Resit Exam Machine Learning
2011-2012

6 June 2012

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** *(15 Points)*

   (a) (**True** or **False**): The back-propagation algorithm learns a globally optimal neural network with hidden layers.

   (b) (**True** or **False**): The error of a hypothesis measured over its training set provides a pessimistically biased estimate of the true error of the hypothesis (i.e., the real-world performance of the hypothesis would be better than the training set performance).

   (c) (**True** or **False**): Overfitting is more likely when the hypothesis space is small

   (d) (**True** or **False**): Because gradient descent is guaranteed to find a local optimum, it cannot overfit.

   (e) (**True** or **False**): When training neural networks, the network weights are updated more often in incremental or stochastic gradient descent than in batch gradient descent.

2. **Decision Trees** *(20 Points)*

   Summertime is festival time: Pinkpop, Pukkelpop, Parkpop, to name but a few. To help you choose which of the acts to go see at the next festival you're visiting, you will develop a model based on the data in table 1.

   (a) What are the entropies $H(Like|Style = Britpop)$ and $H(Like|Set\ length = 30)$? [1]

[1]Note: if your calculator can't do log_2, use one of the following equations: $log_2(x) = 1.44 \cdot ln(x)$ or $log_2(x) = 3.32 \cdot log_{10}(x)$. If you didn't bring a calculator, you may give the answer as an expression.

Table 1: Band data

Style	Seen before	Set length	Like
Grunge	Yes	20	No
RnB	Yes	20	No
Grunge	No	20	No
Britpop	No	20	No
Britpop	No	20	Yes
Grunge	No	30	Yes
RnB	No	20	Yes
RnB	Yes	30	Yes

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

(c) Draw the full decision tree that would be learned for this data (assume no pruning)

(d) What would be this decision tree's training set error? Express your answer as the percentage of records that would be misclassied.

3. **Bayesian Classifiers** *(20 Points)*

(a) Consider a naive Bayes classifier trained on the same dataset from Table 1. How would that predict `Like` given the input `Style = Grunge, Set length = 20`? Show your calculations.

(b) Suppose that in a classification problem one of the possible outcomes is extremely uncommon. Would you prefer to use MAP or MLE in this case? Why?

(c) When building a naive Bayes classifier, we assume that the attributes are conditionally independent given the class label. Why is that?

4. **Neural Nets and Regression** *(15 Points)*

(a) Does increasing the number of hidden nodes in a multilayer perceptron improve generalisation? Why (not)?

(b) Explain how you would use cross-validation to determine an appropriate number of hidden nodes in a neural network.

5. **Instance-Based Learning** *(15 Points)*

(a) The following data set (Fig. 1) shows classification data with two classes: plus and minus. Which of the two nearest neighbours classifiers has the largest Leave-One-Out Cross-validation error on this data set?

i. 1-NN

ii. 3-NN

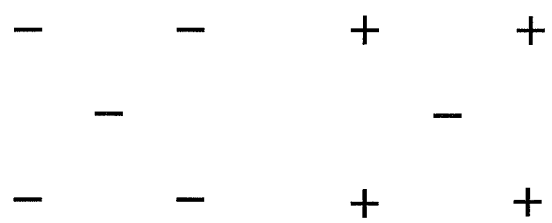

Figure 1: A classification data set

(b) Suppose that you want to use k-nearest neighbour on a data set that contains a non-numeric attribute 'weight' which can be *light*, *medium* or *heavy*. Could you re-code this attribute to be able to use it? How?

(c) The same for an attribute 'material' which can have the values *aluminium* or *plastic*.

6. **Hypothesis Comparison and Cross Validation** *(15 Points)*

(a) On a much larger data set than that in table 1, someone has developed a classifier that correctly classifies 700 out of 1000 test examples. Give an estimate of the true error of this classifier, and a 90% confidence interval around that estimate (you may give your confidence interval in the form of an expression). [2]

(b) In the above problem, if you increase the size of the test set by a factor of 10, assuming a similar error rate of 30%, what happens to the confidence interval?

(c) When tuning the parameters of a learning algorithm (e.g., the number of neighbours in kNN) to a given application, why is it necessary to use cross-validation or a test set?

[2]The table of z-values:

$N\%$:	50%	68%	80%	90%	95%	98%	99%
z_N:	0.67	1.00	1.28	1.64	1.96	2.33	2.58