



Module: Statistical Learning Methods  
Academic Year: 2024-25  
Lecturers: Dr. I. Petrunin  
Title: **Statistical Learning for Predictive Maintenance**  
Date issued: 18 October 2024  
Submission date: *As indicated on Canvas.*  
Estimated time required: 50 hours

---

The assignment should be submitted electronically as a pdf, in the form of a brief individual report (2500 - 3000 words limit applies) explaining the methods and giving a clear statement of the principal results. Students should highlight the physical basis for any assumptions made wherever possible.

The final completed assignment should be submitted to Turnitin following the standard procedure on or before the submission date.

You are reminded that in the absence of exceptional circumstances (supported by written evidence) late submissions will be penalised.

## Aim

The aim of the work is to propose, implement and discuss solutions aiming to enhance the maintenance operations and planning of time-based preventive maintenance of the aircraft engine by selecting and applying statistical learning methods concerning the prediction of the remaining life and the prediction of current status of the engines.

## Problem Definition

Failure prediction is a major topic in predictive maintenance in many industries. Aircraft manufacturers, OEMs and end users are highly interested in prediction of component failures during the operation so that they can plan maintenance operations and reduce losses due to the time aircraft has spent at the ground.

Monitoring of the engine health and current condition is based on sensor data analysis and telemetry from the engine sub-systems. It is supposed to promote predictive maintenance by estimating either Time-To-Failure (TTF) or Remaining Useful Life (RUL) for aircraft components that are currently in-service and may be fully functional at the time of testing.

Based on the measurements from the sensors of the aircraft engine, the developed analysis framework should provide the following predictions, which are the objectives of this assignment:

1. Time-To-Failure (TTF) prediction for the engine
2. Classify which engine will fail in the analysed time period

## Data

The data for this assignment are extracted from the dataset that simulates run-to-failure scenario of engine operation and constitute the anonymised part of the bigger dataset generated by Microsoft that were used, in larger extent, for one of the projects within Springboard DS Career Track Bootcamp.

Text file '**train\_selected.csv**' contains 4 simulated sensors measurements for 100 aircraft engines running up to a failure. It is assumed that the engine progressing degradation pattern is reflected in its sensor measurements. An example of the data visualisation for the engine 1 is shown in Figure 1 that is obtained using the provided data importer for MATLAB '**dataImport.m**'.

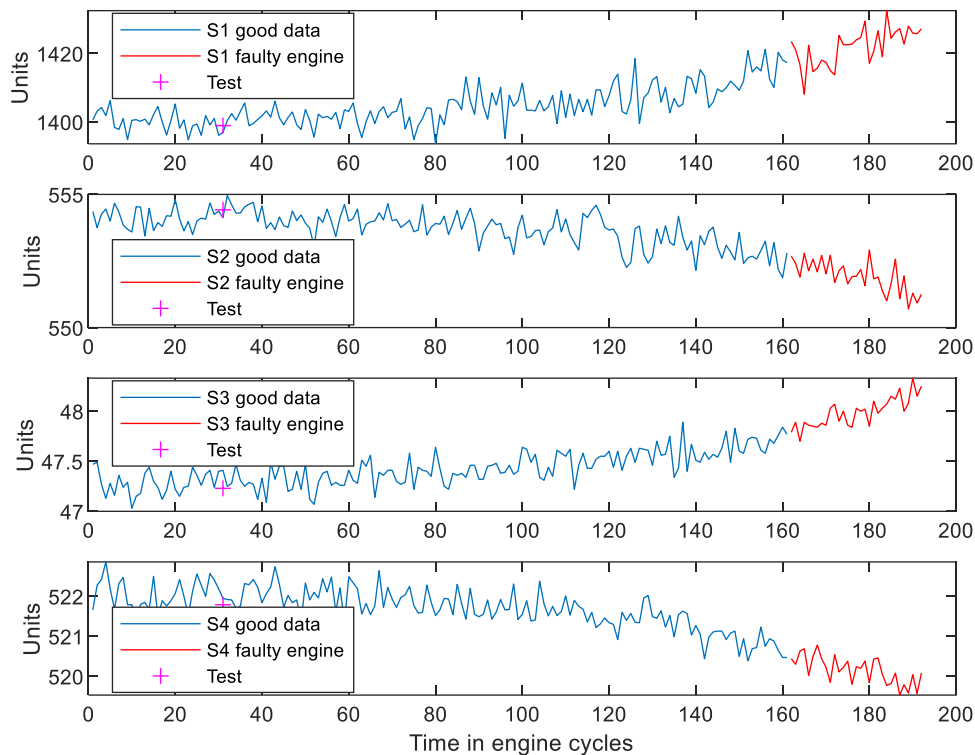


Figure 1 Training and test data for engine 1

In the training data the values of the expected Time-To-Failure are available along with the classification label that is set to 1 when TTF is within the last 30 cycles of engine operation.

File '**test\_selected.csv**' contains measurements from 4 sensors performed at a randomly selected cycle of engine operation. Engine degradation for the test data is expected to be following the same pattern as reflected in the training data, but TTF generally will be different to the one available through training.

True values for the TTF prediction are available separately in the file '**PM\_truth.txt**' and can be used for quantification of the prediction and classification results.

## Methodology

To achieve the objectives of this work you need to

1. Select and implement a regression method that will predict the number of remaining cycles before the failure of the engine utilizing training data for fitting the model or training the classifier and test data for prediction or classification.

While fitting the model you need to give a background to model selection and discuss achieved quality of fit using quantitative evaluation metrics.

After the prediction is done, it should be also discussed using the appropriate quantitative metrics and ground truth data.

2. Select and implement a binary (2-class) classification method that will classify whether or not the engine can be considered as faulty within the current operation cycle.

Similar to the previous case, the background for the method selection should be given along with the discussion of the results in a quantitative manner.

It is expected that during the assignment you will not be limited to the lecture materials only and will use other sources, such as books (both from the essential and additional reading lists) and online resources. For example, it is recommended to read chapter 7 of the book G. James, D. Witten, T. Hastie, R. Tibshirani, *An Introduction to Statistical Learning: with Application in R*. Springer, New York, 2013, ISBN 978-1-4614-7137-0, ISBN 978-1-4614-7138-7 from the recommended list for better understanding of possible approaches to solving a regression part of the assignment.

Any analysis tools are permitted – there is no requirement to use MATLAB or any other tool, but it is recommended to include the code in the report to enable markers to see possible reasons for the problems with results (if this happens) to account for this during the marking process.

The word count for the report is low (2500 - 3000 words, excluding references), therefore, results and discussion should be predominantly presented in a graphical or tabulated way. Reproduction of the problem statement and data description will not be included in the word count.

The assignment report should be prepared using the supplied template and include the word count at the beginning.

## Assessment

Assignment marking will be focused on the ability discuss the appropriateness of the techniques with the following selection and implementation and quality of the results assessment.

Use of the programming languages or tools is not assessed, i.e. type of the tool used or length of the code will not affect marks. Problem statement and data description (in case they are included in the report) will not be assessed too.

It is expected that in order to pass *at least one technique should be selected, implemented and discussed for both regression analysis and classification*.

The marks for the assignment will be distributed as follows:

1. Discussion and selection of the techniques for analysis that include data observation (descriptive analysis is welcomed) identification of the appropriate type of technique, model selection.

- a) Regression techniques [10 marks],
- b) Classification techniques [10 marks].

[Total: 20 Marks]

2. Work carried out, efforts and results that include implementation of techniques (including parameter initialisation, tuning and cross-validation considerations), completeness and qualitative correctness of results.

- a) Regression part [20 marks],
- b) Classification part [20 marks].

[Total: 40 Marks]

3. Analysis, discussion and conclusions that include selection, justification and application of metrics for analysis, comparison with ground truth, performance considerations, comparison between techniques and concluding remarks related to assumptions made during the selection process.

- a) TTF prediction [12 marks],
- b) Classification [12 marks].

[Total: 24 Marks]

4. Style and presentation that includes the presence of logical structure, appropriate citation style for references used, quality of graphical material (labels, legends, titles and captions as appropriate), readability of the text material, and clarity of results presentation in the text.

- a) Structure [6 marks],
- b) Clarity [10 marks].

[Total: 16 Marks]

## Marking rubric

	Fail (0-49%)	Pass/Satisfactory (50-59%)	Good (60-69%)	Excellent (70-100%)
Content (60)	Demonstrates inadequate knowledge of the subject	Demonstrates sufficient knowledge to address ILOs	Demonstration of knowledge meets all and exceeds some ILOs	Demonstration of knowledge exceeds many ILOs

Argument (24)	Absence of critique of the subject matter	Some critique of the subject matter	Good capacity for critical evaluation	High capacity for critical evaluation
Presentation (16)	A poorly structured and communicated piece of work. A large number of spelling or grammar errors; references incorrectly cited; Poor or no use of titles, subtitles, figures or tables; Lack of legends and labelling.	Simple structure with adequate communication skills Most spelling and grammar is correct; other presentational aspects generally correctly applied	Well-structured work with good communication skills Minor errors	Well-structured work with excellent communication skills No mistakes in spelling or grammar; references correctly and consistently cited; appropriate use of titles and subtitles; creative use of figures and tables to complement the text and are correctly labelled and referred to