University of Central Lancashire UCLan School of Engineering	**UCLan Coursework Assessment Brief**	Academic Year: 2022-23
	Module Title: Advanced Topics in Machine Intelligence Module Code: ER4165	Level: 7
	Mechatronic System Diagnostics	This assessment is worth 50% of the overall module mark

THE BRIEF/INSTRUCTIONS

The increase in the requirements for resiliency and sustainability in Mechatronics Systems leads to new challenges in maintainability and manageability issues. Most of the new mechatronics subsystems nowadays have the necessary sensors to log their status. These new intelligent machines allow the decision-making algorithm to help with diagnostics and condition monitoring of these mechatronics systems.

You have been tasked to analyse and test datasets from different mechatronics subsystems. Your answer should provide the necessary code, plots, comments on results, and other requirements.

Mechatronics subsystems are usually equipped with sensors to report and detect faults. When a system stops due to a failure, the data acquisition may stop, reducing the opportunity to collect enough data about that failure. Comparing the system failure to a normal operation, the data collected during system failure is quite small.

Diagnostic systems, condition monitoring, predictive maintenance, and many other current mechatronic applications, utilising machine learning (ML) algorithms, extensively use data with characteristics similar to the one described above.

In a normal operation of a system, the data collected from the system are assigned as negative cases (cases with no-fault). Usually, the system reports many of these samples. When the system experiences and reports failures, which are usually few, the cases are referred to as positive cases (failure cases). The dataset, in this case, consists of a large number of samples of normal system operation ('negative samples') and only a few of failure, i.e., 'positive samples'.

In this assignment, you will be asked to deal with such cases, and you will need to provide the right solution for each task specified below.

These tasks will test your ability to:

- Design ML methods and processes necessary to deploy an artificial intelligence system for quantitative evaluations of a mechatronic system.
- Recognize software design challenges behind implementations of machine learning algorithms.
- Design and optimise software to meet specified requirements.
- Design and provide a working solution for system analysis and diagnosis in a Mechatronics system.

These correspond to points 1, 2, 3, and 4 of the module learning outcomes.

Please note

1. **This coursework consists of three tasks. Provide answer for all the tasks.**
2. **Graphical Tools like Classification Learner, Clustering, and Curve Fitters are not allowed in this coursework. Only command-based solutions will be evaluated.**
3. **The provided Matlab file "ER4165_Assignment_22_23mlx", in the Online Editor file format, has designated spaces for you to write your answers, including texts and**

relevant matlab code. The code space can be expanded to fit your specific code requirements.

4. **For further information about editing the Matlab Online Editor, please check the following two external links: Link1 the Link2.**
5. **You should submit the Word version of this file in addition to the Matlab xlm file format to TurnitIn links on Blackboard.**

Task 1: Data Fitting (35 Marks)

Data fitting is one of the early methods of modeling datasets. One of the data fitting methods is Polynomial regression. Polynomial regression is a form of regression analysis in which the relationship between the independent variable x and the dependent variable y is modelled as an nth-degree polynomial in x. Although polynomial regression fits a nonlinear model to the data, as a statistical estimation problem, it is linear in the sense that the regression function $p(a, x)$ is linear in the unknown parameters that are estimated from the data. For this reason, polynomial regression is considered to be a special case of multiple linear regression.

The polynomial regression model of order n is defined as:

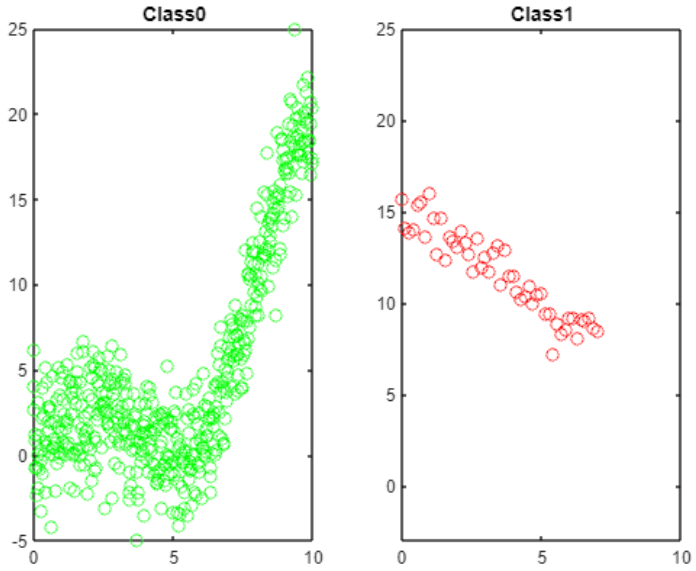
$$p(x) = a_0 + a_1x^1 + a_2x^2 + a_3x^3 + ... + a_nx^n$$

In the above equation, the input is the x, and the $a_0 \dots a_n$ are the parameters that need to be found. The function $p(x)$ can be written as a function of $p(x, a)$. Although the function is not linear with respect to x, it is linear with respect to the parameters $a_0 \dots a_n$, and the solution of this function that can minimize the error with respect to the dataset is to find the optimum parameter vector $a = [a_0 \dots a_n]$.

For this task :

You will use polynomial regression to model the datasets in the dataset0 and dataset1 files. These two files are also avilabe on Blackboard.
dataset0 and dataset1 are shown below:

```
Data0 = readtable('/MATLAB Drive/Examples/ER4165/Assigment/dataset0.csv','Delimiter',',');
Data1 = readtable('/MATLAB Drive/Examples/ER4165/Assigment/dataset1.csv','Delimiter',',');
subplot(2,1,1)
plot(Data0.Var1, Data0.Var2, 'og')
title('dataset0')
subplot(2,1,2)
plot(Data1.Var1, Data1.Var2, 'or')
title('dataset2')
```



1.1 Plot the dataset by combining the two datasets, Dataset1 and Dataset2.

(5 Marks)

1.2. Use regularised least square error, find the optimal degree of the regression polynomial (n) that produces the minimum error for the dataset0, and plot the fitted polynomial. Try different polynomial degrees to find the optimal value of (n)

(5 Marks)

1.3 Choose 20 random samples from dataset0 and plot the error at each sample with respect to the fitted polynomial.

(5 Marks)

1.4. Use regularised mean square error, find the optimal degree of the regression polynomial (n) that produces the minimum error for the dataset1, and plot the fitted polynomial. Try different polynomial degrees to find the optimum value of (n)

(5 Marks)

1.5 Choose 20 random samples from dataset1 and plot the error at each sample with respect to the fitted polynomial.

(5 Marks)

Comment on the results obtained on task 1 by justifying the obtained optimal degrees of polynomials, in particular compare and discuss the results for tasks 1.2 and 1.4.

(10 Marks)

Task 2: Clustering Algorithms Comparison (35 Marks)

Clustering is one of the most widely used techniques for data exploration in many application areas. Researchers and engineers may try to identify a meaningful grouping of their dataset. Applications span areas like predictive maintenance, social sciences, biology, computer science, and condition monitoring. People try to get a first intuition about their data by identifying meaningful groups among the data points. Many examples can be listed where clustering is heavily utilized.

Clustering is an unsupervised learning task as it usually lacks label data or what is also referred to as ground truth. In general, a multi-cluster problem describes a dataset defined as $D = \{x_n\}_{n=1}^{N}$ where

$x_n \in \mathbb{R}^D$ is a vector that needs to be assigned to a particular class. One of the well-known clustering algorithms is K-means clustering.

K-means is a heuristic method that is based on estimating the Euclidean distance with a suggested cluster representative μ_k. The algorithm maintains during its estimation a vector that assigns each data sample x_n to a cluster (hard allocation or assignment). This vector can be thought of in terms of quantization: all points within a given group can be represented by (quantized to) the cluster prototype μ_k with minimal quadratic loss. This relationship can be expressed by the following optimization problem over the cluster assignment variables z_n and the cluster prototype μ_k:

$$\min_{z_n, \mu_k} \sum_{n=1}^{N} \sum_{k=1}^{K} z_{kn} d(x_n, \mu_k)$$

where the d function can be defined as the Euclidean distance $d(x, \mu) = ||x - \mu||^2$.

Fuzzy C-Mean, Mean-Shift Clustering, DBSCAN, and Expectation Maximization with mixture of Gaussian

are other clustering algorithms that compete with the K-Means clustering. Each one of these algorithms has different computational characteristics and performance. Furthermore, these algorithms can perform very differently on the data drawn from the same distribution.

For this task :

The main aim of this task is to compare the performances of different clustering algorithms on a particular dataset.

You are asked to use the Clusters_Dataset data, which is a combination of dataset1 and dataset2 from the task one above. The data should be assigned into **two clusters**. The dataset will also be available on Blackboard.

You will follow the steps below to perform your task.

2.1 Show (Plot) the dataset (Clusters_Dataset) on a graph and use K-Means clustering to perform clustering of the Clusters_Dataset data.

(5 Marks)

2.2 Test the K-Means clustering performance using any appropriate evaluation method and explain any issue the algorithm experiences while clustering Clusters_Dataset **using a proper score metric.**

(5 Marks)

2.3 In light of the estimated score you obtained in step 2.2, identify the issue that usually arises with such dataset characteristics.

(5 Marks)

Explain Here:

2.4 Use another clustering algorithm of your choice that can overcome the issue identified in step 2.3 and obtain a better result than the K-Means clustering using a similar scoring that you performed in step 2.2

(10 Marks)

2.6 Explain the following regarding the algorithm that you chose in step 2.4:

1. **How does your chosen algorithm work?**

(4 Marks)

2. **What are the parameters used in the algorithm that you chose?**

(4 Marks)

3. **How is this algorithm different from the K-Means clustering?**

(2 Marks)

Task 3: Supervised and Unsupervised Learning (30 Marks)

In addition to regression and clustering, another machine learning task is classification. Classification is a fundamental task with a supervised learning requirement. Classification aims to partition the observation (predictor variables) space into regions. Each region is assigned one of the output classes.

Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are known as dimensionality reduction algorithms but can both be utilised for classification. The PCA is an unsupervised learning algorithm, while the LDA is a supervised learning algorithm.

In this task, you will compare the Principle Component Analysis (PCA) and the Linear Discriminant Analysis (LDA) and ask to classify the data available in two files class1 and class2. Each data point consists of two features that are defined as, $D = \{x_n\}_{n=1}^{N}$, where $x_n \in \mathbb{R}^2$. The first feature (x_1) is a temperature reading of a thermocouple sensor attached to fluid chamber, and the second feature (x_2) is a pressure reading obtained from a strain gauge sensor attached to a diaphragm on the base of the chamber. The obtained fluid that reaches the chamber is an output of a chemical process, and the characteristics of the obtained fluid (temperature and pressure) are collected into two datasets

Both datasets are available on Blackboard.

For this task :

You will use **class1** and **class2** datasets to test the performance of both PCA and the LDA to classify the data under test.

3.1 Combine the data from the two datastes (class1 and class2) into one dataset, call it *Fluid_DS,* and show the newly created Fluid_DS on a graph.

(5 Marks)

3.2 Plot the PCA components for the Fluid_DS and show the dominating component that can be potentially used for classification.

(5 Marks)

3.3 Project the Fluid_DS on the principal or primary component of the PCA and obtain a classification threshold between the two classes. Perform a performance test.

(5 Marks)

3.4 Perform classification on the Fluid_DS using the LDA and draw the LDA model against the projected Fluid_DS.

(5 Marks)

3.5 Compute the performance of the LDA model.

(5 Marks)

3.6 Comment on the result obtained and in particular, compare the result obtained for the PCA and the LDA algorithms.

(5 Marks)

Answer here:

Marking Scheme

Marking Scheme is provided within the context of the assignment.

Resources:
Bishop, C. M. (2006) Pattern Recognition and Machine Learning. Springer
Marshal, S. (2009) Machine Learning, an Algorithmic Perspective. CRC Press

PREPARATION FOR THE ASSESSMENT

The design study is to be introduced during the laboratory session on Tuesday the 7th of November. During that session, the **Mechatronic System Diagnostics** Problem will be introduced, the training/test datasets structures will be explained, and the expected results will be elucidated with examples. The set of software tools available for the assignment will also be briefly described. During November laboratory classes, additional support will be provided in the form of Q&A sessions.

You may also benefit from reviewing data visualisation techniques available in Matlab. It is expected that you will do this during your independent studies with support provided during class time if needed.

Your report submission will contain two portions. The first part is the **Matlab LIVE Editor xlm** file and the Word document or PDF vesion of the same xlm file. Each of these files will have its TurnitIn upload link.. The provided xlm file will be tested by the marker to confirm the integreity of the provided solution.

RELEASE DATES AND HAND-IN DEADLINE
Assessment release date:1st of November 2022 Assessment Deadline Date and Time: 23:59 of 16th of December 2023
Please note that this is the final time you can submit- not the time to submit!
Your feedback/feedforward and mark for this assessment will be provided on Blackboard

SUBMISSION DETAILS

Submission of assignment work
This design study will constitute 50% of the total module assessment mark. You should write a report for this design study explaining the algorithm, show and discuss the results, and consider any sources of error in the results. The reportshould be approximately 1500 words long plus relevant materials (References and Appendices). You should use Harvard referencing system for this report. The report should be submitted electronically toTurnitin through Blackboard.
In addition to this briefing, you will download the 'ER4165_Assessment_22_23.mlx' file that will be your main file providing answers and where you must document your solutions. The file allows you to write your code and comment on your results. Once you complete your answers, you will also save that file as a word document and submit it through TurnitIn for plagiarism checking. You will also need to submit the 'ER4165_Assessment_22_23.mlx' to the TureitIn link "*Upload your Matlab mlx File format here*". You will need to submit both files. **You will lose your CW mark if you do not submit both these files.**

Late work
Work submitted electronically may be submitted after the deadline to the same Turnitin assignment slot and will be automatically flagged as late.
Penalties for late submission
Except where an extension of the hand-in deadline date has been approved, lateness penalties will be applied in accordance with University policy as follows:

(Working) Days **Late Penalty**

1 - 5	maximum mark that can be achieved: 50%
more than 5	0% given

Plagiarism

During the induction and via your student handbook, you were informed of the serious consequences of using or attempting to use unfair means to enhance performance. This includes plagiarism. The work submitted must be your own and any information and material used properly identified and acknowledged.

The University operates an electronic plagiarism detection service where your work may be uploaded, stored and cross-referenced against other material. The software searches the World Wide Web and extensive databases of reference material to identify duplication.

For detailed information on the procedures relating to plagiarism, please see the current version of the University Academic Regulations.

HELP AND SUPPORT

- The support for this assignment will be provided during scheduled laboratory sessions.
- For support with using library resources, please contact <SubjectLibrarians@uclan.ac.uk>. You will find links to lots of useful resources in the My Library tab on Blackboard.
- If you have not yet made the university aware of any disability, specific learning difficulty, long-term health or mental health condition, please complete a Disclosure Form. The Inclusive Support team will then contact to discuss reasonable adjustments and support relating to any disability. For more information, visit the Inclusive Support site.
- To access mental health and wellbeing support, please complete our online referral form. Alternatively, you can email wellbeing@uclan.ac.uk, call 01772 893020 or visit our UCLan Wellbeing Service pages for more information.
- If you have any other query or require further support you can contact The <i>, The Student Information and Support Centre. Speak with us for advice on accessing all the University services as well as the Library services. Whatever your query, our expert staff will be able to help and support you. For more information , how to contact us and our opening hours visit Student Information and Support Centre.
- If you have any valid mitigating circumstances that mean you cannot meet an assessment submission deadline and you wish to request an extension, you will need to apply online prior to the deadline.

Disclaimer: The information provided in this assessment brief is correct at time of publication. In the unlikely event that any changes are deemed necessary, they will be communicated clearly via e-mail and a new version of this assessment brief will be circulated.	Version: 1