The University of Melbourne

School of Computing and Information Systems

COMP90051
Statistical Machine Learning
2020 Semester 2 – Practice Exam

Identical examination papers: None

Exam duration: 180 minutes

Reading time: 15 minutes

Length: This paper has 5 pages (real exam is longer) including this cover page.

Authorised materials: Lecture slides, workshop materials, prescribed reading, your own project reports.

Calculators: permitted

Instructions to students: The total marks for the real exam is 180 (practice exam has fewer questions and marks), corresponding to the number of minutes available. The mark will be scaled to compute your final exam grade.

This paper has three parts, A-C. You should attempt all the questions.

This is an open book exam. You should enter your answers in a Word document or PDF, which can include typed and/or hand-written answers. You should answer each question on a separate page, i.e., start a new page for each of Questions 1–6 – parts within questions do not need new pages. Write the question number clearly at the top of each page. You have unlimited attempts to submit your answer-file, but only your last submission is used for marking.

You must not use materials other than those authorised above. You are not permitted to communicate with others for the duration of the exam, other than to ask questions of the teaching staff via the Piazza 'exam' folder. Your computer, phone and/or tablet should only be used to access the authorised materials, enter or photograph your answers, and upload these files.

Library: This paper is to be lodged with the Baillieu Library.

 Continued overleaf ...

COMP90051 Statistical Machine Learning
Practice Exam

Semester 2, 2020

Total marks: 180 in real exam based on more questions; 90 in this practice exam

Students must attempt all questions

Section A: Short Answer Questions [25 marks]

Answer each of the questions in this section as briefly as possible. Expect to answer each question in 1-3 lines, with longer responses expected for the questions with higher marks.

Question 1: [25 marks]

(a) In words or a mathematical expression, what quantity is minimised by *linear regression*? [5 marks]

We would like to minimise the sum of square error

(b) In words or a mathematical expression, what is the *marginal likelihood* for a *Bayesian probabilistic model*? [5 marks] $\int p(x|\theta)P(\theta)d\theta = P(x)$

(c) In words, what does $\Pr(A, B \mid C) = \Pr(A \mid C)\Pr(B \mid C)$ say about the *dependence* of A, B, C? [5 marks] Given event C, event A and B is independent.

(d) What are the *free parameters* of a *Gaussian mixture model*? What algorithm is used to fit them for *maximum likelihood estimation*? [10 marks]

For k cluster, we have one probabilities for each clusters(in total k probabilities and they can sum to 1), a mean vector for each of the k cluster, and k symmetric PD covariance matrix of the k clusters each.

expectation maximisation

Section B: Method & Calculation Questions [45 marks]

In this section you are asked to demonstrate your conceptual understanding of methods that we have studied in this subject, and your ability to perform numeric and mathematical calculations. NOTE: in the real exam, a small number of questions from this section will be a bit harder/longer than others.

Question 2: [10 marks]

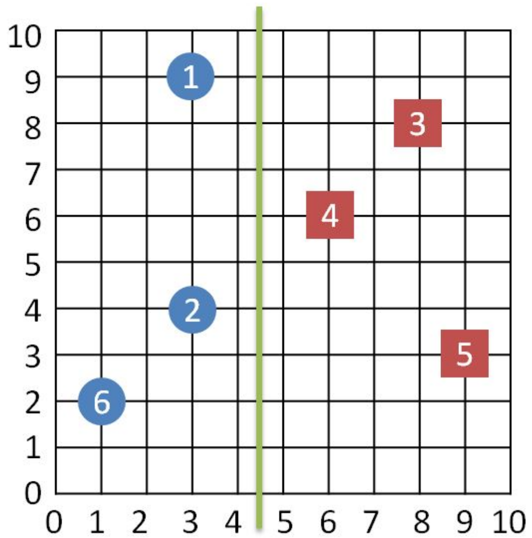
(a) Consider a 2-dimensional *dataset*, where each point is represented by two *features* and the *label* (x_1, x_2, y). The features are binary, the label is the result of XOR function, and so the data consists of four points $(0,0,0)$, $(0,1,1)$, $(1,0,1)$ and $(1,1,0)$. Design a *feature space transformation* that would make the data *linearly separable*. [5 marks] Use the radial basis function

(b) How does SVM handle data that is not linearly separable? List two possible strategies [5 marks]

1. use pre-defined kernel method that map the data to other dimension and hope they are linear separable at that dimension.
2. Use soft margin that have some degree of fault tolerance

Question 3: [10 marks]

Consider the data shown below with *hard-margin linear SVM decision boundary* shown between the classes. The right half is classified as red squares and the left half is classified as blue circles. Answer the following questions and explain your answers.



1, 2, 4 as they all one the margin

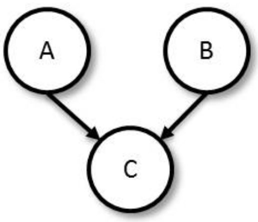
(a) Which points (by index 1–6) would be the *support vectors* of the *SVM*? [5 marks]

(b) What is the value of the *hard margin SVM loss* for point 3? [5 marks]

Zero, since point 3 is on the right side

Question 4: [15 marks]

Consider the following directed PGM



where each random variable is Boolean-valued (True or False).

(a) Write the format (with empty values) of the *conditional probability* tables for this graph. [5 marks]

P(C | A, B)

(b) Suppose we observe n sets of values of A, B, C (complete observations). The *maximum-likelihood principle* is a popular approach to training a model such as above. What does it say to do? [5 marks] It says to choose values in the tables that maximise the likelihood of the data, which in this case $\arg\max_{tables} \prod_{i=1}^{n} \Pr(A = a_i) \Pr(B = b_i) \Pr(C = c_i \mid A = a_i, B = b_i)$

(c) Suppose we observe 5 training examples: for (A, B, C) — (F, F, F); (F, F, T); (F, T, F); (T, F, T); $(T, T; T)$. Determine *maximum-likelihood estimates* for your tables. [5 marks]

A: T 0.4, F 0.6

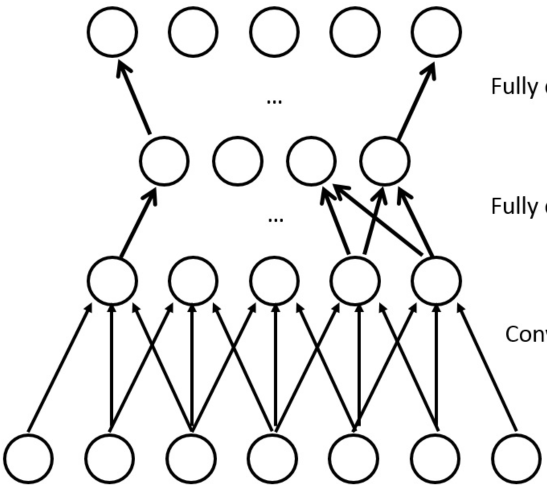
B: T 0.4, F 0.6

C:

A	F	F	T	T
B	F	T	F	T
F	0.5	0.5	0	1
T	1	0	0	1

Question 5: [10 marks]

How many *parameters* does the following *convolutional neural network* have (exclude the *bias*)? Show your working.



$5 \times 4 = 20$

$4 \times 5 = 20$ +

3 +

43

Section C: Design and Application Questions [20 marks]

In this section you are asked to demonstrate that you have gained a high-level understanding of the methods and algorithms covered in this subject, and can apply that understanding. Answers should be about 1 page in length for each question.

Question 6: [20 marks]

Your task is to design a system for alerting residents of the Dandenongs that they should evacuate for an impending bushfire. The Dandenongs is an area of Victoria that suffers from regular bushfires in the summer when temperatures are high, and humidity low. When fires are close to a fictional town called Bayesville, you must alert residents that they should evacuate. If fires are too close, then you should not advise evacuation as residents are safer if they stay where they are (at home).
The Country Fire Association has deployed sensors around the area that monitor whether a fire is in progress at each sensor's location; in particular if any of three sensors S_1, S_2, S_3 are 'on' then residents should evacuate. However if either of the closer sensors C_1, C_2 are 'on' then residents should stay put.

(a) Model the above problem as a *directed probabilistic graphical model* (PGM). In particular, you need not provide any probabilty tables, just the graph relating random variables S_1, S_2, S_3, C_1, C_2 and an additional r.v. A for alerting residents to evacuate. [5 marks]

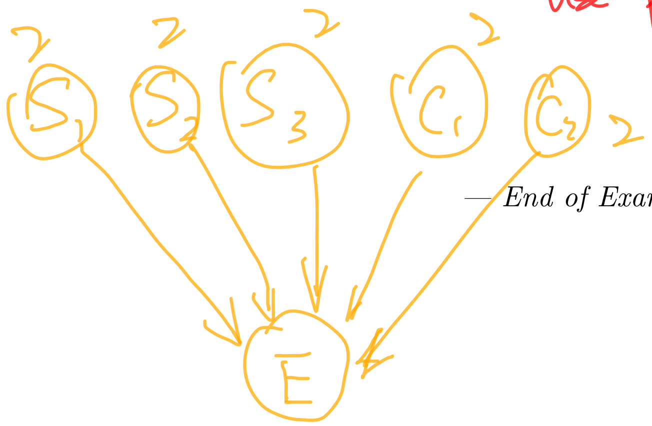
(b) How many *conditional probability tables (CPTs)* should be specified for your model, and what should these tables' dimensions be? [4 marks]

(c) Suppose you are told by the Fire Commissioner that the sensors are always accurate. What could you say about your CPTs? [4 marks] Since always correct then $(S_1 \vee S_2 \vee S_3) \wedge \neg(C_1 \vee C_2)$

(d) How would you change your model if the Commissioner then tells you that the sensor S_1 is not perfectly accurate? [4 marks]

(e) Given this final model, assuming you have trained it and completed all the necessarily CPTs, how would you use it to drive the alarm to evacuate? [3 marks] use probability inference to determine $P(E=\text{True})$ from the observation

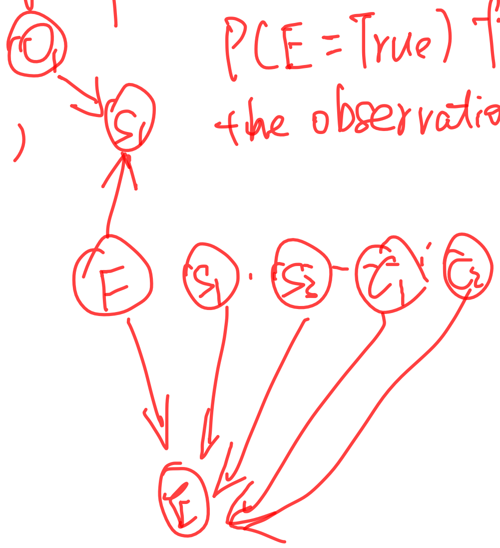
a.



5 dimension

6 variables

d)



— End of Exam —