# Class scribe-1 lecture 15

Observation of mutual information and information information gain on the tennis example demonstration in the class

CS 7840: Foundations and Applications of Information Theory (Fall 2024)

Instructors: Wolfgang Gatterbauer, Javed Aslam

Pranjal Umesh Kalekar

Master of science Northeastern University

# Understanding of the mutual information

**Mutual information/Information Gain** = entropy(parent) – [average entropy(children)]

- Based on the impurity we decide splits
- To decide on splits, we use various impurity measures (such as entropy and Gini index) to calculate how "pure" the subsets are after a split.
- The concept of information gain is introduced as a measure to decide the optimal feature for splitting. Information gain compares the entropy before and after a split, with a higher information gain suggesting a more informative split.

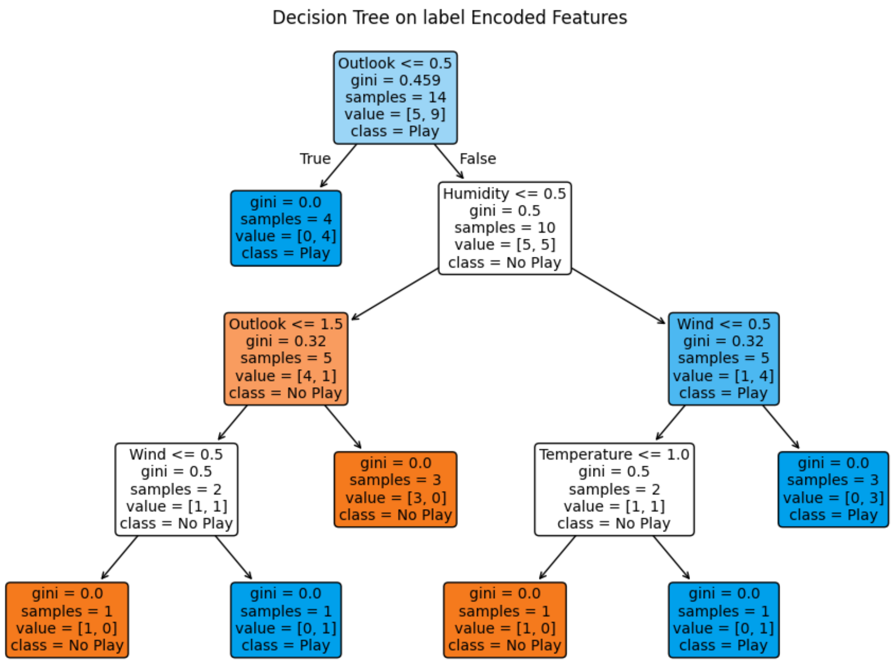
$$\text{Gain ratio} = \frac{\Delta_{\text{info}}}{\text{split info}}$$

- Gain ratio is discussed as an alternative to information gain, accounting for the "split information" or how uniformly an attribute splits the data, which avoids over-favoring attributes with many unique values.

# Implementation on the Tennis classification example

| day | Predictors | | | | Response |
|---|---|---|---|---|---|
| | **(O)utlook** | **(T)emp.** | **(H)umidity** | **(W)ind** | **(P)lay** |
| 1 | sunny | hot | high | weak | no |
| 2 | sunny | hot | high | strong | no |
| 3 | overcast | hot | high | weak | yes |
| 4 | rain | mild | high | weak | yes |
| 5 | rain | cool | normal | weak | yes |
| 6 | rain | cool | normal | strong | no |
| 7 | overcast | cool | normal | strong | yes |
| 8 | sunny | mild | high | weak | no |
| 9 | sunny | cool | normal | weak | yes |
| 10 | rain | mild | normal | weak | yes |
| 11 | sunny | mild | normal | strong | yes |
| 12 | overcast | mild | high | strong | yes |
| 13 | overcast | hot | normal | weak | yes |
| 14 | rain | mild | high | strong | no |

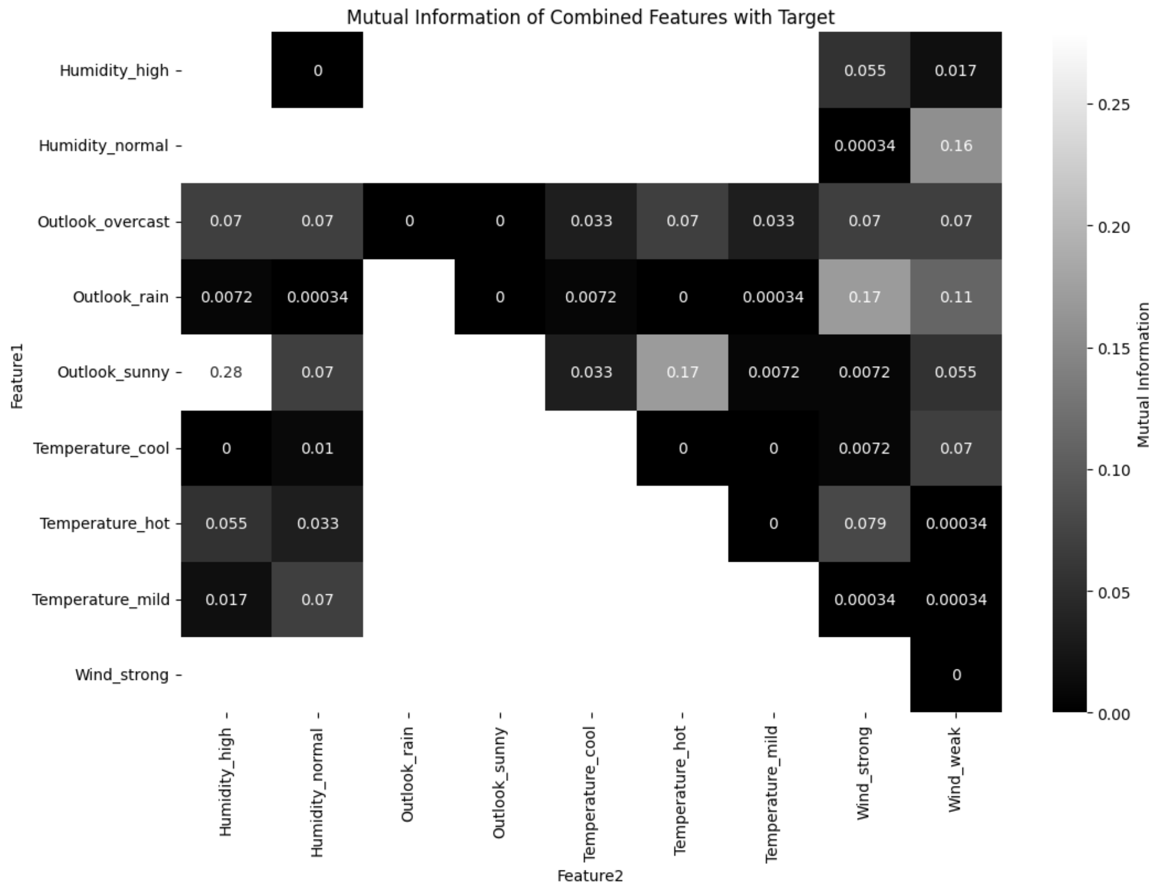
| day | Predictors | | | | Response |
|---|---|---|---|---|---|
| | **(O)utlook** | **(T)emp.** | **(H)umidity** | **(W)ind** | **(P)lay** |
| 1 | sunny | hot | high | weak | no |
| 2 | sunny | hot | high | strong | no |
| 8 | sunny | mild | high | weak | no |
| 9 | sunny | cool | normal | weak | yes |
| 11 | sunny | mild | normal | strong | yes |
| 3 | overcast | hot | high | weak | yes |
| 7 | overcast | cool | normal | strong | yes |
| 12 | overcast | mild | high | strong | yes |
| 13 | overcast | hot | normal | weak | yes |
| 6 | rain | cool | normal | strong | no |
| 14 | rain | mild | high | strong | no |
| 4 | rain | mild | high | weak | yes |
| 5 | rain | cool | normal | weak | yes |
| 10 | rain | mild | normal | weak | yes |




$$I(P;O) = 0.246$$
$$I(P;H) = 0.152$$
$$I(P;W) = 0.048$$
$$I(P;T) = 0.029$$

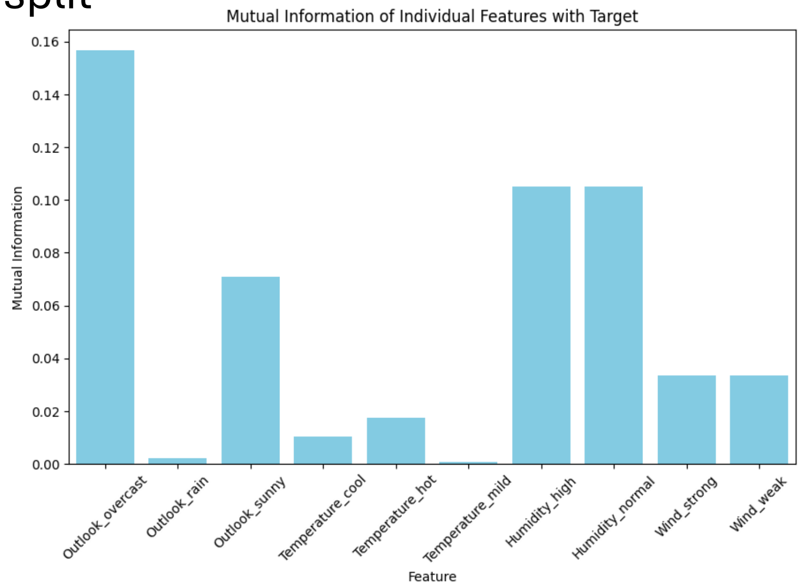
- Based on the mutual information we got following order of feature preferences for top down classifictaion approach
- I tried visualizing the same using scikit learns top down visualization of the decision tree with label encoded data
- And we can see the same order in the tree
- Ultimately ‘Day’ was the most informative feature

# Combining Multiple features and observe information gain spread through out




Calculate mutual information for each pairwise combination of features to see if combining features increases the information gain.

Humidity and Oulook in combination show the highest information gain – again following on to the top down decision tree split




mutual information values for each feature individually.

# Combining features further:

cumulative mutual information as more feature combinations are added, providing a sense of how feature combinations improve information gain.