Started on	Thursday, 29 March 2018, 4:14 PM
State	Finished
Completed on	Thursday, 29 March 2018, 4:42 PM
Time taken	27 mins 56 secs
Grade	**11.0** out of 11.0 (**100**%)
Feedback	Well done!

Information

Athletics

age	club	wins
child	red	yes
child	blue	no
youth	red	yes
adult	red	no
adult	blue	yes

Consider the Athletics dataset, D, above which is a training set for a decision tree classifier aiming to predict the variable "wins" from age and club membership.

Let p_i be the probability that an arbitrary tuple in D belongs to class C_i, of m classes estimated by $p_i = \frac{|C_{i,D}|}{|D|}$.

Then the information needed to classify a tuple in D is defined by
$Info(D) = -\sum_{i=1}^{m} p_i log_2(p_i)$
and after using attribute A to split D into v partitions, the information needed is

$$Info_A(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} \times Info(D_j)$$

So the information gain by splitting is defined as

$$Gain(A) = Info(D) - Info_A(D)$$

Question 1

Correct

Mark 1.0 out of 1.0

What is the information gain for splitting D on the three categories of attribute "age"?

Give your answer to 2 decimal places.

Answer: 0.17

That is, I(p,n) = - (p/(p+n) * log_2 (p/(p+n)) + n/(p+n) * log_2 (n/(p+n)))

Info(D) = I(3, 2) = -(3/5 * log_2 (3/5) + 2/5 * log_2 (2/5)) = 0.971

Now checking for split on age,

child: I(1,1) = - (1/2* log_2(1/2) + 1/2*log_2 (1/2))= -1

youth: I(1,0) = -log_2(1) = 0

adult = I(1,1) =-1

Info_{age}(D) = - (2/5 * I(1,1) + 1/5 * I(1,0) + 2/5* I(1,1)) = 2/5 + 0 + 2/5 = 4/5 = 0.8

Gain_age = Info(D) - Info_{age}(D) = 0.971 - 0.800 = 0.171

p_{yes}= 3/5. p_{no}= 2/5.

Define I(p, n) to be Info(D) with p is the number of "yes" tuples in D and n is the number of "no" tuples in D.

That is, I(p,n) = - (p/(p+n) * log_2 (p/(p+n)) + n/(p+n) * log_2 (n/(p+n)))

Info(D) = I(3, 2) = -(3/5 * log_2 (3/5) + 2/5 * log_2 (2/5)) = 0.971

Now checking for split on age,

child: I(1,1) = - (1/2* log_2(1/2) + 1/2*log_2 (1/2))= -1

youth: I(1,0) = -log_2(1) = 0

adult = I(1,1) =-1

Info_{age}(D) = - (2/5 * I(1,1) + 1/5 * I(1,0) + 2/5* I(1,1)) = 2/5 + 0 + 2/5 = 4/5 = 0.8

Gain_age = Info(D) - Info_{age}(D) = 0.971 - 0.800 = 0.171

The correct answer is: 0.171

Correct

Marks for this submission: 1.0/1.0.

Question 2

Correct

Mark 1.0 out of 1.0

What is the information gain from splitting on the two categories for attribute "club"?

Give your answer to 2 decimal places.

Answer: 0.02

I(p,n) = - (p/(p+n) * log_2 (p/(p+n)) + n/(p+n) * log_2 (n/(p+n)))

club:

red: I(2,1) = - (2/3 * log_2(2/3) + 1/3 * log_2(1/3)) = (2/3 *0.585) + 1/3 *1.585 = 0.390 + 0.528 = 0.918

blue: I(1,1) = -(1/2 log_2(1/2) + 1/2 log_2(1/2)) = 1

Info_{club}(D) = 3/5* I(2,1) + 2/5 * I(1,1) = 3/5 * 0.918 + 2/5 * 1 = 0.551 + 0.4 = 0.951

Gain_{club} = Info(D) - Info_{club}(D) = 0.971 - 0.951 = 0.02

I(p,n) = - (p/(p+n) * log_2 (p/(p+n)) + n/(p+n) * log_2 (n/(p+n)))

club:

red: I(2,1) = - (2/3 * log_2(2/3) + 1/3 * log_2(1/3)) = (2/3 *0.585) + 1/3 *1.585 = 0.390 + 0.528 = 0.918

blue: I(1,1) = -(1/2 log_2(1/2) + 1/2 log_2(1/2)) = 1

Info_{club}(D) = 3/5* I(2,1) + 2/5 * I(1,1) = 3/5 * 0.918 + 2/5 * 1 = 0.551 + 0.4 = 0.951

Gain_{club} = Info(D) - Info_{club}(D) = 0.971 - 0.951 = 0.02

The correct answer is: 0.02

Correct

Marks for this submission: 1.0/1.0.

Question 3

Correct

Mark 1.0 out of 1.0

So the best split to do is on age, and not club, right?

Select one:

- (•) True ✓
- ( ) False

Choose the attribute with maximum gain to split on.

The correct answer is 'True'.

Correct

Marks for this submission: 1.0/1.0.

Question 4

Correct

Mark 1.0 out of 1.0

Decision trees are classically suited to nominal ✓ attributes whereby at any node in the tree, an attribute can be split several ways, one for each value of the attribute.

For numeric (a k.a continuous) ✓ attributes, a split-point is calculated at run time that partitions the data at the node into those with lower values and those with higher values than the split point. The split is selected at run-time to maximise the attribute selection heuristic, computed by evaluating the heuristic for splitting half-way between every pair of adjacent (sorted) attribute values.

ordinal ✓ attributes can be treated either of those two ways.

In some tree methods only binary ✓ splits are allowed, in which case the latter method can also be used for nominal ✓ data by splitting the possible attribute values into two sets, and using set membership rather than numeric comparison to evaluate the split.

When the target variable for a decision tree is a numeric (a k.a continuous) ✓ variable rather than a nominal class, the tree is called a regression tree, and the predicted value for the leaf node may be the average value for the objects in the node.

Your answer is correct.

The correct answer is:
Decision trees are classically suited to [nominal] attributes whereby at any node in the tree, an attribute can be split several ways, one for each value of the attribute.

For [numeric (a k.a continuous)] attributes, a split-point is calculated at run time that partitions the data at the node into those with lower values and those with higher values than the split point. The split is selected at run-time to maximise the attribute selection heuristic, computed by evaluating the heuristic for splitting half-way between every pair of adjacent (sorted) attribute values.

[ordinal] attributes can be treated either of those two ways.

In some tree methods only [binary] splits are allowed, in which case the latter method can also be used for [nominal] data by splitting the possible attribute values into two sets, and using set membership rather than numeric comparison to evaluate the split.

When the target variable for a decision tree is a [numeric (a k.a continuous)] variable rather than a nominal class, the tree is called a regression tree, and the predicted value for the leaf node may be the average value for the objects in the node.

Correct
Marks for this submission: 1.0/1.0.

Question 5
Correct
Mark 1.0 out of 1.0

Overfitting in a decision tree model

can be ameliorated by	stopping construction early ✓
causes	poor accuracy on unseen data ✓
when fresh (unseen but labelled) data is available, can be fixed by	postpruning branches ✓
occurs when a tree has	very high accuracy on the training data ✓

Your answer is correct.

The correct answer is: can be ameliorated by → stopping construction early, causes → poor accuracy on unseen data, when fresh (unseen but labelled) data is available, can be fixed by → postpruning branches, occurs when a tree has → very high accuracy on the training data

Correct
Marks for this submission: 1.0/1.0.

Information

Working from the following dataset

age	credit	buys_computer
youth	fair	no
youth	fair	yes
middle_aged	excellent	yes
middle_aged	fair	no
youth	excellent	no
middle_aged	excellent	no
middle_aged	fair	yes

And using Naive Bayes to predict the class that maximises

$P(C_i|X) \propto P(X|C_i)P(C_i)$

where

$$P(\boldsymbol{X}|C_i) = \prod_{k=1}^{n} P(x_k|C_i)$$

$$= P(x_1|C_i) \times P(x_2|C_i) \times \cdots \times P(x_n|C_i).$$

Question 6

Correct

Mark 1.0 out of 1.0

Calculate the classification of a customer X = (middle-aged, fair) as a computer-buyer or not in the following steps.

1. Calculate the prior probability of an arbitrary customer buying a computer (i.e belonging to the class buys_computer = yes) regardless of other information about the customer. 3/7 ✓

2. Calculate the prior probability of an arbitrary customer *not* buying a computer (i.e belonging to the class buys_computer = no) regardless of other information about the customer. 4/7 ✓

3. Calculate the likelihood probability P(age=middle_aged)| buys=yes) 2/3 ✓

4. Calculate the likelihood probability P(credit=fair| buys=yes) 2/3 ✓

5. Calculate the likelihood P(X = (middle-aged, fair) | buys = yes) 4/9 ✓

6. Using $P(C_i|X) \propto P(X|C_i)P(C_i)$

Calculate P(X = (middle-aged, fair) | buys = yes) * P(buys = yes) 12/63 ✓

Now, repeat for buys = no

3b. Calculate the likelihood probability P(age=middle_aged| buys=no) 1/2 ✓

4b. Calculate the likelihood probability P(credit=fair| buys=no) 1/2 ✓

5b. Calculate the likelihood P(X = (middle-aged, fair) | buys = no) 1/4 ✓

6b. Using $P(C_i|X) \propto P(X|C_i)P(C_i)$

Calculate P(X = (middle-aged, fair) | buys = no) * P(buys = no) 1/7 ✓

Your answer is correct.

The correct answer is: 1. Calculate the prior probability of an arbitrary customer buying a computer (i.e belonging to the class buys_computer = yes) regardless of other information about the customer. → 3/7, 2. Calculate the prior probability of an arbitrary customer *not* buying a computer (i.e belonging to the class buys_computer = no) regardless of other information about the customer. → 4/7, 3. Calculate the likelihood probability P(age=middle_aged)| buys=yes) → 2/3, 4. Calculate the likelihood probability P(credit=fair| buys=yes) → 2/3, 5. Calculate the likelihood P(X =

(middle-aged, fair) | buys = yes) → 4/9, 6. Using $P(C_i|X) \propto P(X|C_i)P(C_i)$ Calculate P(X = (middle-aged, fair) | buys = yes) * P(buys = yes) → 12/63, **Now, repeat for buys = no**

3b. Calculate the likelihood probability P(age=middle_aged| buys=no) → 1/2, 4b. Calculate the likelihood probability P(credit=fair| buys=no) → 1/2, 5b. Calculate the likelihood P(X = (middle-aged, fair) | buys = no) → 1/4, 6b. Using $P(C_i|X) \propto P(X|C_i)P(C_i)$ Calculate P(X = (middle-aged, fair) | buys = no) * P(buys = no) → 1/7

Correct

Marks for this submission: 1.0/1.0.

Question 7

Correct

Mark 1.0 out of 1.0

So X = (middle_aged, fair) is classified as someone who buys a computer, right?

Select one:

- (•) True ✓
- ( ) False

Relies on 12/63 > 1/7

The correct answer is 'True'.

Correct

Marks for this submission: 1.0/1.0.

Question 8

Correct

Mark 1.0 out of 1.0

NEGATIVE MARKS ARE AWARDED FOR WRONG ANSWERS

In the confusion matrix below,

Actual Class\Predicted class	cancer = yes	cancer = no	Total	Recognition(%)
cancer = yes	**90**	**210**	300	30.00 (*sensitivity*
cancer = no	**140**	**9560**	9700	98.56 (*specificity*)
Total	230	9770	10000	96.40 (*accuracy*)

Select one or more:

- [ ] a. True Positives = 9560 and True Negatives = 90
- [x] b. Accuracy = Correct predictions / Size of the data ✓
- [x] c. True Positives = 90 and True Negatives = 9560 ✓
- [ ] d. Sensitivity is the recognition rate of negative examples
- [x] e. Error Rate = (False Positives + False Negatives)/ Total Size of Data ✓
- [x] f. Error Rate = 3.6% ✓
- [ ] g. Specificity is the recognition rate of positive examples
- [ ] h. False Positives = 210
- [x] i. False Positives = 140 ✓
- [ ] j. Recall is another name for specificity

Your answer is correct.

The correct answers are: True Positives = 90 and True Negatives = 9560, False Positives = 140, Accuracy = Correct predictions / Size of the data, Error Rate = 3.6%, Error Rate = (False Positives + False Negatives)/ Total Size of Data

Correct

Marks for this submission: 1.0/1.0.

Information

Say you have a labelled dataset D of 50 labelled objects and you want to evaluate the performance of your fantastic new learning algorithm you wrote yourself, called Z. So you decide to evaluate the performance of Z using 5-fold cross-validation.

Question 9

Correct

Mark 1.0 out of 1.0

Is this a good idea? Answer "yes" or "no" and justify your answer.

Answer: no. 10 is better.

This would be a very small dataset and you need to get all you can out of the data to train your models. Maybe use leave-one-out, or better, bootstrap.

The correct answer is: no

Correct

Marks for this submission: 1.0/1.0.

Question 10

Correct

Mark 1.0 out of 1.0

So how do you build your training sets?

Select one:

- a. You partition into 5 mutually exclusive sets and you use each of those to train.
- b. You need 5 disjoint subsets and you make 5 training sets by each combination of 4 of them.
- c. You partition the 50 tuples into training, test and validation sets.
- d. You make 50 training sets by leaving one out one different tuple from each set.
- e. You partition the 50 tuples into a training set of about 85% and keep the rest back for a test set.

Your answer is correct.

The correct answer is: You need 5 disjoint subsets and you make 5 training sets by each combination of 4 of them.

Correct

Marks for this submission: 1.0/1.0.

Question 11

Correct

Mark 1.0 out of 1.0

So now you have used Z to train some classifiers with % accuracies 20, 20, 50, 60, 70. What do you report as the performance accuracy of Z on D? Give your answer to 1 decimal place.

Answer: 44.0

average

The correct answer is: 44

Correct

Marks for this submission: 1.0/1.0.