THE UNIVERSITY OF ADELAIDE
AUSTRALIA

Primary Examination, Semester 2, 2014

Introduction to Statistical Machine Learning
COMP SCI 4401, 7401

Official Reading Time:	10 mins
Writing Time:	120 mins
Total Duration:	130 mins

Questions	Time	Marks
Answer all 6 questions	120 mins	100 marks 100 Total

Instructions

- Begin each answer on a new page in the answer book.
- Examination material must not be removed from the examination room.

Materials

- Books, lecture notes, slides print-out, calculators and paper dictionaries (English to foreign language) are permitted.
- The use of internet is not permitted.

DO NOT COMMENCE WRITING UNTIL INSTRUCTED TO DO SO

Machine Learning Definition

Question 1

(a) Explain the definition of machine learning in one sentence.

[4 marks]

(b) What are the three main types of learning? Explain the difference in their training data.

[6 marks]

(c) Assume data $(\mathbf{x}, y)$ are from an unknown but fixed distribution q, *i.e.* $(\mathbf{x}, y) \sim q$. The training (empirical) data are $\{(\mathbf{x}_i, y_i)\}_{i=1}^m$, and the loss function is $\ell(\mathbf{x}, y, \mathbf{w})$. Write down risk and empirical risk.

[4 marks]

(d) Explain the difference between regression and classification.

For the following problems, do you want to model as a regression problem or a classification, and why?

1. estimate rain or not tomorrow
2. estimate the amount of rainfall tomorrow
3. estimate the height and weight of an elephant.

[6 marks]

[Total for Question 1: 20 marks]

Support Vector Machines (SVMs) and Kernels

Question 2

(a) Write down the primal form of the soft margin SVMs.

[4 marks]

(b) Write down Lagrangian function of above problem.

[4 marks]

(c) Write down the dual form of the soft margin SVMs.

[4 marks]

(d) Given kernel $\kappa(\mathbf{x}, \mathbf{x}')$, set of support vectors S, please write the decision function $\text{sign}[\langle \mathbf{w}, \mathbf{x} \rangle + b]$ in terms of kernel and support vectors.

[4 marks]

(e) Let $\mathbf{u} = [\mathbf{w}; b]$ and $\mathbf{z} = [\mathbf{x}; 1]$. We can rewrite $(\langle \mathbf{w}, \mathbf{x} \rangle + b)$ as $\langle \mathbf{u}, \mathbf{z} \rangle$. This means if we augment the training data $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ to $\{(\mathbf{z}_i, y_i)\}_{i=1}^n$, where $\mathbf{z}_i = [\mathbf{x}_i; 1]$, we only need to learn one parameter $\mathbf{u}$ instead of two parameters $\mathbf{w}$ and b.

1. Please write down the primal form of the soft margin SVMs using decision function $\text{sign}[\langle \mathbf{u}, \mathbf{z} \rangle]$.

2. Is the new primal form equivalent to the old primal form?
3. Please prove your answer for above question (*i.e.* using derivation to show why or why not equivalent).

[4 marks]

[Total for Question 2: 20 marks]

Boosting

Question 3

(a) The AdaBoost learning algorithm takes an input training dataset $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$. Briefly describe the algorithm.

[4 marks]

(b) What is a weak classifier (or base classifier)?

[4 marks]

(c) What is a strong classifier?

[4 marks]

(d) How does AdaBoost select weak classifiers?

[4 marks]

(e) In AdaBoost, how does one update the weak learner's weight? Write down the relevant formula. (4 points)

[4 marks]

(f) A requirement for the weak learner used in AdaBoost is that at each iteration, the selected best weak learner must perform better than random guess. In other words, the weighted error of the selected weak learner must be less than 0.5. Mathematically explain this. (6 points)

[5 marks]

[Total for Question 3: 25 marks]

Principal Component Analysis (PCA)

Question 4

(a) What are the main purposes of a principal component analysis?

[5 marks]

(b) In the case of using PCA for dimensionality reduction, How does one determine the appropriate number of principal components to retain?

[5 marks]

[Total for Question 4: 10 marks]

Learning Theory

Question 5

(a) Assume $(\mathbf{x}, y) \sim P$, and given training data $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$, please write down the generalisation error, and training error.

[4 marks]

(b) Explain growth function and VC dimension.

[2 marks]

(c) Explain VC dimension.

[3 marks]

(d) Give an example to each of the following types of generalisation bounds (including their names and the mathematical statements).

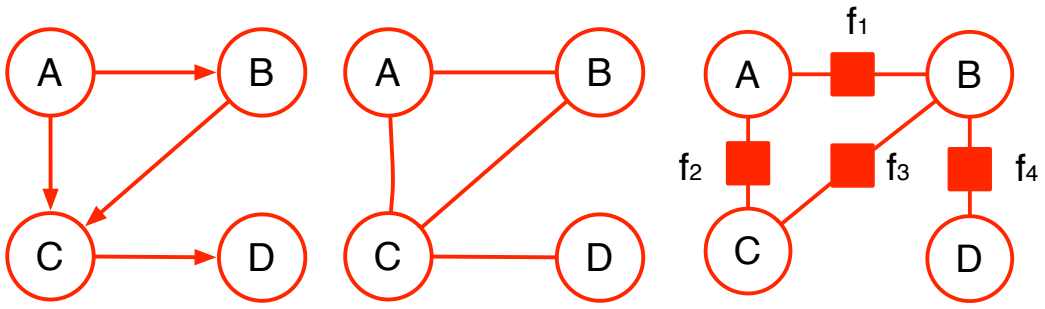
1. Counting the hypotheses
2. Counting the outputs

[6 marks]

[Total for Question 5: 15 marks]

Probabilistic Graphical Models

Question 6



(a) Directed graph (b) Undirected graph (c) Factor graph

Figure 1: Three graphical models

(a) write down general factorisation rules of graphical models including directed acyclic graphs, undirected graphs and factor graphs.

[3 marks]

(b) factorise the joint distribution for the graphical model in Figure 1(a) (hint: apply above rules).

[2 marks]

(c) write down MAP inference and Marginal inference in the context of Figure 1(a).

[2 marks]

(d) write down variable elimination for marginal inference to compute $P(B)$ in Figure 1(a).

[3 marks]

[Total for Question 6: 10 marks]

End of exam