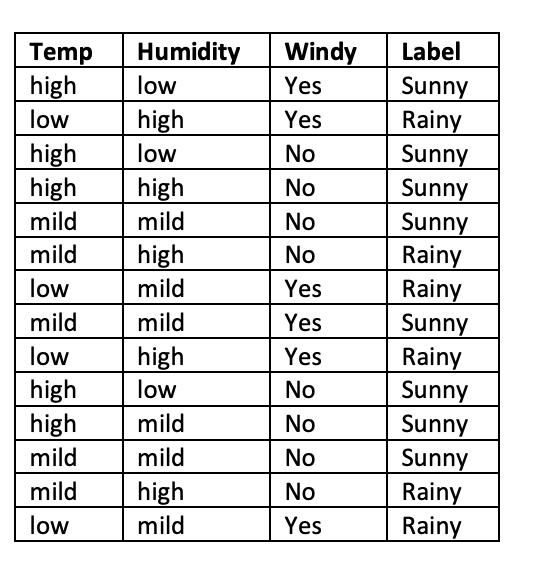
Adan Constanzo

CS 4661

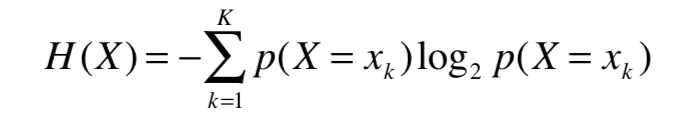
Homework 02

**Question 1**: Decision Tree for Weather Forecasting: Suppose that we want to build a decision tree classifier to perform weather forecasting! We have 3 features (Temp, Humidity, Wind), and a binary Label (sunny/rainy). The following table includes the data collected over the past two weeks. Based on this data, which feature is the best feature to put on the top of the tree? Justify your answer by providing detailed Entropy calculations.



Number of days is 14.

8 of those days were Sunny and the other 6 were Rainy.

Our “expected information” is 

Therefore,

Our Entropy before splitting is

First we split by wind,

* Windy: 6 samples of widny: 2 sunny, 4 rainy
* Not Windy: 8 samples of windy: 6 sunny, 2 rainy

Windy:

Not Windy:

Weighted Average

Entropy Before Split: 0.99

Entropy After Split: 0.86

Information Gain: 0.99 - 0.86 = 0.13

Second with Humidity,

* Low: 3 samples of low: 0 rainy, 3 sunny
* High: 5 samples of high: 4 rainy , 1 sunny
* Mild: 6 samples of mild: 2 rainy, 4 sunny

Low:

High:

Mild:

Weighted Average

Entropy Before Split: 0.99

Entropy After Split: 0.65

Information Gain: 0.99 - 0.65 = 0.34

Lastly with Temp,

* Low: 4 samples of low: 4 rainy, 0 sunny
* High: 5 samples of high: 0 rainy, 5 sunny
* Mild: 5 samples of mild: 2 rainy, 3 sunny

Low:

High:

Mild:

Weighted Average

Entropy Before Split: 0.99

Entropy After Split: 0.35

Information Gain: 0.99 - 0.35 = 0.54

By going through all the Entropy levels, the best feature to split on top of the tree is Temp because it provides the maximum information gain at this level.