UNIVERSITY
OF WOLLONGONG
AUSTRALIA

School of Computing and Information Technology

Student to complete:

Family name	
Other names	
Student number	

CSCI933

Machine Learning: Algorithms and Applications

South West Sydney and Wollongong

Examination Paper

Autumn 2022

Exam duration	3 hours
Weighting	45%
Items permitted by examiner	OPEN BOOK - any reference materials permitted
Aids supplied	Nil.
Directions to students	Four (4) questions are to be answered.
	Marks allocated to each question are as indicated in the paper.
	Any act of copying from lecture notes or world wide web will incur a zero mark. All answers must be in your own words. Please be warned.
	Answer each question neatly and legibly. ONLY PDF documents can be submitted.
	You may use **legible handwriting** to present any answer requiring mathematical equations and attach it as a **neat and readable** scanned page in your PDF submission.
	Submit your completed response/answers to the examination questions in the Submission Box provided on Moodle.

Question 1 (10 marks)

Imagine you are the data scientist in a top car insurance company. Your immediate supervisor would like you to lead a new project in which the premium payable by customers will be set based on a number of variables. These variables are known to determine the risk carried by a given customer. In a previous project meeting with your supervisor it was determined that a regression model will be most appropriate for this project.

You are required to write a one-page report for your supervisor discussing the issues that need to be considered in building this model. This report must clearly show that you understand the necessary fundamental concepts in machine learning and how they inform the design of your predictor model.

Your report should contain considerations for machine learning concepts including *model complexity, number of variables, model capacity, under-fitting, over-fitting, data scaling, missing data, number of example data, PAC learning, training error, cross validation,* and *generalization error.* In essence your report should explore theoretical aspects of machine learning that will guide your design decisions and successful execution of the project. **The report must not be a list of points. It must be written as a well thought-out report and must not exceed one page**.

Question 2   10 marks

Assume that the following 10 features vectors originate from two classes ω_1 and ω_2. You can assume that the two features are denoted by x_1 and x_2.

$$\omega_1 : [0.1 \quad -0.2]^t, [0.2 \quad 0.1]^t, [-0.15 \quad 0.2]^t, [1.1 \quad 0.8]^t, [1.2 \quad 1.1]^t$$
$$\omega_2 : [1.1 \quad -0.1]^t, [1.25 \quad 0.15]^t, [0.9 \quad 0.1]^t, [0.1 \quad 0.2]^t, [0.2 \quad 0.9]^t$$

(a)   [5 marks] Check whether these two classes are linearly separable. If they are not, write down the two linear discriminant functions $g_1(x_1, x_2)$ and $g_2(x_1, x_2)$ that when combined will separate the two classes. [Hint: plot the vectors in 2-dimensional space, and write equation of two staright lines]

(b)   [5 marks] Assume that the function required to combine $g_1(x_1, x_2)$ and $g_2(x_1, x_2)$ in part a is given by

$$f(g_1, g_2) = g_1 - 2g_2 - \frac{1}{2}$$

design a two-layer perceptron that nodes having step function activation function to classify the vectors into the two classes. You **must** draw the two-layer perceptron with appropriate wieghts and bias. [Hint: think about how the signal flows in a perceptron/fully connected neural network]

Question 3 (13 marks)

Four measurements are made on each of a random sample of 500 animals. The first three variables were different linear dimensions, measured in centimetres, while the fourth variable was the weight of the animal measured in grams. The sample covariance was calculated and its four eigenvalues found to be $14.1, 4.3, 1.2,$ and 0.4. The eigenvectors corresponding to the first and second eigenvalues were:

$$\boldsymbol{u}_1^t = \begin{bmatrix} 0.39 & 0.42 & 0.44 & 0.69 \end{bmatrix}$$
$$\boldsymbol{u}_2^t = \begin{bmatrix} 0.40 & 0.39 & 0.42 & -0.72 \end{bmatrix}$$

where t denotes transpose.

(a)   [3 marks]

Comment on the use of the sample covariance matrix for the principal components analysis for these data.

(b)   [3 marks]

What is percentage of the variance in the original data accounted for by the first two principal components? Describe the results.

(c) [3 marks] The undercomplete autoencoder can be used to achieve feature extraction and dimensionality reduction. What is the relationship between the autoencoder and the principal component analysis? Draw an autoencoder that can accomplish similar task as the principal component analysis described in this question. Indicate the number of nodes at each layer of your autoencoder.

(d) [4 marks] The following equation represents an infinite impulse response filter:

$$y[n] = y[n-1] + y[n-2] + \sum_{n=0}^{N} x[n] + C$$

where $x[n]$ is the input sample and $y[n]$ is the output sample at time instance n; C is a constant. The notation $y[n-1]$ indicates an output delayed by one time unit. It can be thought of as a memory.

   (i) By considering the equation of the filter given above, explain how it achieves the memory using the following input, $x[n] = [0.5, 0.4, 0.7, 0.5, 0.6, 0.4, 0.8, 0.5], n = 0, ...7, C = 0.1, N = 2, y[-1] = 0.0, y[-2] = 0.0$. [Hint: calculating $y[n]$ up to $n = 5$ should be sufficient.]

   (ii) In what way does this filter function similarly to an LSTM cell? You need to explain how LSTM achieves "memory".

Question 4 (15 marks)

Answer the following questions using appropriate equations and diagrams where they help the examiner understand your response better. Note that you are not being asked to solve the optimization problems.

(a) [6 marks]

Regression is a very useful machine learning algorithm and has various formulations in which additional constraints are added to the basic formulation. Ridge regression and LASSO (Least Absolute Shrinkage and Selection Operator) are two well known forms and the equations describing the associated optimization problems are given below.

$$\min_{\boldsymbol{W}} F(\boldsymbol{W}) = \lambda||\boldsymbol{W}||^2 + ||\boldsymbol{X}^t\boldsymbol{W} - \boldsymbol{Y}||^2 \tag{1}$$

Ridge regression

$$\min_{\boldsymbol{W}} F(\boldsymbol{W}) = \beta||\boldsymbol{W}||_1 + ||\boldsymbol{X}^t\boldsymbol{W} - \boldsymbol{Y}||^2 \tag{2}$$

LASSO regression

where λ and β are positive parameters that determine the trade-off between the respective the regularization terms, $||\boldsymbol{W}||^2$, $||\boldsymbol{W}||_1$, and the empirical mean squared error; $\boldsymbol{X} \in \mathbb{R}^{N \times m}$ is the matrix of feature vectors; $\boldsymbol{Y}$ is the vector of associated predictors; $\boldsymbol{W}$ is the vector of weights; $()^t$ indicates transpose in these equations.

Discuss the similarities and differences between the two formulations (Equations 1 and 2) with respect to the properties of the norms used in the regularization and the data fidelity terms, and the nature of the solution weight vector $\boldsymbol{W}$. How do these two forms of regression affect the feature vectors in their contribution to the prediction? Use appropriate diagrams to illustrate your answer.

(b) [4 marks]

Explain how kernelization affects the conventional PCA (principal component analysis) and SVM (support vector machine).

(c) [5 marks]

You are required to design of a neural network-based classifier. Explain why you will consider each of the following in your network architecture:

   (i) Convolutional layer

(ii) Pooling

(iii) Dropout layer

(iv) Layer normalization

(v) Skip connection

Your explanation of each item must not be more than $1/3$ of a page and should illustrate the depth of your understanding.

-END OF EXAMINATION PAPER-