Coversheet for centrally scheduled online non-invigilated exams

Exam information	
Course code and name	COMP4702 / COMP7703 / DATA7703 Machine Learning / Machine Learning for Data Scientists
Semester	Semester 1, 2020
Exam type	Online, non-invigilated
Exam date and time	Please refer to your personalised timetable
Exam duration	• Working time (180 minutes) + additional online allowance: 30 minutes = TOTAL exam duration: 3 hrs 30 minutes within a 24 hour window
Exam window	• You have a 24 hour window in which you must complete your exam. Once you open your test you will have the duration of the exam to complete and submit/upload your response. The latest possible starting time is 180 minutes prior to the end of this 24 hours.
Late submission of exams	If you have a technical problem, you should collect evidence (photo/screenshot if possible) and make a statement about the incident. You should submit as soon as the problem is alleviated. Should the problem extend beyond 30 minutes, you should gather evidence at regular intervals (e.g. a screenshot every 1 hour). If you do not submit the exam on time, or within the approved extended time, you will receive '0' mark, unless there is sufficient evidence that the late submission is beyond your control. Without a reason preventing submission there is a 100% penalty. Course coordinators decide on the evidence presented whether they will accept a late submission. • You can appeal to the Associate Dean (Academic).
Reading time	Reading time has not been formally allocated for online exams, however students are encouraged to review and plan their approach for the exam before they start. The total exam time should be sufficient to do this.
Additional time	30 minutes additional time has been incorporated in recognition of the online environment and the different circumstances that students face in their home environments. This includes allowances for network or connection issues.
Weighting	COMP4702 - This exam is weighted at 45% of your total mark for this course. COMP7703 - This exam is weighted at 40% of your total mark for this course. DATA7703 - This exam is weighted at 45% of your total mark for this course.
Permitted materials	This is an open book exam – all materials permitted.

Required/recommended materials	Laptop or desktop computer with software used during the course. Scientific calculator or calculator app. Paper and writing implements for rough working. If the student chooses to complete hand-written answers for some of the exam, scanning technology (e.g. mobile phone, camera, scanner) will be required to include this material in a pdf submission.
Instructions	Answer all questions. Part A is to be completed in the test tool. Part B is to be completed in the test tool. Part C is to be completed in a document, and uploaded to the Assignment link for the exam.
Who to contact	*Given students may not all undertake the online exam at the same time, or in the same time zone, and that some questions may be randomised, responding to student queries and/or relaying corrections to exam content during the exam will not be feasible.* If you need to make assumptions or have any queries or concerns about a particular exam question, state these at the start of your solution to that question. You may also include queries you may have made with respect to a particular question, should you have been able to 'raise your hand' in an examination room. If you experience any **technical difficulties** during the exam, contact the Library AskUs service for advice. You should also ask for an email documenting the advice provided so you can provide this to the course coordinator immediately at *marcusq@itee.uq.edu.au*.
Important exam condition information	This is an open book exam. You will have access to your own notes, course texts, and other materials. The normal academic integrity rules apply. • You cannot cut-and-paste material other than your own work as answers. • You are not permitted to consult any other person – whether directly, online, or through any other means – about any aspect of this assessment during the period that this assessment is available. If it is found that you have given or sought outside assistance with this assessment then that will be deemed to be cheating and will result in disciplinary action. By undertaking this online assessment you will be deemed to have acknowledged UQ's academic integrity pledge to have made the following declaration: *"I certify that my submitted answers are entirely my own work and that I have neither given nor received any unauthorised assistance on this assessment item"*.

Preview Test: COMP4702/7703/DATA7703 Semester One 2020 Final Examination

Test Information

Description	COMP4702/7703/DATA7703 - Machine Learning Semester One 2020 Final Examination
Instructions	Answer all questions. There are 60 marks in total.
Timed Test	This test has a time limit of 3 hours and 30 minutes.You will be notified when time expires, and you may continue or submit. Warnings appear when **half the time**, **5 minutes**, **1 minute**, and **30 seconds** remain.*[The timer does not appear when previewing this test]*
Multiple Attempts	Not allowed. This test can only be taken once.
Force Completion	This test can be saved and resumed later. The timer will continue to run if you leave the test.

QUESTION 1 **2 points** Save Answer

Regarding the k-nearest neighbour (kNN) classifier, which of the following statements is **incorrect**?

- ○ For even values of k, a majority vote can be used directly to determine the final class prediction.
- ○ kNN is a non-parametric model.
- ○ kNN is sometimes called a “lazy learning” method because no training is required.
- ○ The choice of a distance metric is an important factor in applying kNN.

QUESTION 2 **2 points** Save Answer

When Principal Component Analysis (PCA) is performed by calculating the eigenvectors and eigenvalues of the covariance matrix of a dataset, which of the following statements is true?

- ○ The principal components are always orthogonal to each other in the transformed feature space.
- ○ The eigenvalues always sum to equal 1.
- ○ By retaining more principal components the amount of variance explained in the projection is strictly increases.
- ○ The area under the scree graph is proportional to the computational complexity of solving the eigensystem.

QUESTION 3 **2 points** Save Answer

Which of the following is not a hyperparameter or experimental factor when training neural networks?

- ○ Prior probability coefficients.
- ○ Batch size.
- ○ Learning rate.
- ○ Weight initialization range/distribution.

QUESTION 4 **2 points** Save Answer

If a supervised learning model with high complexity (e.g. the number of parameters in a polynomial regression model) is trained on a relatively small dataset (i.e. number of data points) the it is most likely to have the following characteristics:

- ○ Low bias, high variance, low training error, high test error.
- ○ High bias, low variance, high training error, low test error.
- ○ Low bias, high variance, high training error, low test error.
- ○ High bias, high variance, high training error, low test error.

QUESTION 5 — **2 points** Save Answer

Match the following key terms in k-means clustering with their closest corresponding term in fitting a Gaussian mixture model using the EM algorithm:

- - Mean vector.
- - Data point ownership.
- - Local optimization.
- - Covariance matrix.

A. Cluster center.
B. Responsibility value.
C. Expectation maximization.
D. Distance metric.

QUESTION 6 — **2 points** Save Answer

Support Vector Machines:

- ○ Use slack variables to penalize data points which fall inside or on the incorrect side of the margin.
- ○ Are trained using stochastic gradient descent.
- ○ Solve classification problems by minimizing the margin between the discriminant and the training points.
- ○ Use a kernel function to determine the likelihood of each datapoint in the training set.

QUESTION 7 — **2 points** Save Answer

A Bayesian network is designed which implements the following probability distribution for a given problem:
$P(A,B,C,D,E,F,G) = P(A)P(B)P(C|A)P(D|A,B)P(E|D)P(F|C)P(G|E,F)$

The number of edges in the graph representation of the network is:

QUESTION 8 — **2 points** Save Answer

Calculate the overall error rate (as a percentage, to 2 decimal places) given by the following confusion matrix, where the entry in the j-th colun indicates the number of data points predicted to belong to class i that actually belong to class j:

$$\begin{pmatrix} 47 & 0 & 0 \\ 7 & 38 & 2 \\ 0 & 1 & 46 \end{pmatrix}$$

QUESTION 9 — **2 points** Save Answer

Which of the following equations could describe the k-th output of a multilayer perceptron with 64 inputs and 128 hidden units?

- ○ $y_k(\mathbf{x},\mathbf{w}) = \sigma\left(\sum_{j=1}^{129} w_{kj} h\left(\sum_{i=1}^{65} w_{ji} x_i + w_{j0}\right) + w_{k0}\right)$
- ○ $y_k(\mathbf{x},\mathbf{w}) = \sigma\left(\sum_{j=1}^{65} w_{kj} h\left(\sum_{i=1}^{129} w_{ji} x_i + w_{j0}\right) + w_{k0}\right)$
- ○ $y_k(\mathbf{x},\mathbf{w}) = \sigma\left(\sum_{j=1}^{64} w_{kj} h\left(\sum_{i=1}^{128} w_{ji} x_i + w_{j0}\right) + w_{k0}\right)$
- ○ $y_k(\mathbf{x},\mathbf{w}) = \sigma\left(\sum_{j=1}^{128} w_{kj} h\left(\sum_{i=1}^{64} w_{ji} x_i + w_{j0}\right) + w_{k0}\right)$

QUESTION 10

2 points Save Answer

Given the following matrices:

$$H = \begin{pmatrix} 3 & 1 & 4 & 3 \\ 2 & 1 & 4 & 3 \\ 1 & 4 & 1 & 3 \\ 2 & 2 & 3 & 1 \end{pmatrix}, K = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$$

where H is the 4x4 input data (e.g. pixel values) and K is a 2x2 kernel, perform a convolution operation (as it is done in convolutional neural networks) on H using K, with a stride of 2 and padding of 1 all around H.

- ☐ $\begin{pmatrix} 2 & 3 \\ -1 & -3 \end{pmatrix}$
- ☐ $\begin{pmatrix} 6 & -1 & 7 & 2 & -3 \\ 1 & 2 & 4 & 3 & 0 \\ 0 & 8 & -5 & 6 & 0 \\ 3 & -1 & 7 & -3 & 2 \\ -2 & 0 & -1 & 2 & 1 \end{pmatrix}$
- ☐ $\begin{pmatrix} 6 & 7 & -3 \\ 0 & -5 & 0 \\ -2 & 0 & 1 \end{pmatrix}$
- ☐ $\begin{pmatrix} 6 & 7 & -3 \\ 0 & -5 & 0 \\ -2 & -1 & 1 \end{pmatrix}$

QUESTION 11

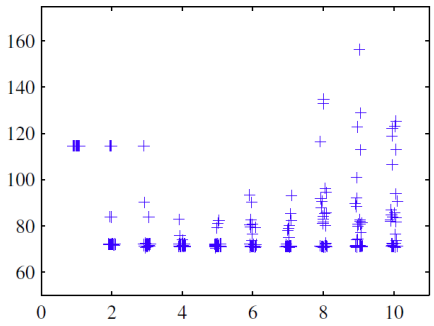
5 points Save Answer

In the textbook by Chris Bishop, the following Figure is presented:

a. What do you think this graph is showing about the training of a multi-layer perceptron? Discuss in a few sentences. (3 marks)

b. Bishop states that the lowest validation error obtained over all the training runs was for 8 hidden units (he says "validation" in the text but I assume he means the test error values in this graph). Based on the results shown, what value would you choose for the number of hidden units and why? (2 marks)

c. **Figure 5.10** Plot of the sum-of-squares test-set error for the polynomial data set versus the number of hidden units in the network, with 30 random starts for each network size, showing the effect of local minima. For each new start, the weight vector was initialized by sampling from an isotropic Gaussian distribution having a mean of zero and a variance of 10.



QUESTION 12

5 points Save Answer

The Figure below is taken from the Hastie et al. textbook. The Bayes error rate here is the lowest possible error that could be achieved for this particular example.

While this Figure presents a specific example, an important, general point can be drawn from it regarding the setting of hyperparameters in Machine Learning models. Discuss in your own words what this general point is by referring to the Figure. (5 marks)

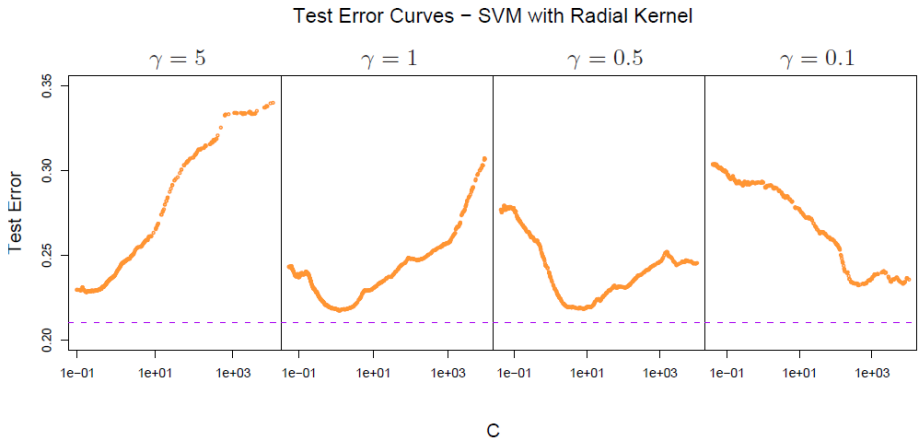
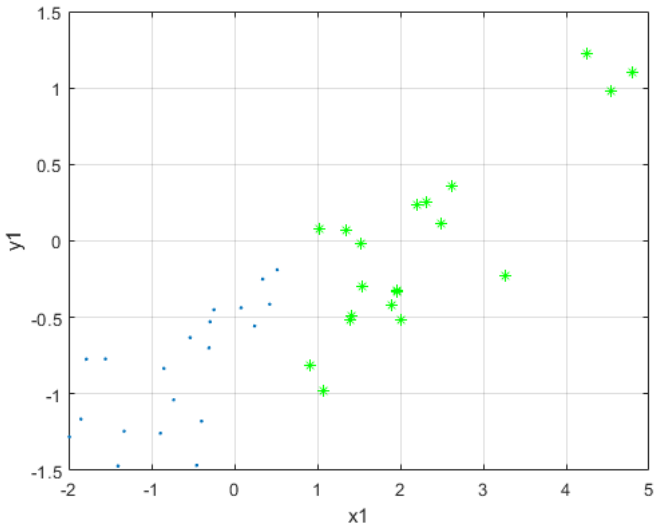


FIGURE 12.6. *Test-error curves as a function of the cost parameter C for the radial-kernel SVM classifier on the mixture data. At the top of each plot is the scale parameter γ for the radial kernel: $K_\gamma(x, y) = \exp -\gamma||x - y||^2$. The optimal value for C depends quite strongly on the scale of the kernel. The Bayes error rate is indicated by the broken horizontal lines.*

QUESTION 13

5 points Save Answer

The following Figure shows a set of data for a classification problem:



Consider using rectangles as a model to classify this data, where a rectangle is specified by [a<=x1<=b; c<=x2<=d]. A rectangle predicts that a point belongs to its designated class if the point lies within or on the perimeter of the rectangle.

a. Specify a set of model parameter values that will perfectly classify all of the data shown for both classes. (2 marks)

b. Give an example of a test data point (feature values and class label) that would be incorrectly classified by the models you have specified. (1 mark)

c. Consider performing logistic regression on the data from Figure 1. How many parameters would the trained linear regression model have? (2 marks)

QUESTION 14

5 points Save Answer

The Figure below is from the textbook by Chris Bishop:

a. Which plot do you think is most representative of overfitting? Explain your choice. (3 marks)

b. Referring to the two green points towards the top of the plots, explain how the K-nearest-neighbour algorithm can be said to be robust to noise for large values of K. (2 marks)

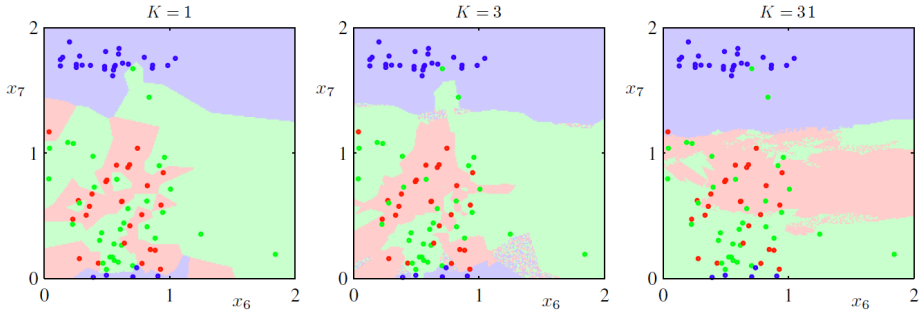


Figure 2.28 Plot of 200 data points from the oil data set showing values of x_6 plotted against x_7, where the red, green, and blue points correspond to the 'laminar', 'annular', and 'homogeneous' classes, respectively. Also shown are the classifications of the input space given by the K-nearest-neighbour algorithm for various values of K.

QUESTION 15

20 points Save Answer

In the Practical classess for this course, you have implemented and/or applied a range of different machine learning models and algorithms. For this question, you are required to demonstrate your ability to apply one or more of these techniques to a supplied dataset.

Key points:

- You are expected to spend about 60 minutes on this question.
- You can use any code that you developed during the pracs, as well as built-in Matlab and python functions/libraries. If you use other libraries then you must reference them and they must be publicly available.
- Explain key steps of your analysis (data preprocessing, training, testing, visualization, etc.).
- Present any results and output from your work. Add comments and discussion to demonstrate that you understand the results and output.
- State any assumptions you make.
- Don't spend time introducing background concepts or describing the theory of models from the course.
- Make sure your work is understandable and readable but don't spend a lot of time making the presentation neat and polished.

Your answer for this question should be uploaded as a pdf file in the turnitin submission for the exam.