Supplementary Exam: Machine Learning (COMP30027_2021_SM1)

Started: Jul 20 at 15:00

Quiz Instructions

The University of Melbourne
School of Computing and Information Systems

Final Examination, Semester 1, 2021
COMP30027 Machine Learning

Reading Time: 15 minutes. **Writing Time:** 2 hours

Total Time: 2.25 hours

Instructions to Students:

The exam includes 20 questions worth a total of 120 marks, making up 60% of the total assessment for the subject.

- This exam includes a combination of multiple-choice questions, short-answer questions, and longer-response questions. Please answer all questions in the fields provided.
- This is a timed quiz. The time remaining is shown in the quiz window and will continue to count down even if you leave the Canvas site.
- It is recommended that you do not close your browser while working on this quiz.
- At the end of the time limit, your answers will be submitted automatically.

Authorised Materials: This exam is open-book. While undertaking this assessment you ***are permitted to***:

- make use of textbooks and lecture slides (including electronic versions) and lecture recordings
- make use of your own personal notes and material provided as part of tutorials and practicals in this subject
- make use of code that has been provided as part of this subject, or that you have written yourself
- use calculators, code, or mathematical software to compute numeric answers

While you are undertaking this assessment you ***must not***:

- make use of any messaging or communications technology
- make use of any world-wide web or internet-based resources such as Wikipedia, Stack Overflow, or Google and other search services
- act in any manner that could be regarded as providing assistance to another student who is undertaking this assessment, or will in the future be undertaking this assessment.

The work you submit ***must be based on your own knowledge and skills***, without assistance from any other person.

Technical support

This exam is a Canvas Quiz. Technical support for this exam can be accessed at: **https://students.unimelb.edu.au/your-course/manage-your-course/exams-assessments-and-results/exams/technical-support (https://students.unimelb.edu.au/your-course/manage-your-course/exams-assessments-and-results/exams/technical-support)** .

Academic Integrity Declaration

By commencing and/or submitting this assessment I agree that I have read and understood the **University's policy on academic integrity. (https://academicintegrity.unimelb.edu.au/#online-exams)**

I also agree that:

1. Unless paragraph 2 applies, the work I submit will be original and solely my own work (cheating);
2. I will not seek or receive any assistance from any other person (collusion) except where the work is for a designated collaborative task, in which case the individual contributions will be indicated; and,
3. I will not use any sources without proper acknowledgment or referencing (plagiarism).
4. Where the work I submit is a computer program or code, I will ensure that:
    a. any code I have copied is clearly noted by identifying the source of that code at the start of the program or in a header file or, that comments inline identify the start and end of the copied code; and

b. any modifications to code sourced from elsewhere will be commented upon to show the nature of the modification.

Short response questions

This section asks you to demonstrate your conceptual understanding of various methods we have studied in this subject, your ability to apply them or evaluate them in the context of specific cases, and your ability to perform the numeric calculations involved.

Question 1 — 3 pts

Numeric attributes should be discretised before use in decision trees. Briefly describe one discretisation strategy for decision trees. (1-2 sentences)

Edit View Insert Format Tools Table

20px Paragraph

p 0 words </>

Question 2 2 pts

What is the entropy of a fair six-sided die? A fair six-sided die can produce 6 different outcomes (1,2,3,4,5,6) each with probability 1/6.

- ○ 2.6
- ○ 0.17
- ○ 0.86
- ○ 1

Question 3 3 pts

In what situations would we prefer to use gain ratio instead of information gain? (About 1 sentence)

Edit View Insert Format Tools Table

20px Paragraph B I U A T²

p 0 words </>

Question 4 2 pts

Suppose you wish to use a binary classifier (e.g., SVM or logistic regression) on a multiclass classification problem. Given N classes ($N > 2$), how many models are needed if the one-vs-one classification approach is adopted?

Question 5 3 pts

How can you balance model bias and variance? Give one strategy as example, and briefly describe this strategy. (2-3 sentences)

p 0 words </>

Question 6

3 pts

Adding regularisation in a regression model introduces a hyperparameter λ, and the objective function becomes:

$$\hat{\beta} = argmin(Error(\beta; \{\mathbf{X}, Y\}) + \lambda\psi(\beta))$$

What is the impact of increasing λ? (2-3 sentences)

Edit View Insert Format Tools Table

20px Paragraph B I U A T²

p 0 words </>

Question 7

2 pts

Which of the following statements are TRUE? Select all that apply.

- [ ] A combination of classifiers is always more accurate than the base classifiers.
- [ ] In the bagging algorithm, some instances in the original dataset may not appear in any bagged datasets.
- [ ] A bagging step is involved in several ensemble learning models, such as random forest and boosting.
- [ ] In AdaBoost, the weight of an incorrectly classified instance will be decreased to avoid misleading the classifiers.

Question 8

4 pts

Describe the strategies to improve the generalisability for: (1) SVM; (2) logistic regression, respectively. (1-2 sentences for each model)

Edit View Insert Format Tools Table

20px Paragraph B I U A T²

p 0 words </>

Naïve Bayes

The follow questions relate to probability and naïve Bayes.

Question 9

10 pts

A university conducted a poll of students to gauge support for a new policy. The survey evenly sampled students from three disciplines (Mathematics, Music, Chemistry). The number of students in each group who support or do not support the policy are given in the table below.

	Mathematics	Music	Chemistry
Support	60	20	50
Do not support	40	5	25

(1) [2 point] What is the marginal probability of Support?

(2) [2 point] What is the joint probability P(Music, Do not support)?

(3) [6 points] Suppose we recruited an additional 3 students (2 from Mathematics and 1 from Music), and polling one at random, we find that they support the new policy. Is this student most likely from Mathematics or Music? State your reasoning and provide posterior probability values for each discipline to support your answer.

Edit View Insert Format Tools Table

20px Paragraph B I U A T²

p 0 words

Question 10

3 pts

When training a naïve Bayes classifier, is it better to use stratified cross-validation or randomly-sampled cross-validation? Justify your answer in terms of the properties of naïve Bayes. (1-2 sentences)

Edit View Insert Format Tools Table

20px Paragraph

p 0 words

Decision Trees

The following questions are related to decision tree algorithms.

Question 11

10 pts

Given the following dataset, with four features and one class label PLAY:

ID	Outlook	Temperature	Humidity	Play
A	sunny	cold	low	Yes
B	sunny	hot	low	Yes
C	rainy	hot	high	No
D	rainy	cold	high	No
E	cloudy	hot	high	No
F	cloudy	cold	low	Yes

(1) [8 points] What is the mutual information of Temperature?

(2) [2 points] Briefly describe how to select features using mutual information (2-3 sentences)

Edit View Insert Format Tools Table

20px Paragraph

p 0 words

Evaluation

The following questions pertain to model evaluation.

Question 12

10 pts

A machine learning algorithm was trained to classify the sentiment of Twitter messages into 3 classes (positive, negative, neutral). It was tested on a set of 1000 messages and the performance is summarized in the following confusion matrix (the number in each cell indicates the number of messages of a particular class given a particular label):

		Classifier prediction		
		Positive	**Negative**	**Neutral**
True class	**Positive**	100	50	50
	Negative	50	100	50
	Neutral	100	100	400

(1) [7 points] Is the classifier equally good at classifying each type of tweet? Report one-vs.-rest precision and recall to support your answer.

(2) [3 points] What baseline algorithm would you use for this classification problem and why?

Edit View Insert Format Tools Table

20px Paragraph

p 0 words

Instance-based learning

The following questions are about instance-based learning.

Question 13 10 pts

The following dataset shows product information and class labels for 6 products:

Product_ID	Price ($)	Sales_Quantity	Brand	Class_Label
A	6	2000	Private	-

B	7	1500	Private	-
C	4	3000	National	+
D	8	1500	National	+
E	2	5000	Private	-
F	7	1000	Private	+

Choose sensible distance metrics based on the data above, and use it to predict the **Class_Label** of the test instance (Product_ID = G, Price = $5, Sales_Quantity = 2500, Brand = National) according to the method of 3-NN. Use majority voting (no weighting).

(1) [3 points] Briefly describe your distance metric in 2-3 sentences

(2) [4 points] Show your distance values, give the Product_IDs of the top 3 neighbours, and prediction label for the test instance.

(3) [3 points] If the weighting strategy is used in the voting, such as inverse linear distance, how would it impact the result? Briefly explain at conceptual level in 2-3 sentences, no calculation needed.

Edit View Insert Format Tools Table

20px Paragraph B I U A T²

p 0 words </>

Question 14 — 3 pts

The performance of kNN may be undermined when there is a large number of attributes, a problem known as the curse of dimensionality. Briefly explain how the curse of dimensionality impacts models like kNN. (2-3 sentences)

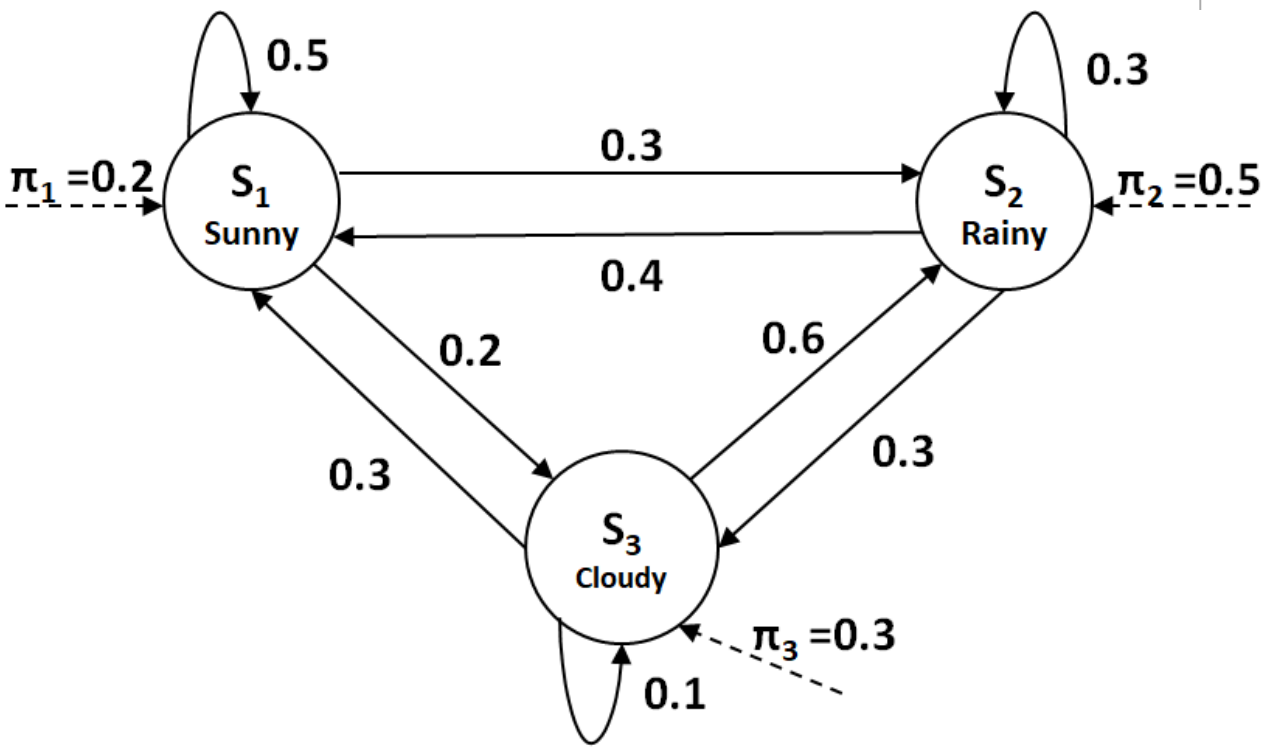
p 0 words </>

Hidden Markov Models

The next two questions relate to hidden Markov models (HMMs).

Question 15 — 7 pts

Suppose we have constructed a hidden Markov model (HMM) to represent the tasks completed on different days. The day's weather (Sunny, Rainy, Cloudy) is the hidden state, and the task completed (Task 1, Task 2, Task 3) is the observation. The transition probabilities and initial state distributions are illustrated in the figure below. The output probability matrix is given in the table below.



	Task 1	Task 2	Task 3
Sunny	0.6	0.3	0.1
Rainy	0.05	0.15	0.8
Cloudy	0.3	0.4	0.3

Given an observation sequece on the first three days (t = 1, 2, 3) as **Task1-Task2-Task2**, what is the most probable state sequece? The initialisation and the induction steps have been partially completed, and the values are given in the table below. Please complete the last column of the table ($\delta_3\ (S_1)$ and $\psi_3\ (S_1)$), and give the most porbable state sequence.

	t = 1	t = 2	t = 3
$\delta_t\ (S_1)$	0.12	0.018	**?**
$\psi_t\ (S_1)$	0	S_1	**?**

$\delta_t\ (S_2)$ $\psi_t\ (S_2)$	0.025 0	0.0081 S_3	0.000864 S_3
$\delta_t\ (S_3)$ $\psi_t\ (S_3)$	0.09 0	0.0096 S_1	0.00144 S_1

Edit View Insert Format Tools Table

20px Paragraph B I U A T²

p 0 words </>

Question 16 5 pts

Part-of-speech is a typical application of hidden Markov model (HMM). Given a paragraph as follows and the part-of-speech tag of each word:

Profits**/N** soared**/V** at**/P** Boeing**/N** Co.**/N** , easily**/ADV** topping**/V** forecasts**/N** on**/P** Wall**/N** Street**/N** , as**/P** their**/POSS** CEO**/N** Alan**/N** Mulally**/N** announced**/V** first**/ADJ** quarter**/N** results**/N** .

(1) [2 points] What are the sates of HMM model in the given example? List all the states.

(2) [3 points] Briefly describe at conceptual level the values stored in the transition probability matrix A for part-of-speech application. (1-2 sentences).

p 0 words </>

Neural networks

The following questions are related to neural networks.

Question 17 5 pts

In the formula for a neural network neuron:

$$y_i = f\left(\left[\sum_j w_j x_{ij}\right] + b\right)$$

(1) [2 points] What is f? Explain its role in a one-unit network (perceptron).

(2) [3 points] When would it be appropriate to use a linear function for f? When is this inappropriate?

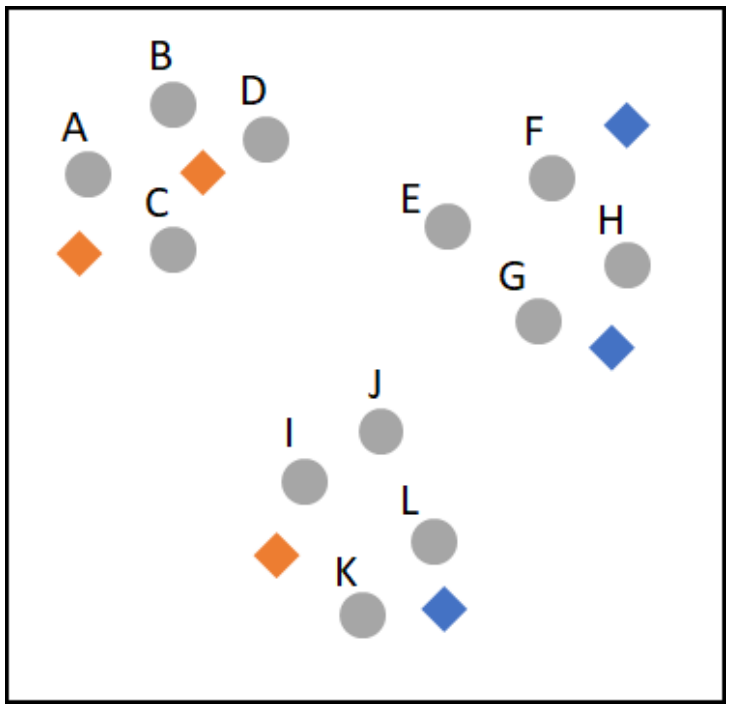
p 0 words </>

Un- and semi-supervised learning

The following questions relate to un- and semi-supervised learning.

Question 18 5 pts

Suppose you are training an algorithm on dataset shown below. Instances have two attributes, which are continuous numeric values, represented by the x and y axes. The blue and orange diamonds represent labelled instances from two classes, and the grey circles (A-L) are unlabelled instances.



Describe how you would construct a query-by-committee active learning system for this problem, and describe its behaviour on this dataset. Which instance(s) would be labelled first and why?

Edit View Insert Format Tools Table

20px Paragraph B I U A T²

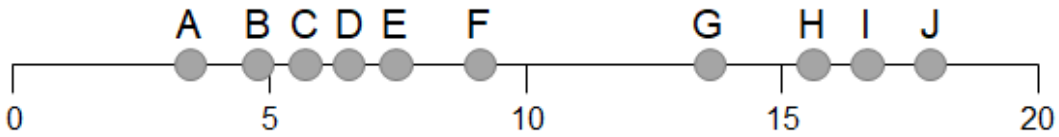
p

0 words

Question 19

5 pts

Suppose you wish to model the probability density function of the dataset shown below using kernel density estimation (KDE). The dataset consists of 10 instances (labelled A-J) with a single attribute shown on the x axis.



Explain how the peak(s) of the probability density function would change as you increase the kernel bandwidth from a very low value (e.g., 0.5) to a very high value (e.g., 50). Do not report formulas or try to compute exact values -- just explain approximately where the peaks would be located for different bandwidths, using the figure. (3-4 sentences)

p 0 words </>

Question 20 25 pts

Design and Application

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 to respond in a full paragraph for each of the questions below. These questions will require significantly more thought than those in the previous portion of the exam and it is recommended to attempt this section only after completing the earlier sections.

Predicting movie viewers' opinions

A popular movie streaming website wants to learn more about its customers' preferences so it can develop targeted advertisements and make recommendations about product placement in upcoming films. The streaming website has partnered with a company (Brand X) and have asked you to help them develop a system that can predict people's opinions about Brand X products based on their movie watching habits.

Last year, the movie streaming website conducted a poll of 10,000 users to measure their opinions of Brand X, which is measured on a 5-point scale (1 = strongly negative opinion, 5 = strongly positive opinion). The streaming website will provide you a dataset consisting of the movie viewing history for each of these users (e.g., a list of which movies these users have streamed on the website, with dates) and their opinion of Brand X (a value from 1-5).

The streaming website will also give you access to their internal database, so you'll have access to a list of all the movies in their

streaming catalogue, some basic information about each movie (e.g., genre, language, date of release, dates available for streaming, box office, etc.), and the movie viewing history of all current users. This dataset is constantly updated, since users stream movies on the website every day, and new movies are added to the website every week. Note that this means that many of the movies currently on the website were added after the user poll was conducted, so viewers who participated in the poll could not have streamed these movies. Unfortunately, the users who participated in the poll are not included in the "live" dataset; you have no way to find out what movies these users streamed after the poll was conducted, and you have no ground truth labels for any users in the "live" dataset.

The objective is to build a system that predicts users' opinion of Brand X based on their movie viewing history. To ensure the system stays-up-to-date with new movie releases, the streaming website would like your algorithm to be able to make predictions based on movies released in the last year, in addition to making predictions based on movies released before the poll.

- What machine learning system would you use? Why? Explain any assumptions you are making.
- How would you train and evaluate this system?
- Suppose you could get labels for a portion of the "live" dataset, by doing another poll of 10,000 users. How would you select users for this poll and why?

Edit View Insert Format Tools Table

20px Paragraph

p

0 words

Not saved

Submit Quiz