

The University of Melbourne

Department of Computing and Information Systems

COMP90049

Introduction to Machine Learning

June 2021

Identical examination papers: None

Exam duration: 120 minutes

Reading time: Fifteen minutes

Length: This paper has 6 pages 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 this paper is 120, 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 Question 1, Question 2 etc, – 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. If your last submission arrives more than 15 minutes after the due time, you will fail the exam.

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 discussion board. Your computer, phone and/or tablet should *only* be used to access the authorised materials, enter or photograph your answers, and upload these files. The work you submit **must be based on your own knowledge and skills**, without assistance from any person or unauthorized materials.

COMP90049 Introduction to Machine Learning Final Exam

Semester 1, 2021

Total marks: 120

Students must attempt all questions

Section A: Short answer Questions [40 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: [40 marks]

- (a) Name three differences between exact optimization and Gradient descent. [6 marks]
- (b) Indicate the best alignment of the concepts under (a) to the concepts under (b). Many-to-one and one-to-many alignments are possible. [3 marks]

(a)	(b)
clustering	supervised
classification	semi-supervised
regression	unsupervised

- (c) [6 marks] On a given test data, a classifier detects 4 TP, 3 TN, 6 FP, and 0 FN. What are precision, recall and F-score (assume $\beta=1$) of the classifier?
- (d) [3 marks] Consider the following set of evaluation metrics

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$
$$\text{Precision} = \frac{TP}{TP + FP}$$
$$\text{Recall} = \frac{TP}{TP + FN}$$
$$\text{Error Rate} = 1 - \text{Accuracy}$$

1. What types of machine learning algorithms can be evaluated with these measures? [1 mark]
 2. Explain why. [2 marks]
- (f) Consider the following two tasks: (1) predicting whether a job applicant is successful based on the characteristics of their CV; (2) Predicting the expected salary of a job applicant based on the characteristics of their CV. (i) For each task, (i) name the corresponding machine learning concept. (ii) Justify your choice. [3 marks]

Section B: Method & Calculation Questions [55 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.

Question 2: K-Nearest Neighbors [8 marks]

With respect to the following data set of 6 instances with 3 attributes and two classes F and T, plus a single test instance labelled "?":

instance #	ele	fed	aus	CLASS
1	1	1	1	F
2	1	0	0	F
3	1	1	0	T
4	1	1	0	T
5	1	1	1	T
6	1	1	1	T
7	0	0	0	?

Explain why a model with $K = 1$ will make a different prediction compared to a model with $K = 3$ on the given test instance. You do not need to show your work for this question, but should provide an explanation which refers to the data.

Question 4: K- Means [10 marks]

Consider the following data set of 6 instances with 3 attributes and two classes F and T, plus a single test instance labelled "?":

instance #	ele	fed	aus	CLASS
1	1	1	1	F
2	1	0	0	F
3	1	1	0	T
4	1	1	0	T
5	1	1	1	T
6	1	1	1	T
7	0	0	0	?

Exclude the class labels from the dataset, and cluster all 7 instances using the method of “k-means”. Apply the Manhattan Distance as a similarity measure; use the second (1,0,0) and third (1,1,0) instances as seeds. Show your mathematical working.

Question 5: Data Sampling and Evaluation [3 marks]

Consider the following data set of instances.

#	X1	X2	y
1	7	0	1
2	9	1	1
3	1	5	0
4	3	4	1

1. Is this data set linearly separable? Graphically demonstrate your answer. [1 mark]

2. Assume that instances 1-2 are the training set and instances 3-4 the test set. Further assume that all parameters initialized to 0.3. Compute the negative conditional log-likelihood of the training data set. [2 marks]

Question 6: Decision Trees [7 marks]

In the following dataset every row represents a patient with three descriptive features, i.e., fever, dry cough, and headache, and Class indicates the label of each instance. Assume we are interested in building a decision tree to determine whether a patient has flu or cold.

Patient #	Fever	Dry cough	Headache	CLASS
1	yes	no	mild	Flu
2	yes	yes	severe	Flu
3	no	yes	moderate	Flu
4	no	no	moderate	Cold
5	yes	no	severe	Cold
6	no	no	severe	Cold

1. Determine the attribute that a decision tree would select first based on the information gain criteria. (Note: you need to provide the results of each step to get full marks. Show your work for computing information gain for all three attributes. [6 marks])

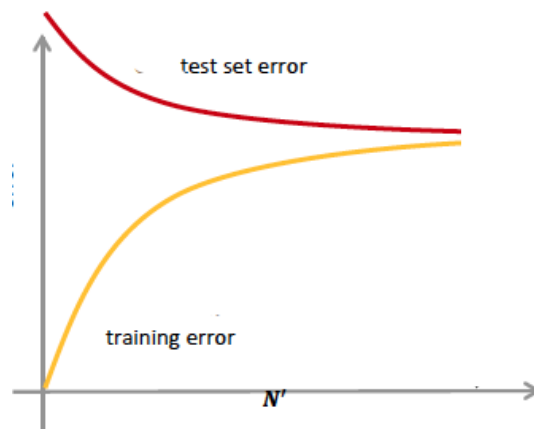
You may need to use the following results:

$$\log_2(1/2) = -1, \log_2(1/4) = -2, \log_2(3/4) = -0.41, \log_2(1/3) = -1.58, \log_2(2/3) = -0.58, \log_2(1) = 0$$

2. Calculate the Total error of the best decision stump you built in the previous step. [1 mark]

Question 7: Evaluation [7 marks]

Given the following learning curve for Naive Bayes, where N' is the number of samples used in the training set, answer the following questions:



1. How can you detect whether a model is overfitting or underfitting the data using the learning curve? [2 marks]

2. Does the Naive Bayes model in the above plot have high bias or high variance? Why? [2 marks]
3. Briefly describe one strategy to overcome underfitting. (1-2 sentences) [3 marks]

Question 8: Multi-layer Perceptron [16 marks]

Consider the following labelled data set of 4 instances, 3 features (X1 ... X3) and label Y. Instances 1 and 2 are training instances, and instances 3 and 4 are test instances.

N.B: Show your mathematical working for all calculations.

ID	X1	X2	X3	Y
1	0.1	0.9	-0.9	B
2	0.1	0.08	-0.5	A
3	6.4	0.9	9.8	A
4	0.3	0.9	4.5	C

The following formulas might be useful for answering the questions:

- Rectified linear unit (ReLU) function: $z = \max(\sum_i a_i, 0)$, i.e., returning either 0 or the summed inputs, whichever is larger.
- Softmax: $\text{softmax}(a_i) = \frac{\exp(a_i)}{\sum_k \exp(a_k)}$, where k ranges over all elements in vector a and i indexes one specific element.

Please answer the following questions.

1. Describe the given machine learning task, making sure to specify the concept, features and labels. Justify your definitions. [2 marks]
2. Construct a multi-layer perceptron which predicts a probability distribution over possible outputs, which consists of an input layer, one hidden layer of width 2, and an output layer. Define all necessary parameters including output functions and loss. Draw your multi-layer perceptron. [3 marks]
3. Initialize all MLP parameters according to the formula $\theta_{in,out}^{layer} = layer + in \times out$. (For example, in weight layer 2 the weight connecting incoming node 1 to outgoing node 2 is $\theta_{1,2}^2 = 2 + 1 \times 2 = 4$. Assume a constant bias of 1.0. (i) Perform *only* the *forward* pass of a single training epoch. For the hidden layers, assume the "Rectified linear unit" (ReLU) activation function. For the remaining functions, use your choices from question 2. (ii) What is the accuracy of your model for the training instances? [7 marks]
4. Compute the loss of your model, given your results in question 3, and choice of loss function in question 2. [4 marks]

Question 9: Feature Engineering [4 marks]

Many machine learning algorithms benefit from *feature normalization* as a pre-processing step. During this step, each feature is normalized to zero mean and unit variance.

- Give the formula for the normalized feature \tilde{x}_j as a function of the original feature x_j and the mean μ_j and standard deviation σ_j of that feature. [2 marks]
- Provide one concrete example machine learning problem (data, features, concepts, ...) where you expect normalisation to be particularly useful. [2 marks]

Section C: Design and Application Questions [25 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. Expect your answer to each question to be from one third of a page to one full page in length. These questions will require significantly more thought than those in Sections A–B, and should be attempted only after having completed the earlier sections.

Question 10: Insurance Policy [25 marks]

You are a manager of a life insurance company and want to provide optimal insurance quotes to your potential customers. The quotes fall into one of three categories ‘high’, ‘medium’ or ‘low’ premium. Your company is so popular that you cannot sort through all applications manually. Instead, you want to pre-sort applications into meaningful groups. Each application comes with features such as

- Name of applicant
- Age of applicant
- Favorite color of applicant
- Longest period spent in hospital
- Marital status of applicant
- Gender of applicant

Please answer the following questions with respect to the machine learning problem introduced above.

1. Describe the machine learning concept and features underlying this task. [3 marks]
2. Assume you have access to the following ML methods: (a) Decision trees; (b) neural networks; (c) k-means. For each algorithm, state whether it is appropriate in this situation as well as a reason for your decision [6 marks]
3. Now assume a slightly different situation where you (a) have access to a set of 50 admission decisions from previous years. Describe how this new information will change (a) your machine learning approach. [8 marks]
4. Further questions e.g., on evaluation or feature selection or bias ... [8 marks]

— End of Exam —