Tutorial 9

5th November 2021

Question 1

In this question, we will use information gain (entropy reduction) as the attribute selection method and build a decision tree.

Question 1 (a)

Before we start considering the features, calculate the entropy of the target variable “Buys Computer”.

Age	Income	Student	Credit Rating	Buys Computer
Young	High	No	Fair	No
Young	High	No	Excellent	No
Middle-aged	High	No	Fair	Yes
Senior	Medium	No	Fair	Yes
Senior	Low	Yes	Fair	Yes
Senior	Low	Yes	Excellent	No
Middle-aged	Low	Yes	Excellent	Yes
Young	Medium	No	Fair	No
Young	Low	Yes	Fair	Yes
Senior	Medium	Yes	Fair	Yes
Young	Medium	Yes	Excellent	Yes
Middle-aged	Medium	No	Excellent	Yes
Middle-aged	High	Yes	Fair	Yes
Senior	Medium	No	Excellent	No

Question 1 (b)

What will be the average entropy if we split on the variable “Age”?

Age	Income	Student	Credit Rating	Buys Computer
Young	High	No	Fair	No
Young	High	No	Excellent	No
Middle-aged	High	No	Fair	Yes
Senior	Medium	No	Fair	Yes
Senior	Low	Yes	Fair	Yes
Senior	Low	Yes	Excellent	No
Middle-aged	Low	Yes	Excellent	Yes
Young	Medium	No	Fair	No
Young	Low	Yes	Fair	Yes
Senior	Medium	Yes	Fair	Yes
Young	Medium	Yes	Excellent	Yes
Middle-aged	Medium	No	Excellent	Yes
Middle-aged	High	Yes	Fair	Yes
Senior	Medium	No	Excellent	No

Question 1 (c)

What is the information gain (or reduction of entropy) if we split on “Age”?

Age	Income	Student	Credit Rating	Buys Computer
Young	High	No	Fair	No
Young	High	No	Excellent	No
Middle-aged	High	No	Fair	Yes
Senior	Medium	No	Fair	Yes
Senior	Low	Yes	Fair	Yes
Senior	Low	Yes	Excellent	No
Middle-aged	Low	Yes	Excellent	Yes
Young	Medium	No	Fair	No
Young	Low	Yes	Fair	Yes
Senior	Medium	Yes	Fair	Yes
Young	Medium	Yes	Excellent	Yes
Middle-aged	Medium	No	Excellent	Yes
Middle-aged	High	Yes	Fair	Yes
Senior	Medium	No	Excellent	No

Question 1 (d)

Calculate $Gain_{Income}$, $Gain_{Student}$ and $Gain_{Credit\ Rating}$
Which attribute should be selected and why?

Age	Income	Student	Credit Rating	Buys Computer
Young	High	No	Fair	No
Young	High	No	Excellent	No
Middle-aged	High	No	Fair	Yes
Senior	Medium	No	Fair	Yes
Senior	Low	Yes	Fair	Yes
Senior	Low	Yes	Excellent	No
Middle-aged	Low	Yes	Excellent	Yes
Young	Medium	No	Fair	No
Young	Low	Yes	Fair	Yes
Senior	Medium	Yes	Fair	Yes
Young	Medium	Yes	Excellent	Yes
Middle-aged	Medium	No	Excellent	Yes
Middle-aged	High	Yes	Fair	Yes
Senior	Medium	No	Excellent	No

Question 1 (e)

The root node branches into 3 (Young, Middle-aged, Old). How many data points will you use to to grow the tree further along each of these branches?

Age	Income	Student	Credit Rating	Buys Computer
Young	High	No	Fair	No
Young	High	No	Excellent	No
Middle-aged	High	No	Fair	Yes
Senior	Medium	No	Fair	Yes
Senior	Low	Yes	Fair	Yes
Senior	Low	Yes	Excellent	No
Middle-aged	Low	Yes	Excellent	Yes
Young	Medium	No	Fair	No
Young	Low	Yes	Fair	Yes
Senior	Medium	Yes	Fair	Yes
Young	Medium	Yes	Excellent	Yes
Middle-aged	Medium	No	Excellent	Yes
Middle-aged	High	Yes	Fair	Yes
Senior	Medium	No	Excellent	No

Question 1 (f) (H.W)

Build the tree to completion.

Age	Income	Student	Credit Rating	Buys Computer
Young	High	No	Fair	No
Young	High	No	Excellent	No
Middle-aged	High	No	Fair	Yes
Senior	Medium	No	Fair	Yes
Senior	Low	Yes	Fair	Yes
Senior	Low	Yes	Excellent	No
Middle-aged	Low	Yes	Excellent	Yes
Young	Medium	No	Fair	No
Young	Low	Yes	Fair	Yes
Senior	Medium	Yes	Fair	Yes
Young	Medium	Yes	Excellent	Yes
Middle-aged	Medium	No	Excellent	Yes
Middle-aged	High	Yes	Fair	Yes
Senior	Medium	No	Excellent	No

Question 2

Which of the following are parallelizable?

- Random Forests at train time
- AdaBoost at train time
- Random Forest at test time
- AdaBoost at test time

Question 3

Consider an alternative way of learning a Random Forest where instead of randomly sampling the attributes at each node, we sample a subset of attributes for each tree and build the tree on these features. Would you prefer this method over the original or not, and why?

1. Yes, because it reduces the correlation between the resultant trees
2. Yes, because it reduces the time taken to build the trees due to the decrease in the attributes considered
3. No, because many of the trees will be bad classifiers due to the absence of critical features considered in the construction of some of the trees

Question 4

Considering the AdaBoost algorithm, which among the following statements is true?

1. In each stage, we try to train a classifier which makes accurate predictions on any subset of the data points where the subset size is at least half the size of the data set
2. In each stage, we try to train a classifier which makes accurate predictions on a subset of the data points where the subset contains more of the data points which were misclassified in earlier stages
3. The weight assigned to an individual classifier depends upon the number of data points correctly classified by the classifier
4. The weight assigned to an individual classifier depends upon the weighted sum error of misclassified points for that classifier

Question 5

In AdaBoost, we re-weight points giving points misclassified in previous iterations more weight. Suppose we introduced a limit or cap on the weight that any point can take (for example, say we introduce a restriction that prevents any point's weight from exceeding a value of 10). Which among the following would be an effect of such a modification?

1. We may observe the performance of the classifier reduce as the number of stages increase
2. It makes the final classifier robust to outliers
3. It may result in lower overall performance