Exam: Machine Learning (COMP30027_2021_SM1)

Started: Jun 9 at 10:00

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 22 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

Given a machine learning task, which regularly adds new observations with labels to dataset, which of the following two models is preferred: decision tree or k-nearest neighbours? Briefly explain your choice. (2-3 sentences).

Edit View Insert Format Tools Table

12pt Paragraph

p

0 words

Question 2

3 pts

For linear regression, why do we need regularisers? When do we prefer L-1 norm to L-2 norm as a regulariser? (2-3 sentences)

Edit View Insert Format Tools Table

12pt Paragraph

p 0 words

Question 3

3 pts

There are different strategies to search for optimal solutions. Explain one benefit of using gradient descent instead of an analytic solution for linear regression. (1-2 sentences)

Edit View Insert Format Tools Table

12pt Paragraph

p

0 words

Question 4

2 pts

Which of the following statements are TRUE about SVM?

- [ ] Support vectors are the training instances which the model can classify with higher confidence than other instances.
- [ ] Support vectors define the hyperplane that is used to classify new instances.
- [ ] The minimum number of support vectors is two.
- [ ] Transforming data into a higher-dimensional space can increase the generalisability of the model.

Question 5

4 pts

Briefly describe how the forward algorithm reduces computation time for the evaluation task of HMM compared to the brute-force method. (2-3 sentences)

Edit View Insert Format Tools Table

12pt Paragraph

p 0 words

Question 6

2 pts

Suppose a neural network takes as input a 1-dimensional vector of size 10. Its first hidden layer consists of 8 convolutional neurons with kernel size 3. How many parameters does this first hidden layer learn?

Question 7

4 pts

Backpropagation is a form of stochastic gradient descent. What makes the training process stochastic, and what is an advantage of this approach (compared to non-stochastic gradient descent)? (2-3 sentences)

Edit View Insert Format Tools Table

12pt Paragraph

p 0 words

Question 8

3 pts

Given a dataset with N instances, what is the expected impact on model variance, if you increase k in k-fold cross validation? Briefly explain why. (1-2 sentences).

Edit View Insert Format Tools Table

12pt Paragraph

p 0 words

Naïve Bayes

The follow questions relate to probability and naïve Bayes.

Question 9

7 pts

A researcher has collected the following dataset on two species of birds (the "?" symbol indicates missing data):

ID	Head	Wing Pattern	Back	Species
1	Red	Spots	Brown	A
2	Orange	Bars	Gray	A
3	Orange	Bars	?	A
4	Orange	Solid	Brown	A
5	Red	Spots	?	A
6	Red	Spots	Brown	B
7	Orange	Spots	Brown	B
8	Orange	?	Brown	B
9	Red	Bars	Gray	B

Use this data to train a naïve Bayes classifier to distinguish the two species (do not use smoothing). Give the classification result for a new instance with the attributes:

Head == Red, Wing Pattern == Spots, Back == Gray

Report the posteriors that the classifier computes for each class (and how they are obtained) to support your answer.

Edit View Insert Format Tools Table

12pt Paragraph B I U A T^2

p 0 words </>

Question 10

4 pts

Suppose you observe a bird with Head == Red, but you are unable to observe its Wing Pattern or Back colour. Which label would your classifier predict for this instance? Report the posteriors that the classifier computes for each class to support your answer.

Edit View Insert Format Tools Table

12pt Paragraph B I U A T²

p 0 words

Question 11

7 pts

Suppose you make additional observations of birds in the wild and observe the following features. Note that you do not know the species label for any of these new instances:

ID	Head	Wing Pattern	Back	Species
10	Red	Bars	Gray	?
11	Orange	Bars	Gray	?
12	Orange	Spots	Brown	?
13	Red	Spots	Brown	?

How could you incorporate this information into your Naïve Bayes model? You do not need to actually recompute your model parameters, but you should explain the steps you would follow to do so. What advantage can be gained by including unlabelled

data in the model?

Edit View Insert Format Tools Table

12pt Paragraph

p 0 words

Decision Trees

The following questions are related to decision tree algorithms and boosting.

Question 12

6 pts

Given a dataset as shown in the following table,

Instance ID	A1	A2	A3	Class
1	T	High	F	Y
2	T	High	F	Y
3	F	Medium	F	N
4	F	Medium	F	N
5	F	Medium	T	Y
6	T	Medium	T	Y
7	T	Low	T	Y
8	T	Low	T	N

What is the information gain of A1? Show your work.

Edit View Insert Format Tools Table

12pt Paragraph

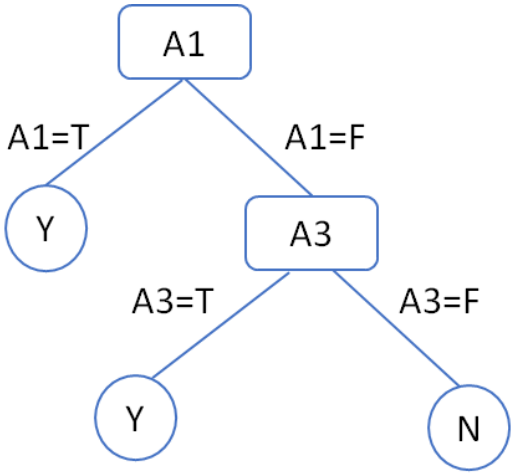
p

0 words

Question 13

8 pts

Assume we are using AdaBoost, with decision tree as the base model. In the first iteration, the dataset sampled will be the set of instances from original training set with ID [1,2,2,3,4,5,6,7]. The classifier for the first iteration C_1 is shown in the following figure:



What are the instance weights for the next iteration? Show your calculation.

Edit View Insert Format Tools Table

12pt Paragraph

p 0 words

Question 14

3 pts

Briefly explain why AdaBoost updates instance weights. (1-2 sentences).

Edit View Insert Format Tools Table

12pt Paragraph

p 0 words

Numeric classification

The following questions are related to numeric classification methods.

Question 15

8 pts

Given a dataset as shown in the following table,

ID	A1	A2	A3	Class
1	2.5	1.5	Low	Y
2	3	1	Medium	Y
3	1	0.5	Medium	N
4	4.5	2	High	Y
5	2	3	Low	N

Test instance: A1 = 2, A2 = 1.5, A3 = Low

If 3-NN is used, what is the prediction for the test instance? What are the instance IDs of the 3 nearest neighbours? Show the distance measure you select and your calculation.

Edit View Insert Format Tools Table

12pt Paragraph

p 0 words

Question 16

6 pts

Suppose you train a logistic regression classifier on the dataset above. How would the parameters $[\beta_0,\ \beta_1, \beta_2, \beta_3]$be updated in one step of the gradient ascent? Assume the learning rate is 0.1, the current parameters are $[\beta_0,\ \beta_1, \beta_2, \beta_3] = [0.1, 0.2, 0.2, -0.1]$ and [Low, Medium, High] is mapped to [1,2,3] for A3.

Note: $\frac{\partial L(\beta)}{\partial \beta_j} = \sum_i (y_i - h_\beta(x_i))x_{ji}$

You should provide a clear explanation of the steps, but you do not need to actually compute the values.

Edit View Insert Format Tools Table

12pt Paragraph

p 0 words

Question 17

5 pts

If logistic regression is used, assuming the final optimisation solution is $[\beta_0, \beta_1, \beta_2, \beta_3] = [0.05, 0.3, 0.1, -0.2]$, and [Low, Medium, High] is mapped to [1,2,3] for A3, what is the prediction result for the test instance? Show your calculation.

Edit View Insert Format Tools Table

12pt Paragraph

p

0 words

Unsupervised learning

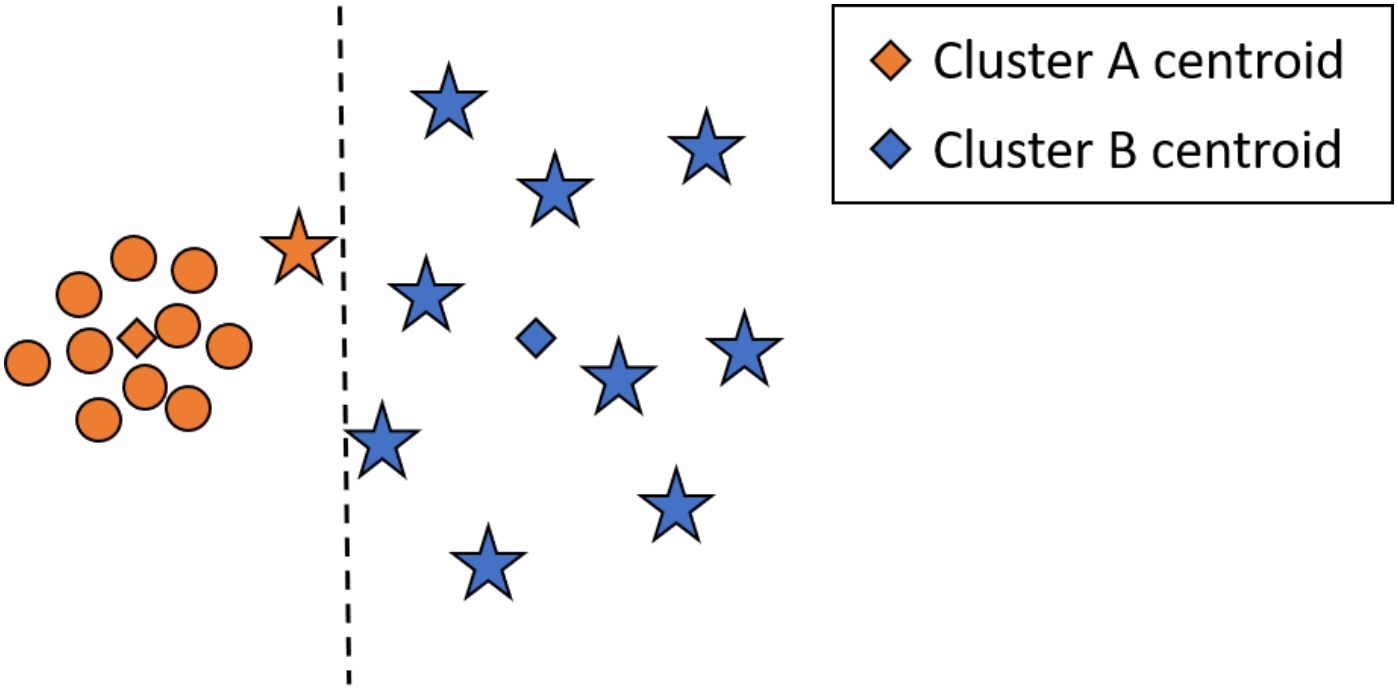
The following questions relate to unsupervised learning algorithms.

Question 18

4 pts

Suppose you run a k-means clustering algorithm (k=2) on a labelled dataset and obtain the clustering result shown below. In this figure, the colour indicates cluster assignment (A or B) and the shape of the data point indicates its ground truth label (circle or star).

(To save you some counting: there are 20 points in this data set, 10 from the ground-truth "star" class and 10 from the ground-truth "circle" class.)



What is the purity and entropy of cluster A? What is the purity and entropy of cluster B?

Edit View Insert Format Tools Table

12pt Paragraph

p

0 words

Question 19

5 pts

Suppose that instead of using k-means, we fit a GMM to represent this data as a mixture of two Gaussians, and further suppose that the GMM identifies the same cluster centroids as the k-means algorithm shown above. Since GMM is a probabilistic clustering algorithm, we obtain labels for each point by assigning it to its highest-probability cluster. Would you expect the cluster labels for the individual instances from this GMM to differ from the labels from k-means shown above? Why or why not? (2-3 sentences)

Edit View Insert Format Tools Table

12pt Paragraph

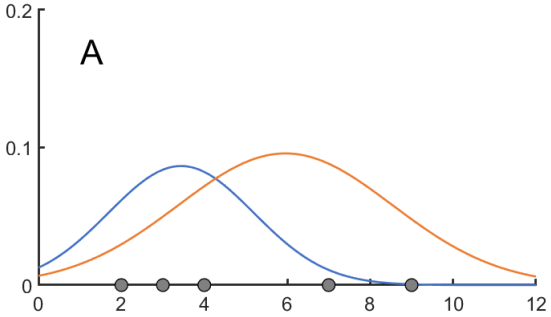
p

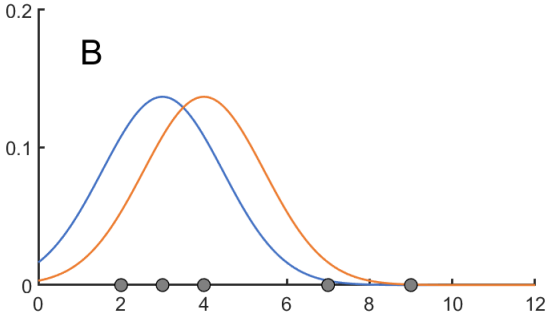
0 words

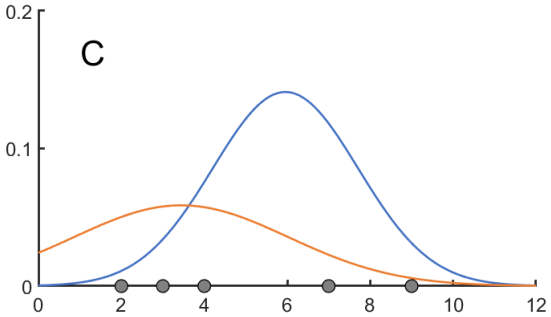
Question 20

3 pts

Two of the figures below illustrate an iteration of the EM algorithm as it fits a Gaussian mixture model to a dataset. In each figure, the gray dots are data points and the blue and orange curves show the probability distribution of the two Gaussians in the mixture model. One figure shows the initial state of the model, and one figure shows the updated state after one Expectation-Maximization step. The third figure is a decoy that doesn't belong.







Which figure is the initial state of the model and which figure shows the updated model after one EM step?

Initial state of the model: [Choose]

Updated model after one EM step: [Choose]

Question 21

5 pts

Justify your answer to the previous question by explaining how the model is updated on this EM step, making specific reference to the figures you have chosen.

Edit View Insert Format Tools Table

12pt Paragraph

p

0 words

Question 22

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.

Which pop songs will be hits?

You have been contracted by a music industry start-up to develop an algorithm that can predict whether a pop song will be a hit. The start-up will provide you with a dataset created from songs released in the U.S. over a 50-year period (1960-2010). Each instance (song) in the dataset includes a large variety of details about the song (e.g., the genre, tempo, duration, artist name, key words from the lyrics, etc.), and a binary label ("hit" or "non-hit"). The "hit" label is based on record sales; a song which ranks in the top 40 best-selling songs in the U.S. for at least one week is labelled as a "hit" and songs which do not achieve this are "non-hits." Your goal is to develop an algorithm that will be able to predict the "hit" label for songs released in the future.

The goal of the start-up is to not only identify which new songs are most likely to be hits, but also to understand what makes a song a hit. In addition to accurately predicting which songs will be hits, the start-up wants to be able to explain the algorithm's decisions to its clients (i.e., explain why a particular song was labelled as a "hit" or "non-hit" according to your model).

a.) [10 marks] Describe the algorithm that you would use to do this classification task. Be specific about your implementation, including any data preprocessing steps, the model structure, and how you would construct train/test/validation sets. What do you consider to be the main advantages and disadvantages of your chosen approach?

b.) [10 marks] Describe the method(s) you would use to interpret the model's decisions for clients. What kind of explanations will your model be able to provide? State any assumptions or limitations in your approach.

c.) [5 marks] The start-up claims that an important advantage of using an algorithm to evaluate songs is that it won't judge music the way humans do, so it can provide a neutral, objective assessment of a song without any bias. For example, the algorithm will

be able to discover new hit songs from genres or styles that haven't traditionally been hits in U.S. Do you agree with this view? Why or why not?

Edit View Insert Format Tools Table

12pt Paragraph

p 0 words

Not saved Submit Quiz