Two hours

UNIVERSITY OF MANCHESTER
SCHOOL OF COMPUTER SCIENCE

M.Sc. in Advanced Computer Science

Machine Learning

Date: Wednesday 20\textsuperscript{th} January 2010

Time: 09.45 – 11.45

---

Answer any THREE questions from the FOUR questions provided

The use of electronic calculators is permitted provided they are not programmable and do not store text

[PTO]
Answer any THREE questions from the questions provided

1. a) State the equation for the Perceptron decision rule, being sure to fully explain the parameters, and their effect on the decision boundary. (4 marks)

b) Describe the general principle of a Support Vector Machine, including how it differs from a Perceptron. Be sure to give a definition of
   i) a support vector,
   ii) the margin,
   iii) a kernel. (6 marks)

c) Read the following passages, and then answer the question that follows it.

James works for a bank. He is given a historical database of 10,002 customers, and told to apply machine learning algorithms, enabling the bank to predict credit-worthiness of future customers. In the database, 502 of the customers have not paid back their loans, and so are considered unworthy of credit. Among these, James finds 2 customers who are bankrupt multi-millionaires, with overdrafts several thousand times the normal amount, so discards them as he thinks they are likely to adversely affect the learning algorithms.

The remaining 10,000 customers are characterised by their age, sex, living conditions, and bank transactions over the last 5 years. When choosing features to give to his algorithms, James includes only the last 6 months of transactions, and summarises the rest of the 5 years by calculating each customer's average overdraft amount.

James uses the 10,000 customers as input to the Perceptron learning algorithm, using the learning rate 0.25. Because of the Perceptron convergence theorem, he calculates that it should correctly classify at least 0.25 x 10,000 = 2,500 of the customers, and in fact it vastly exceeds his expectations, correctly classifying 9,500 of them. James happily concludes that the data must be linearly separable, but would like better performance.

Due to his conclusion on the Perceptron, he decides to apply a linear classifier, so picks the K-NN rule. He sets k=10,000, and finds that he can again only correctly classify 9,500 customers, but is not sure why. He concludes that a Perceptron with learning rate 0.25 is optimal, and recommends this to the bank.

List 4 things that are wrong with James’s methodology and understanding of Machine Learning, giving reasons for each, and explain to him why his K-NN rule behaved in the way it did. (5 marks)
2. a) The SVM optimisation problem is phrased as tradeoff between two objectives:

\[ L = \frac{1}{2} w^T w - \sum_{i=1}^{N} \alpha_i \{t_i f(x) - 1\} \]

What is the purpose of the \( \frac{1}{2} w^T w \) part?

What is the purpose of the \( \sum_{i=1}^{N} \alpha_i \{t_i f(x) - 1\} \) part? (2 marks)

b) Give a definition of “overfitting”, including how we know whether it has occurred. Explain how it manifests in decision tree classifiers, and state one way to control it. (5 marks)

c) Explain the advantages and disadvantages of “filters” versus “wrappers” for feature selection. (4 marks)

d) How is entropy used in feature selection problems? State the equation to calculate the entropy, and apply this to an example feature \( x \) with \( p(x=1) = 0.7 \), and \( p(x=0) = 0.3 \). (4 marks)

3. a) Explain how a probabilistic classifier differs from a non-probabilistic classifier. Give one example of each type of classifier. Give two examples of applications where a probabilistic classifier provides an advantage over a non-probabilistic classifier. (5 marks)

b) Explain the difference between the generative and discriminative approaches to probabilistic classification. Describe the main advantages of each approach relative to the other. (5 marks)

c) Show how Maximum Likelihood (ML) and Maximum a Posteriori (MAP) parameter estimation can be related using Bayesian principles. Give an example of how MAP learning can be used to improve the performance of a classifier relative to ML learning. (5 marks)
4. a) Define a 1st order Markov chain model. Write down a general expression for the probability of a sequence under the model. (3 marks)

Parts (b)-(e) of the question concern the following 1st order Markov chain model for generating phoneme sequences corresponding to the words “hi” and “bye”. Each phoneme can be repeated to indicate differences in duration, e.g. the sequence “hh-hh-ay-ay-ay” corresponds to one possible pronunciation of the word “hi”.

\[
\begin{array}{c}
\text{START} \\
\downarrow \quad 0.5 \quad 0.5 \\
\hline
\text{hh} \\
\downarrow \quad 0.5 \\
\text{ay} \\
\downarrow \quad 0.5 \\
\text{END} \\
\end{array}
\]

\[
\begin{array}{c}
\text{b) } \text{A sequence of three phonemes is generated from the above model. Is the sequence more likely to correspond to the word “hi” or to the word “bye”?} \quad (4 \text{ marks})
\end{array}
\]

\[
\begin{array}{c}
\text{c) You are asked to re-compute the model parameters from a new dataset D containing 6 sequences transcribed from examples of real speech. Given that } D=\{hh-ay, \text{ hh-ay-ay-ay, hh-ay, b-b-ay-ay, b-ay, b-ay-ay}\}, \text{ use maximum likelihood to compute new values for the model parameters}. \quad (3 \text{ marks})
\end{array}
\]

\[
\begin{array}{c}
\text{d) Given the maximum likelihood parameter estimates from part (c), what is the probability of the sequence “hh-hh-ay” under the model? Do you think your answer is a good estimate of the probability of this sequence occurring in some new speech data? (give reasons for your answer).} \quad (2 \text{ marks})
\end{array}
\]

\[
\begin{array}{c}
\text{e) The original parameter values given in the above figure were determined from previous analyses of speech data. Describe a method of parameter estimation which includes consideration of these original estimates while also being informed by new data. Give some justification for your proposed approach.} \quad (3 \text{ marks})
\end{array}
\]