Maxwell Jung
9319674

Homework 4

CMPSC 165B
Nov 23, 2022

1. (a) Decision Tree

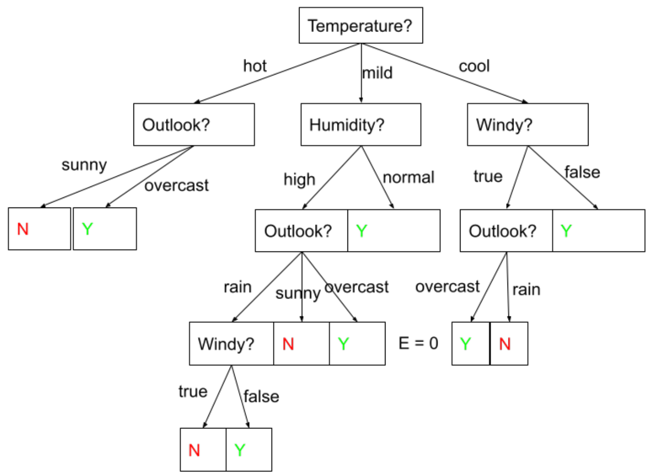

Figure 1: Decision Tree

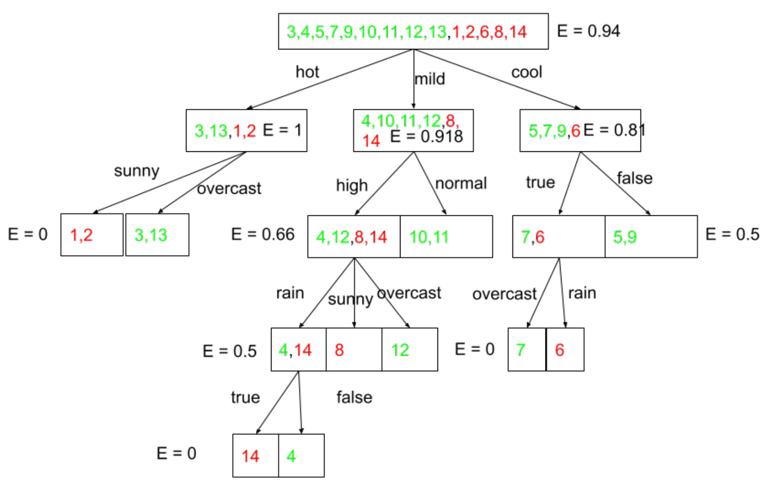

Figure 2: Decision Tree on Training Data

(b) Decision Tree

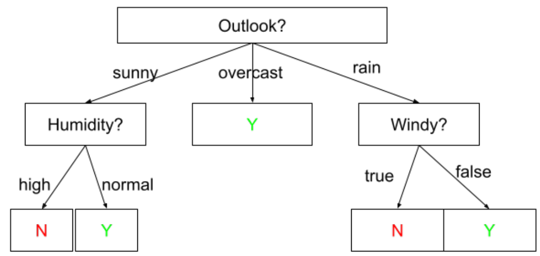

Figure 3: Decision Tree

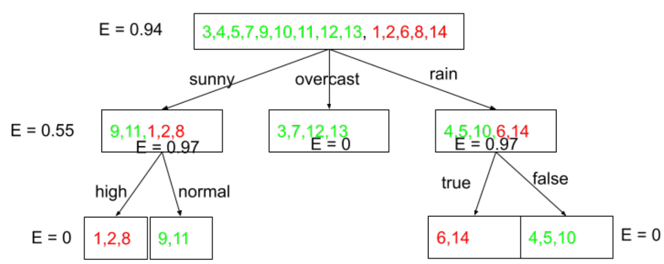

Figure 4: Decision Tree on Training Data

Calculating entropy and information gain:

Only one example of calculations are shown because the rest of the calculations are repetitions of this example.

Entropy is $E = E(p_y, p_n) = p_y \log_2(p_y) + p_n \log_2(p_n)$ where p_y is the proportion of 'yes' (greens) and p_n is the proportion of 'no' (reds) in each node (represented as boxes).

In Figure 4, the top node's entropy is $E_p = E(\frac{9}{14}, \frac{5}{14}) = \frac{9}{14}\log_2(\frac{9}{14}) + \frac{5}{14}\log_2(\frac{5}{14}) \approx 0.94$.

Its child's E from left to right are $E(\frac{2}{5}, \frac{3}{5}) = 0.97$, $E(\frac{4}{4}, \frac{0}{4}) = 0$, and $E(\frac{3}{5}, \frac{2}{5}) = 0.97$

Information gain is the change in entropy from parent to children. The entropy of children is the weighted average entropy of each child under the parent.

For example, the entropy of the three children in the second layer of Figure 4 is $E_c = \frac{5}{14}(0.97) + \frac{4}{14}(0) + \frac{5}{14}(0.97) \approx 0.55$, and the information gain from top node to next node is $E_p - E_c \approx 0.94 - 0.55 \approx 0.39$

The best feature is determined by running the above calculations for each feature and selecting the feature with the largest information gain. For figure 4, out of the 4 features $\{\text{Outlook}, \text{Temperature}, \text{Humidity}, \text{Windy}\}$, Outlook had the highest information gain from first node to second node and was picked as the first feature.

Homework 4

2. (a) The assumption in Naive Bayes classifier is that all the features are conditionally independent of each other with respect to the class label.

   In other words: $p(\boldsymbol{X}|Y) = \prod_j p_j(x^j|Y)$

   (b) The number of parameters to be estimated without simplifying assumption is $2(k^d - 1)$ because there are k^d unique feature vectors and 2 possible labels. The subtraction comes from the fact that the conditional probability of the last vector is equal to the complement of the probabilities of all the previous vectors.

   With the conditional independence assumption, the number of estimations is $2d(k-1)$ because for each label, there are d different attributes, and for each attribute, there are k possible probabilities. The kth probability can be derived from the previous $k-1$ probabilities.

3. (a) $Y = \{\text{Physics}, \text{Biology}, \text{Chemistry}\}$

   $X = \{\text{atom}, \text{carbon}, \text{proton}, \text{life}, \text{earth}\}$

   $A =$ "the carbon atom is the foundation of life on earth"

   $\rightarrow x_A = \{\text{atom}, \text{carbon}, \neg\text{proton}, \text{life}, \text{earth}\}$

   $p(\text{Physics}|x_A) = p(\text{Physics}) \prod_{x_{Ai}}^{\{\text{atom,carbon,}\neg\text{proton,life,earth}\}} p(x_{Ai}|\text{Physics})$

   $= 0.35(0.1 \times 0.005 \times 0.95 \times 0.001 \times 0.005) = 8.3125 \times 10^{-10}$

   $p(\text{Biology}|x_A) = p(\text{Biology}) \prod_{x_{Ai}}^{\{\text{atom,carbon,}\neg\text{proton,life,earth}\}} p(x_{Ai}|\text{Biology})$

   $= 0.40(0.01 \times 0.03 \times 0.97 \times 0.1 \times 0.006) = 6.984 \times 10^{-8}$

   $p(\text{Chemistry}|x_A) = p(\text{Chemistry}) \prod_{x_{Ai}}^{\{\text{atom,carbon,}\neg\text{proton,life,earth}\}} p(x_{Ai}|\text{Chemistry})$

   $= 0.25(0.2 \times 0.05 \times 0.95 \times 0.008 \times 0.003) = 5.7 \times 10^{-8}$

   Document A is most likely Biology related.

   (b) $Y = \{\text{Physics}, \text{Biology}, \text{Chemistry}\}$

   $X = \{\text{atom}, \text{carbon}, \text{proton}, \text{life}, \text{earth}\}$

   $B =$ "the carbon atom contains 12 protons"

   $\rightarrow x_B = \{\text{atom}, \text{carbon}, \text{proton}, \neg\text{life}, \neg\text{earth}\}$

   $p(\text{Physics}|x_B) = p(\text{Physics}) \prod_{x_{Bi}}^{\{\text{atom,carbon,proton,}\neg\text{life,}\neg\text{earth}\}} p(x_{Bi}|\text{Physics})$

   $= 0.35(0.1 \times 0.005 \times 0.05 \times 0.999 \times 0.995) = 8.69754375 \times 10^{-6}$

   $p(\text{Biology}|x_B) = p(\text{Biology}) \prod_{x_{Bi}}^{\{\text{atom,carbon,proton,}\neg\text{life,}\neg\text{earth}\}} p(x_{Bi}|\text{Biology})$

   $= 0.40(0.01 \times 0.03 \times 0.001 \times 0.9 \times 0.994) = 1.07352 \times 10^{-7}$

   $p(\text{Chemistry}|x_B) = p(\text{Chemistry}) \prod_{x_{Bi}}^{\{\text{atom,carbon,proton,}\neg\text{life,}\neg\text{earth}\}} p(x_{Bi}|\text{Chemistry})$

   $= 0.25(0.25 \times 0.2 \times 0.05 \times 0.992 \times 0.997) = 6.1814 \times 10^{-4}$

   Document B is most likely Chemistry related.