UNIVERSITY OF WARWICK	
Department	Computer Science
Module Code	CS909
Module Title	Data Mining
Exam Paper Code	CS9090_A
Duration	2 Hours
Exam Paper Type	24-hour window

STUDENT INSTRUCTIONS

1. Read all instructions carefully. We recommend you read through the entire paper at least once before writing.

2. There are **6** questions. All candidates should **attempt 4 questions in total with at least ONE question from Section B.**

3. You should not submit answers to more than the required number of questions.

4. You should handwrite your answers either with paper and pen or using an electronic device with a stylus (unless you have special arrangements for exams which allow the use of a computer). Start each question on a new page and clearly mark each page with the page number, your student id and the question number.
Handwritten notes must be scanned or photographed and all individual solutions collated into a single PDF with pages in the correct order.

5. Please ensure that all your handwritten answers are written legibly, preferably in dark blue or black ink. If you use a pencil ensure that it is not too faint to be captured by a scan or photograph.

6. Please check for legibility before uploading. It is your responsibility to ensure your work can be read.

7. Add your student number to all uploaded files.

8. You are permitted to access module materials, notes, resources, references and the internet during the online assessment.

9. You must not communicate with any other candidate during the assessment period or seek assistance from anyone else in completing your answers. The Computer Science Department expects the conduct of all students taking this assessment to conform to the stated requirements. Measures will be in operation to check for possible misconduct. These will include the use of similarity detection tools and the right to require live interviews with selected students following the assessment.

10. By starting this assessment, you are declaring yourself fit to undertake it. You are expected to make a reasonable attempt at the assessment by answering the questions in the paper.

IMPORTANT INFORMATION

- We strongly recommend you use Google Chrome or Mozilla Firefox to access the Alternative Exams Portal.

- You are granted an additional 45 minutes beyond the stated duration of this assessment to allow for downloading/uploading your assessment, your files and any technical delays.

- Students with approved Alternative Exam Arrangements (Reasonable Adjustments) that permit extra time and/or rest breaks will have this time added on to the stated duration.

- You must have completed and uploaded the assessment before the 24-hour assessment window closes.

- Late submissions are not accepted.

- If you are unable to submit your assessment, you must submit Mitigating Circumstances immediately, attaching supporting evidence and your assessment. The Mitigating Circumstances Panel will consider the case and make a recommendation based on the evidence to the Board of Examiners.

SUPPORT DURING THE ASSESSMENT

Operational Support

- Use the Alternative Exams Portal to **seek advice immediately if during the assessment period:**
    - you cannot access the online assessment
    - you believe you have been given access to the wrong online assessment

  Operational support will be available between 09:00 and 17:00 BST for each examination (excluding Sunday)

Technical Support

- If you experience any technical difficulties with the Alternative Exam Portal please contact helpdesk@warwick.ac.uk

Technical support will be available between 09:00 and 17:00 BST for each examination (excluding Sunday)

Academic Support

- If you have an academic query, contact the invigilator (using the 'Contact an Invigilator' tool in AEP) to raise your issue. Please be aware that two-way communication in AEP is not currently possible

  Academic support will normally be provided for the duration of the examination (i.e. for a 2 hour exam starting at 09:00 BST, academic support will normally be provided between 09:00 and 11:45 BST). Academic support beyond this time is at the discretion of the department.

Other Support

- **if you cannot complete your assessment for the following reasons submit Mitigating Circumstances immediately**:

    - you lose your internet connection
    - your device fails
    - you become unwell and are unable to continue
    - you are affected by circumstances beyond your control

Section A

1. (Linear Discriminants)

a. **Show** whether the following dataset is linearly separable or not. **Write** the classification constraints for positive and negative examples and **show** whether those constraints can be satisfied by a linear classifier or not. **Verify** your reasoning by calculating score of the linear discriminant for each example. **[5] -x1-x2-x3>2.5**

Index	Feature-1	Feature-2	Feature-3	Label
1	-1	-1	-1	+1
2	-1	-1	+1	-1
3	-1	+1	-1	-1
4	-1	+1	+1	-1
5	+1	-1	-1	-1
6	+1	-1	+1	-1
7	+1	+1	-1	-1
8	+1	+1	+1	-1

b. **Show** whether the following dataset is linearly separable or not. **Write** the classification constraints for positive and negative examples and **show** whether those constraints can be satisfied by a linear classifier or not. **Verify** your reasoning by calculating score of the linear discriminant for each example. **[5]**

Index	Feature-1	Feature-2	Feature-3	Label
1	-1	-1	-1	+1
2	-1	-1	+1	-1
3	-1	+1	-1	-1
4	-1	+1	+1	-1
5	+1	-1	-1	-1
6	+1	-1	+1	-1
7	+1	+1	-1	-1
8	+1	+1	+1	+1

c. The following dataset is not linearly separable. Can you write an expression for an additional feature (called feature-3) based on Feature-1 and Feature-2 that would make this dataset linearly separable in the resulting 3-dimensional feature space? Please verify your answer showing that the dataset is indeed linearly separable in the resulting 3-dimensional feature space by writing the set of classification constraints for each example and showing that they can be separated by a linear discriminant. **[5]**

Index	Feature-1	Feature-2	Label
1	-1	-1	+1
2	-1	+1	-1
3	+1	-1	-1
4	+1	+1	+1

d. Write a kernel function and its corresponding feature transformation that would make the data in Question-1 part (c) linearly separable. **[5]**

e. The linear discriminant function with weight $\boldsymbol{w}$ and bias b for an example $\boldsymbol{x}$ can be written as $f(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + b$. How can we change the feature representation of the example and the weight vector so that an explicit bias term is not required? **[5] w+b, x +1**

2. **(SVM and Performance Evaluation)**

a. Write the **optimization problem** of a support vector machine with discriminant function $f(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + b$ and its **weight update equation** for a single example using gradient descent. **[5]**

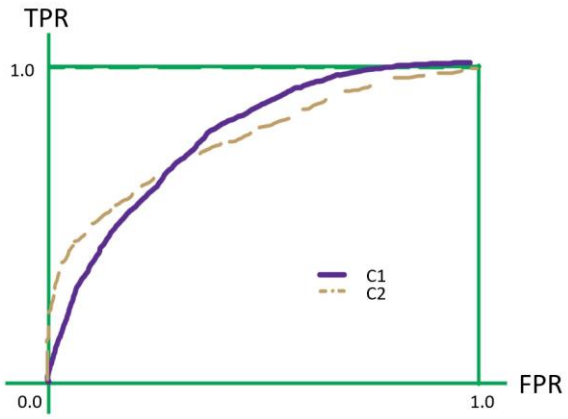
b. **What** is the difference between stratified and non-stratified cross-validation? Consider a dataset with 10,000 examples (9,000 negative, 1,000 positive). **What** is the total number of positive and negative examples in each fold of 10-fold stratified cross-validation? **Why** would you use stratified cross-validation as opposed to non-stratified cross-validation? **[5]**

straitified
保持正负比例

c. You are given a validation dataset for a classification problem with 10,000 examples (9,000 negative, 1,000 positive). **What performance metric** would be ideal to use for this purpose to measure the predictive performance of a binary classifier irrespective of the exact decision value threshold? **Give proper justification** of the exact reason for this choice. **[5]**

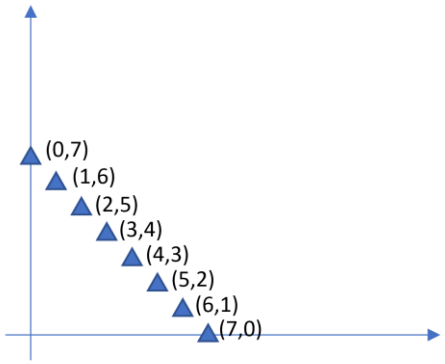
d. For a binary classification problem, all points in the Receiver Operating Characteristic Curve (ROC Curve) of a classifier C1 lie either on or above the corresponding points in the ROC curve of a classifier C2. **Which** of these two classifiers will have a larger Area Under the ROC Curve (AUROC)? **Would** all points in the Precision-Recall (PR) Curve of Classifier C1 lie on or above the corresponding points of the PR curve of classifier C2 or not? **Justify** your answer. **Which** of these two classifiers will have a larger Area Under the Precision-Recall curve (AUC-PR)? **[5]**

e. Consider the ROC curves given in the plot below. Assuming both classifiers (C1 and C2) in the plot have similar area under their ROC curves, which of these classifiers (C1 or C2) is a better classifier for an application in which the number of negative examples is much larger than the number of positive examples? **Why**? **[5]**

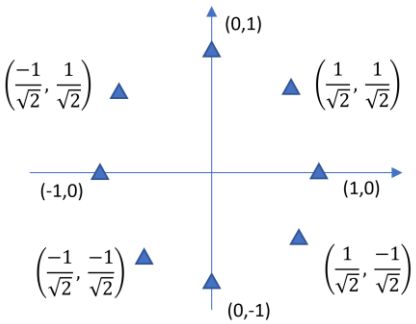


3. **(Dimensionality Reduction and SRM)**

a. **Write** the first principal component as a unit vector for the data points given below in the plot. **Project the data** onto the first principal component and **write the corresponding projection values** for each point. **What** is the percentage of variance captured along the first principal component for this data? **Would** it be possible to reconstruct the 2-D data points from their projection onto the first principal component without any loss? **Give reasoning** for all answers. **[5]**



b. **Write the unit vector corresponding to the direction of maximum variance for the dataset in the plot below. What** is the percentage of variance captured along the first principal component for this data? **Would** it be possible to reconstruct the 2-D data points from their projection onto the first principal component without any loss? **Give reasoning** for all answers. **[5]**



c. Which of the following two linear discriminants will have a larger margin. **Why**? **[5]**

$$f_1(\boldsymbol{x}) = 0.45x_1 + 0.43x_2 - 1.19$$
$$f_2(\boldsymbol{x}) = 1.79x_1 - 0.77x_2 - 0.23$$

d. Assume that for a given example with a target value of y = +1.0, the prediction score for a learning model is z = 0.1. Write the formulae for each of the following loss functions in terms of z and y and calculate the loss functions for this example: **[10]**

i. Squared Error Loss
ii. Hinge Loss
iii. Perceptron Loss
iv. 0-1 Loss
v. Epsilon insensitive loss with epsilon set to 0.1

4. (Optimization)

a. Given a set of d-dimensional examples $\boldsymbol{x_i} \in \mathbb{R}^{\boldsymbol{d}}, i = 1 \ldots N$ with associated labels y_i, derive the analytical solution for the optimal weight parameters for a linear model $f(\boldsymbol{x}; \boldsymbol{w})$ with the following objective function. **[10]**

$$\min_{\boldsymbol{w}} \lambda \|\boldsymbol{w}\|^2 + \sum_{i=1}^{N} (f(x_i; \boldsymbol{w}) - y_i)^2$$

b. How can you solve the objective function in Question 4(a) using gradient descent? Specifically, calculate the gradient of the objective function with respect to the weight parameters for a single example and write at least one step of the gradient-based weight update equation. **[10]**

c. How does the convexity of the loss function affect convergence of a machine learning model? **[5]**

Section B: Answer at least ONE question from Q5 or Q6.

5. (Neural Networks)

a. What is the difference between ranking, regression and classification? **Write loss functions** for each of these problems and explain their difference in detail. [**5**]

b. What is the expected impact of increasing the number of layers in a fully connected neural network on the norm of the gradients of the first layer in backpropagation? Explain your answer in the context of vanishing and exploding gradients. [**5**]

c. Assume that we have a convolutional neural network with the input and filter shown below. **What** value of 'w' can be chosen so that the convolution of the input and the filter results in a value of 0.0 at the center? **Verify** your answer by providing the result of the aforesaid convolution. [**5**]

Input

2	10	3
7	5	3
1	5	3

Filter

0	1	0
1	w	1
0	1	0

d. **Why** does neural network training operate in a batched fashion? **How** does batch size effect batch normalization in neural network training? [**5**]

e. Discuss the impact of each of the following on convergence rate of neural networks by specifying what choices are expected to increase convergence rate and why: [**5**]
   i. Bipolar vs. Binary Input Values
   ii. Bipolar vs. Binary Target Values
   iii. Learning Rate
   iv. Convexity of the loss function
   v. Saturating activation functions vs. non-saturating activation functions

6. **(Application)**

a. **Explain** and apply the following natural language processing steps on the following string as specified in the questions below: **[5]**

"It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of incredulity, it was the season of light, it was the season of darkness, it was the spring of hope, it was the winter of despair."

i. Lemmatization (show output after stemming lemmatization for at least 2 words)
ii. Stop word removal (show stop words)
iii. Term Frequency of each word
iv. Explain why TF-IDF features can be effectively used for comparing this string to other strings

b. Assume that you are given the following dataset from an institution with the goal of prioritizing how closely each individual should be monitored based on historical data of emergency visits. Answer the following questions **with appropriate justification**:

i. How would you **represent** these variables (Gender, Chest Pain, Race, Family History, Exercise Endurance, Age, BMI) as features? Give a justification of your design choice. **[4]**
ii. What is the **target variable** for this problem and **how** would that be used for prioritization of monitoring? **[4]**
iii. Would you model this as a classification problem or a regression problem? **Explain** the **reasoning** behind your choice. **[4]**
iv. What **additional features** do you think can be relevant for this problem? What **features** or **information** should not be used? **[4]**
v. How will you chose a particular machine learning model for this problem? How will you ensure that your predictor generalizes well on data that is not part of your training dataset? **[4]**

Index	Patient-ID	Gender	Chest Pain	Race	Family History	Exercise Endurance	Age	BMI	Emergency Visits
1	X-123	M	No	Asian	Multiple	5 min	79	38	2
2	Z-871	M	Yes	Caucasian	Yes	20 min	67	23	0
3	P-456	F	Yes	Unknown	No	30 min	84	15	3
So on									

END OF ASSESSMENT