Introduction to Artificial Intelligence
Autumn Semester 2017
October 13, 2017

The University of Birmingham
School of Computer Science
Jeremy L Wyatt

Exercise 2 - Machine Learning

Q1 Suppose that you have to classify animals using the decision-tree learning algorithm. The data to learn from is in the table:

#	Animal	Teeth	Size	Furry	Class
1	Tiger	Sharp	Big	Yes	Scary (+)
2	Piranha	Sharp	Small	No	Scary (+)
3	Elephant	Not sharp	Big	No	Not scary (-)
4	Orangutang	Not sharp	Big	Yes	Not scary (-)
5	Shark	Sharp	Big	No	Scary (+)
6	Cat	Sharp	Small	Yes	Not scary (-)

Recall that the formula for entropy is

$$H = -\sum_{i=1}^{n} p_i log_2(p_i)$$

(i) Which feature to split on reduces the entropy of the classification by the most? To answer this question you must calculate the weighted average entropy of the child nodes resulting from the split.

(ii) Why will the entropy reduction algorithm used above not always lead to a tree that reduces entropy by the most? Your answer must be less than 70 words.

(Total for this Question is 3 marks)

Q2

(i) Explain the difference between regression and classification in machine learning in less than 50 words.

(ii) Explain the difference between supervised learning, unsupervised learning and reward based learning.

(Total for this question is 2 marks)

Q3 Now download both the **simple** cancer data set from Canvas and the decision tree learning application from aispace.org. Follow the tutorials so that you know how to train and test decision trees. Then select a random 20% of the data for testing. Now train with the information gain rule selected under splitting functions, and the sole stopping condition being the maximum tree depth. Start by setting this as 2.

(i) Run training and then record the performance on the test set. This is the percentage of test examples predicted correctly. If you see that there are examples without predictions, count them as incorrect predictions. Now reselect a different 20% of the data and repeat. Do this to obtain 10 trials in total. Now repeat this whole experiment with depths = 2,4,8 and 16. Plot a graph of the average error rate against the tree depth, including standard error bars.

(ii) Does the tree depth make a difference? Explain the result.

(iii) Run learning one more time with maximum tree depth 2 again. You can see that the leaf nodes are not consistent in most cases. Using the information available at the node devise a rule that can still produce a classification for a node even when it is not consistent.

(2.5 marks for this question)

Q4 (optional)

(i) There are two kinds of errors the classifier can make. A false positive is when you classify a 0 (not cancer) as 1 (cancer). A false negative is when you classify a 1 (cancer) as 0 (not cancer). Suppose that a false negative is 100 times more costly than a false positive. Devise a new classification rule for inconsistent nodes that takes this fact into account and which is guaranteed to minimise the expected cost of misclassifications.

(1 bonus mark for this question up to a total of 7.5)