CS 515: Assignment 1

Cam J. Loader

October 16, 2024

- Calculating Information Gain for each:

  Entropy on whole tree:

  $$-9/14log_2 9/14 - 5/14log_2 5/14 = 0.940$$

  First split calculations:

  Split on Outlook:

  $$.940 - (5/14) * (-2/5log2/5 - 3/5log3/5)+$$
  $$(4/14) * (-4/4log4/4 - 0/4log0/4)+$$
  $$(5/14) * (-2/5log2/5 - 3/5log3/5) = 0.246$$

  Split on Temperature:

  $$.940 - (5/14) * (-2/5log2/5 - 3/5log3/5)+$$
  $$(4/14) * (-2/4log2/4 - 2/4log2/4)+$$
  $$(5/14) * (-2/5log2/5 - 3/5log3/5) = 0.029$$

  Split on Humidity:

  $$.940 - (8/14) * (-4/8log4/8 - 4/8log4/8))+$$
  $$(6/14) * (-4/8log4/8 - 4/8log4/8) = 0.152$$

  Split on Wind:

  $$.940 - (8/14) * (-5/8log5/8 - 3/8log3/8))+$$
  $$(6/14) * (-3/8log3/8 - 3/8log3/8) = 0.048$$

  Based on Information Gain, we can do a split on Outlook. The second split will be calculated next.

Split on Humidity:

$$0.246 - (-4/9 * (3/9log3/9 - 6/9log6/9)+$$
$$(-5/9) * (2/5log2/5 - 3/5log3/5) = 0.2432$$

Split on Wind:

$$0.246 - (-3/8 * (2/8log2/8 - 6/8log6/8)+$$
$$(-5/8) * (3/5log3/5 - 2/5log2/5) = 0.229$$

Split on Temperature:

$$0.246 - (-2/6) * (2/2log2/2 - 0/2log0/2)+$$
$$(-2/6) * (2/2log2/2 - 0/2log0/2)+$$
$$(-2/6) * (2/2log2/2 - 0/2log0/2) = 0.216$$

From the Information above, we split on Humidity.

The tree starts at Outlook, branches 3 ways, sunny, outcast, rainy, where outcast is a leaf node [0,4]. We then break it down on Hunidity.

- Cosine Similarity
  i.

$$< 1, 1, 1, 1, 1, 1, 0, 0, 0, 0 >$$
$$< 1, 1, 0, 0, 0, 1, 1, 1, 1, 1 >$$
$$\frac{1 * 1 + 1 * 1 + 1 * 0 + 1 * 0 + 1 * 0 + 1 * 1 + 0 * 1 + 0 * 1 + 0 * 1 + 0 * 1}{(1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 0^2 + 0^2 + 0^2 + 0^2)(1^2 + 1^2 + 0^2 + 0^2 + 0^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2)}$$
$$CosineSim = 0.46291$$

  ii.

$$< 1, 1, 1, 0, 0, 0 >$$
$$< 0, 0, 0, 1, 1, 1 >$$
$$\frac{1 * 0 + 1 * 0 + 1 * 0 + 0 * 1 + 0 * 1 + 0 * 1}{(1^2 + 1^2 + 1^2 + 0^2 + 0^2 + 0^2)(0^2 + 0^2 + 0^2 + 1^2 + 1^0 + 1^2)}$$
$$CosineSim = 0$$

- $$P = (1, 1, 1, 1, 0, 1)Q = (1, 0, 0, 1, 1, 0)$$

  Simple Matching Coefficient

$$(0 + 2)/(0 + 2 + 3 + 1) = 0.167$$

  Jaccard Coefficient

$$2/(1 + 2 + 2) = 0.40$$

- Refer to ipynb