19T2: COMP9417 Machine Learning and Data Mining

Lectures: Tree Learning
Topic: Questions from lectures
Last revision: Fri Jun 28 09:57:48 AEST 2019

Introduction

Some questions and exercises from the course lectures covering aspects of supervised tree learning for classification and regression.

Expressiveness of decision trees

Question 1 Give decision trees to represent the following Boolean functions, where the variables `A`, `B`, `C` and `D` have values `t` or `f`, and the class value is either `True` or `False`:

a) $A \wedge \neg B$

b) $A \vee [B \wedge C]$

c) $A \ \ X\text{OR} \ \ B$

d) $[A \wedge B] \vee [C \wedge D]$

Can you observe any effect of the increasing complexity of the functions on the form of their expression as decision trees ?

Decision tree learning

Question 2 Here is small dataset for a two-class prediction task. There are 4 attributes, and the class is in the rightmost column. Look at the examples. Can you guess which attribute(s) will be most predictive of the class ?

species	**rebel**	**age**	**ability**	**homeworld**
pearl	yes	6000	regeneration	no
bismuth	yes	8000	regeneration	no
pearl	no	6000	weapon-summoning	no
garnet	yes	5000	regeneration	no
amethyst	no	6000	shapeshifting	no
amethyst	yes	5000	shapeshifting	no
garnet	yes	6000	weapon-summoning	no
diamond	no	6000	regeneration	yes
diamond	no	8000	regeneration	yes
amethyst	no	5000	shapeshifting	yes
pearl	no	8000	shapeshifting	yes
jasper	no	6000	weapon-summoning	yes

You probably guessed that attributes 3 and 4 were not very predictive of the class, which is true. However, you might be surprised to learn that attribute "species" has higher information gain than attribute "rebel". Why is this ? Refer to slides 71–73 on "Attributes with Many Values" in the lecture notes.

Suppose you are told the following: for attribute "species" the Information Gain is 0.52 and *Split Information* is 2.46, whereas for attribute "rebel" the Information Gain is 0.48 and *Split Information* is 0.98.

Which attribute would the decision-tree learning algorithm select as the split when using the *Gain Ratio* criterion instead of Information Gain ? Is Gain Ratio a better criterion than Information Gain in this case ?

Question 3 Assume we learn a decision tree to predict class Y given attributes A, B and C from the following training set, with no pruning.

A	B	C	Y
0	0	0	0
0	0	1	0
0	0	1	0
0	1	0	0
0	1	1	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	0	1
1	1	1	0
1	1	1	1

What would be the training set error for this dataset ? Express your answer as the number of examples out of twelve that would be misclassified.

Question 4 One nice feature of decision tree learners is that they can learn trees to do *multi-class* classification, i.e., where the problem is to learn to classify each instance into exactly one of $k > 2$ classes.

Suppose a decision tree is to be learned on an arbitrary set of data where each instance has a discrete class value in one of $k > 2$ classes. What is the maximum training set error, expressed as a fraction, that any dataset could have ?

Model trees

Question 5 In learning a model tree, which is a piecewise-linear approximation to a function, it may be necessary to balance predictions made by linear models at the leaves, which possibly are based on small samples of the training set, with the "coarser" linear models fitted at internal nodes. Apply the *smoothing heuristic* shown on slide 93 of the lecture notes on "Tree Learning" for the following situation.

Suppose you have a univariate target function which for the value $x = 2$ is well approximated by $y = 2x + 1$. A model tree has been learned with an internal mode containing the linear model $y = 1.5x + 1$. At this internal node there is a split with condition $x \leq 2$ below which is the leaf node containing the linear model $y = 5x + 1$. Say the number of examples from the training set used to learn the leaf node was 2, and let's say the user has decided to apply a smoothing constant of 5. Compute the true value of y for $x = 2$, the value that will be returned by the leaf node, and the value that will be pased up the tree by the internal node. Has the smoothing improved the prediction ? Comment on the role of n and k as they are used in this formula.

Extending Tree Learning

Question 6 Propose an algorithm extending decision trees with Naive Bayes classification. Explain what changes to the basic TDIDT learning algorithm: a) during training; and b) at classification time.

Question 7 How could you combine tree learning with *local regression* as described on slides 101–105 of the week 1 lecture "Regression" ?

Question 8 The space of possible decision trees can be enormous, so searching for the "best" tree to fit a given dataset is usually not doneIn fact, under certain conditions on what the "best" tree is the problem is known to be NP-complete., with the most common solution being the greedy search of the basic TDIDT algorithm. Suggest how you could extend the basic greedy approach to allow a form of backtracking search.