Final Exam Machine Learning
2009

14 December 2009, 08:45–11:30

This exam is **open book**: you can use Tom Mitchell's "Machine Learning" as well as the lecture slides and any notes you've taken. You can use a calculator.
Answers are allowed in Dutch and English.

Good luck!

Questions

1. **Short answers** *(18 Points)*

   (a) (**True** or **False**): If $P(A|B) = P(A)$, then $P(A \wedge B) = P(A)P(B)$.

   (b) (**True** or **False**): Because decision trees learn to classify using only discrete-valued variables, they cannot overfit.

   (c) (**True** or **False**): When using a kernel function, instance-based learning can use all data points instead of only the k nearest ones.

   (d) (**True** or **False**): Because Bayes Classifiers work independently of an hypothesis space, the training-set performance is an unbiased estimator of true performance.

   (e) (**True** or **False**): Genetic Algorithms are guaranteed to converge to a local optimum.

   (f) (**True** or **False**): Overtting is more likely when the set of training data is small

2. **Decision Trees** *(23 Points)*

   Gordon Ramsey is coming to lunch and you want to make him a sandwich. Gordon's not a man to rub the wrong way, so to make sure that he'll enjoy your culinary creation, you've asked around and assembled the data-set in Table 1. You will use it to learn a decision tree for predicting whether Gordon enjoys a sandwich based on the type of bread, whether or not it's buttered, the topping and whether aubergine, lettuce or tomato is added.

   (a) What are $H(Enjoy|Bread = White)$ and $H(Enjoy|Topping = Ham)$?

Table 1: sandwich enjoyment training data

Bread	Buttered	Topping	Veg	Enjoy
White	Yes	Cheese	Lettuce	Yes
Brown	No	Cheese	Lettuce	Yes
White	No	Cheese	Tomato	Yes
Brown	Yes	Cheese	Tomato	Yes
Brown	No	Cheese	Tomato	No
White	No	Ham	Lettuce	No
White	Yes	Ham	Lettuce	No
Brown	Yes	Cheese	Aubergine	No

(b) Which attribute would the decision tree building algorithm choose to use for the root of the tree (assume no pruning). Show your calculations.

(c) Draw the full decision tree that would be learned for this data (assume no pruning). Show your calculations. *Hint: Use the statement from question (e) to reduce the number of calculations.*

(d) Suppose we have a validation set as in Table 2. What will be the training set error and validation set error of the tree? Express your answer as the number of examples that would be misclassied.

Table 2: sandwich enjoyment validation data

Bread	Buttered	Topping	Veg	Enjoy
White	Yes	Ham	Tomato	Yes
Brown	Yes	Ham	Lettuce	No
Brown	Yes	Cheese	Lettuce	Yes

(e) Suppose that a binary-valued attribute splits a set of examples E into subsets E_1 and E_2, and that the subsets have p_1 and p_2 positive examples and n_1 and n_2 negative examples, respectively. Show that the attribute has zero information gain if the ratios $\frac{p_1}{p_1+n_1}$ and $\frac{p_2}{p_2+n_2}$ are the same.

3. **Bayesian Learning** *(23 Points)*

(a) Consider a naive Bayes classifier trained on the same dataset from Table 1. How would that predict `Enjoy` given the input `Bread = White, Veg = Tomato`? Show your calculations.

(b) Why can't such a classifier give a prediction about the input `Veg = Aubergine`?

(c) Model the following problem using a Bayesian network (only the structure is required, the quantitative part should not be given):

> If my car does not start, it is either due to the starter motor not turning over or I have run out of petrol. I can read the level of petrol in the tank from the fuel gauge. The most common reason for the starting motor to fail is a flat battery. I can check the battery by looking at the head-lights: if they burn brightly the battery is O.K., otherwise the battery is usually flat.

4. **Instance-based Learning** *(18 Points)*

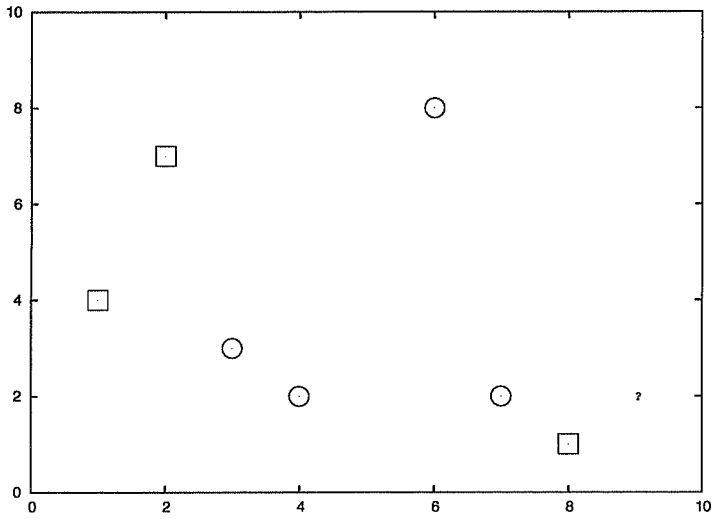

Figure 1: kNN sample: a data-set that maps two inputs (real-valued, range from 0 to 10) to a class. The two classes are denoted by squares and circles.

(a) Consider Fig. 1. Would 1-nearest neighbour classify the record `(9,2)` (denoted by the question mark in the figure) as "square" or as "circle"? Assume unweighted Euclidean distance.

(b) Same question, but for 3-nearest neighbour.

(c) Suppose you want to use k-nearest neighbour with an attribute *size*, that can be on of `small, medium, large`. Can you you define a distance measure for this attribute and if so, how?

(d) Explain why the nearest neighbour algorithm gets into trouble when using a large number of input variables.

5. **Hypothesis Testing** *(18 Points)*

(a) You have trained some classifier to predict loan repayment from past credit data. Over the 100 training examples, it correctly classifies 90. Over a separate set of 50 test examples, it correctly classifies 43. Give an **unbiased** estimate of the true error of this classier, and a 90% confidence interval around that estimate (you may give your confidence interval in the form of an expression). Note: $Z_{90} = 1.64$.

(b) What is the meaning of the 90% confidence interval you just calculated? Be sure to state any assumptions.

(c) What is "Leave One Out Cross-Validation"?