A30406

Calculators may be used in this examination but must not be used to store text. Calculators with the ability to store text should have their memories deleted prior to the start of the examination.

UNIVERSITY OF BIRMINGHAM

School of Computer Science

Third Year – BSc Artificial Intelligence and Computer Science
First Year – UG Affiliate Computer Science/Software Engineering
Third Year – BSc Natural Sciences
Third Year – BSc Computer Science
Third Year – MSci Computer Science
Third Year – MEng Computer Science/Software Engineering
Third Year – BSc Mathematics and Computer Science
First Year – UG Affiliate Science without Borders Computer Science
Fourth Year – BSc Computer Science with Industrial Year
Third Year – MEng Computer Science/Software Engineering with Industrial Year

06 26428

Machine Learning

Summer Examinations 2015

Time allowed: 1 hour 30 minutes

[Answer ALL Questions]

[Answer ALL Questions]

1. (a) Your roommate, who is a bit of a slacker, is trying to convince you that money cannot buy happiness, citing a Harvard study showing that only 10% of happy people are rich. After giving it some thought, it occurs to you that this statistic is not very compelling. What you really want to know is what percent of rich people are happy. This would give a better idea of whether becoming rich might make you happy.

   (i) Use Bayes Rule to calculate this latter, reversed statistic, assuming that the Harward study is correct, and using two additional pieces of information: [10%]
   40% of people are happy; and
   5% of people are rich

   (ii) Based on the same information, what percentage of unhappy people are rich? [5%]

   (b) Is it possible to construct a Gaussian classifier for univariate inputs x, such that when it is used it will predict Class 1 if $x < -1$ or $x > 1$, and Class 2 if $-1 < x < 1$? You can draw a figure to help you explain. [10%]

2. (a) The following data is to be used for training an SVM:
Class one: (1,1), (2,2), (2,0)
Class two: (0,0), (1,0), (0,1).

(i) Plot the training points and, by inspection, determine the position of the optimal, maximum margin, decision boundary. [5%]
(ii) List the support vectors. [5%]

(iii) Explain in which way would a kernel be useful for classification? [5%]

(b) Answer the following questions about k nearest neighbours

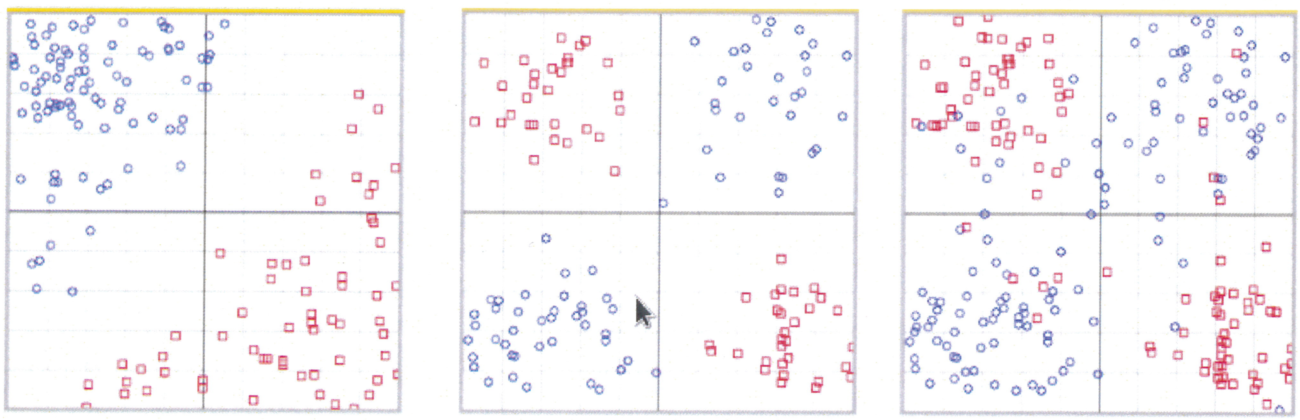
(i) Assume that the distance measure is not explicitly specified to you. Instead, you are given a "black box" where you input a set of instances P_1, P_2, .. P_n and a new example Q, and the black box outputs the nearest neighbour of Q, say P_i and its corresponding class label C_i. Is it possible to construct a k-NN classification algorithm based on this black box alone? If so, how and if not, why not? [5%]
(ii) The curse of dimensionality generically refers to a set of problems associated with data that is high dimensional, i.e. data that has a large number of attributes. Briefly describe 3 specific problems that affect the k nearest neighbours method when the data is high dimensional?

[5%]

(c) AdaBoost is susceptible to outliers.

(i) Explain what is the reason for this. [5%]
(ii) Suggest a simple heuristic modification to adaboost that relieves this problem. [5%]

3. Consider the three data sets (A), (B), (C) in the figure, where the circles are positive examples with two numeric features (i.e. the two coordinates of the points), and the squares are negative examples.



(A) (B) (C)

(a) For each of these three data sets describe the classification methods that you think will perform well, and say why. [10%]

(b) For each of these three data sets describe the classification methods that you think will perform poorly or fail, and say why. [10%]

4. (a) Define training error and generalisation error. [4%]

(b) Consider a finite set of functions, *H*, that map an input set *X* into the set of labels *{0,1}*. Let *L* be an algorithm that for any function c from *H*, and any training set *S* of *N* training points, drawn independently from some unknown distribution *D* over *X*, returns a hypothesis, h_S, that is consistent with the given training set. In this case it is known (as proved in the class) that for any choice of $\varepsilon > 0$ the generalisation error of h_S is upper bounded by ε with probability of at least $1-|H|exp(-N\varepsilon)$. Here, $|H|$ denotes the cardinality of a set (that is the number of elements in the set *H*). Using this result, give an upper-bound on the generalisation error of a consistent hypothesis that holds with probability at least 0.95. [6%]

5. (a) Draw a data set with 2 clusters that k-means would fail to discover and explain why it would fail. [5%]

   (b) Draw a data set with 3 clusters that could be discovered by k-means. Can k-means still occasionally fail to discover these clusters? Why or why not? [5%]