Student Assignment Brief

This document is intended for Coventry University Group students for their own use in completing their assessed work for this module. It must not be passed to third parties or posted on any website. If you require this document in an alternative format, please contact your Module Leader.

Contents:

The work you submit for this assignment must be your own independent work, or in the case of a group assignment your own groups' work. More information is available in the 'Assignment Task' section of this assignment brief.

Assignment Information

Module Name: Modelling and Optimisation under Uncertainty

Module Code: 7135CEM

Assignment Title: Written Coursework

Assignment Due: 26/07/2024, 18:00 UK time

Assignment Credit: 15 credits

Word Count (or equivalent): 6000 words +/- 10%

Assignment Type:

Percentage Grade (Applied Core Assessment). You will be provided with an overall grade between 0% and 100% (**each task of this CW worth 50%)**. You have one opportunity to pass the assignment at or above 40%.

Assignment Task

Task and Mark distribution:

This coursework consists of two tasks, and you should attempt both and submit one Word or pdf file (or similar) for each task. Each task is worth 50 marks and the marks breakdown for each task is provided with each task. This coursework contributes 100% to your overall module mark.

Task 1: The Machine learning algorithms for solving real-world problems in Regression, Classification, modelling data, and text mining

Individual Research Paper: 50% of the module mark

Context

During this module, you learned about different advanced machine learning techniques, associated concepts and applications. We explored the Gaussian process model, which is computationally efficient method for Regression, Classification, optimization, etc. We have also covered the Bayesian networks as promising tools for modelling the data with complex dependency structure. Finally, you have learned how to use Dirichlet Latent processes for unsupervised learning applications, particularly text mining.

In this assignment, you will have to select an application related to a regression, classification, modelling un-structured data, or text mining problem, and explore how best to apply the machine learning algorithms to solve it. The selected application for each of the methods mentioned above should have the following features:

1. **Gaussian Process regression and Classification**: The application selected for any of these two methods must consist of **at least four input variables** and a **single output variable**. You must also implement Gaussian process classification by appropriately define a threshold on the output variable to create a binary or multiple classes first, and then apply the Gaussian process classification on the categorized output.
2. **Bayesian network**: If you are choosing an application for this method, this application must consist of at least **eight random variables**. The random variables could be all discrete or continuous or hybrid.
3. **There is no restriction** on selecting the application to apply the Latent Dirichlet allocation model for topic modelling.

There are **some potential projects** listed below, which could be studied to get some ideas. However, I strongly recommend you come up with **your own idea(s)** by reviewing these project and some other relevant and recent articles.

1. This dataset from the UCI repository is quite interesting. The task is to predict the depth in the body (effectively, the depth along the spine) given the properties of a two-dimensional "slice" of the body. The hard part about this problem is that it is actually the output causing the input rather

than the other way around. I have not had luck designing a good regression method for this data. Can you do this?

2. Find a Bayesian interpretation of elastic net regularization, and compare this method for regression against "standard" Bayesian regression (with a Gaussian prior) on a dataset of your choosing.
3. Probabilistic PCA using Gaussian Process is a Bayesian interpretation of the classical PCA algorithm for dimensionality reduction. Implement Gaussian Process based PPCA in Python, R or Matlab, and compare its performance with other methods (such as "standard" PCA) on a dataset of your choosing.
4. Bayesian optimization is very important issue with a wide range of applications. However, this was not fully studied during lectures, but it can be easily implemented using Gaussian Process. The Python codes and some examples can be found here!
5. The squared exponential covariance is widely used for Gaussian process regression. It is probably used in 90+% of all GP publications. That said, it is widely believed to be "too smooth" for many real-world regression tasks. Compare the squared exponential covariance versus the Matéern covariance on several datasets via Bayesian model selection. How often is the squared exponential the right choice?
6. Latent Dirichlet allocation (LDA) is a Bayesian method for creating "topic models" of text documents. There are plenty of interesting text datasets available (e.g., DBpedia could be a good resource!). One idea would be to compare the behavior of LDA with other techniques, such as latent semantic analysis.

You may be able to get **relevant dataset and ideas** by visiting the following sites:

- This competition site consists of some relevant data, and the relevant ideas could be developed by analysing this data. Also check dataset in Kaggle competitions.
- This website has a fantastic compilation of 100 interesting, relevant datasets from all sorts of application areas.
- The creators of libSVM have also compiled a great list of datasets, all in a standardized format. The libSVM codebase also includes *libsvmread* for reading these in MATLAB.
- The UCI Machine Learning Repository is a mainstay in machine-learning research. There is a wide range of datasets there from many different application areas and with many different properties (large, small, high-dimensional, low-dimensional, classification, regression, etc.).
- DBpedia is an amazing resource that automatically extracts structured data from Wikipedia. They have all sorts of data available for download in convenient formats. This tool can be used to extract labeled graphs from DBpedia, but there is so much more you could do.

The purpose of the first TASK of this coursework is to

- Examine the fundamental concepts of machine learning, their implementation and application.
- Perform appropriate preparation of a dataset and evaluate the performance of different learning algorithms on this dataset.
- Gain practical experience in selecting machine learning algorithms for solving a real-life Regression, classification, modelling data with complex dependency structure, or text-mining problems.
- Demonstrate effectiveness in project teamwork and leadership.

You will be required to:

- Work individually, developing a paper/report by considering the following instruction. You need to consider developing **at least one common methodology and one individual techniques for your analysis**.
- Consult with your tutor about your project work if needed during the Theta hrs**.**

- You will write a proposal (maximum of 1 A4 page), giving the title of the project, your name, the description of the problem and the plan of the work. You will need to **submit this proposal** to your tutor by **Friday 19th of July** via 7135CEM Aula submission link. In case of any required changes, you will receive feedback from your tutor.
- **Note: The proposal is not a separate assessment component, and its mark will be included in the final CW mark. The separate Aula submission link is only for you to submit your Proposal and for your Tutor to approve the selected problem and data for "TASK 1", and also provide you with any additional comments to improv your work before conducting the research.**

- Your final submission on **TASK 1** will include **a scientific paper (in 6 pages A4, up to 4000 words), written individually based on the experience and the derived results by fitting Machine learning methods covered in this module.**

You are encouraged to target a certain conference or journal and submit the proposed paper to it. You can either use the template of Machine Learning Journal, or any other single column formats from other relevant journal or conference sites.

The paper should broadly include the following sections:

- Abstract
- Introduction (where you introduce the problem along a short literature review of related work; if the literature review is longer, it is recommended to be a section on its own)
- Problem and Data set(s) description (where you describe in detail the problem you want to solve and its significance)
- Methods (where you shortly describe the machine learning methods and/or other methods employed to solve the problem)
- Experimental setup (including data pre-processing, feature selection and extraction)
- Results
- Social, ethical, legal and professional considerations
- Discussion and Conclusions
- References

These are generic section titles, which you may adapt appropriately to the application/problem that is being investigated. You may include sections describing modifications of algorithms or developments that are novel and specific to your work. You may include figures, tables, pseudo-code, and appendices with the actual code that has been developed.

The project general guidelines and milestones:

Please note, the following guidelines are good practice and should lead to better result, but you have the freedom to pick whatever is suitable for your style:

- You have to select a challenging real world problem and one (or more) appropriate data set(s) as suggested above. You could also use the following links, which have numerous problems and data sets:
    1. Learning Repository: http://archive.ics.uci.edu/ml/;
    2. Kaggle competitions: http://www.kaggle.com/competitions;
    3. PhysioBank Databases https://archive.physionet.org/physiobank/database/
    4. kdnuggets https://www.kdnuggets.com/datasets/index.html
    5. DATA CATALOG https://catalog.data.gov/dataset?tags=diabetes

6. Vision & Eye Health Data Portal
https://chronicdata.cdc.gov/browse?category=Vision+%26+Eye+Health

Marking Scheme for Task 1	**Mark**
1) Proposal • The description of the problem and the initial plan of the work. • The initial and correct application of the proposed methods for the selected dataset. Notes: **1. You will not get the full marks of this section if you submit your proposal late.** **2. If the final submission of your CW is the different to what you propose in your proposal, you will not get any marks for parts 2 & 3.**	5
2) Technical quality 1. Rigour and extent of the experiments. 2. Correct application of the selected algorithms and suitability of the methods. 3. Data preparation - technical quality. 4. Extent of evidence of running the experiments provided in appendices.	5 5 5 5
3) Evaluation 1. Evaluation and discussion of the results. Why the results are important? How would the results be useful to other researchers or practitioners? 2. Is this a "real" problem or a small "toy" problem? How does the paper advance the state of the art? **Notes about Valid codes**: • All your programming code should be included in the Appendix of your report or provided via a valid GitHub link. Please display them in a structured way (put headings for each Task implemented in your CW), with appropriate comments/annotations. • You need to attach the original R code (or Python/Matlab), **NOT the screenshots of the code**. • The code will be marked as part of the above marking scheme (for all the Tasks in this coursework, you will need to provide the corresponding code; when you describe/discuss the Tasks in the main text of the report, please reference the corresponding code section in the Appendix or link).	7 2
4) Social, ethical, legal and professional considerations related to the problem in question.	2

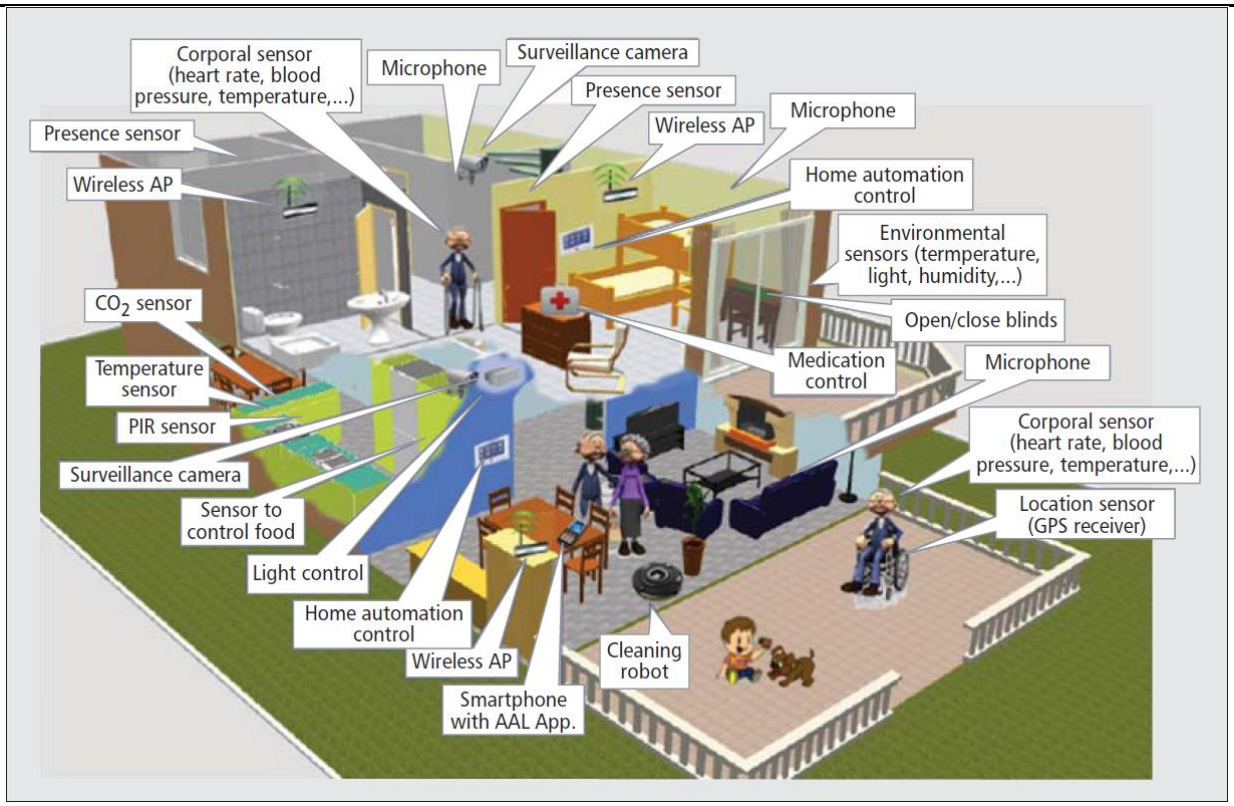
5) Clarity of the writing: 1. Is there sufficient information for the reader to reproduce the results? Is the language used in the paper good? 2. References and general presentation; Are results clearly presented, with appropriate visualizations?	5 3	
6) Originality: 1. Is there some original approach to the problem, original use of techniques? 2. Is there any (and how much) difference from previous contributions?	3 3	

Task 2: Evolutionary and Fuzzy Systems

Fuzzy Logic Optimized Controller (FLC) for an Intelligent Assistive Care Environment

Design and Implement an FLC for controlling the environmental parameters of an intelligent flat for disabled residence, where the system needs to automate the regulation of environmental conditions and user preferences or the operation of assistive equipment, ramps, auto adjusting furniture, kitchen worktops, HAVC, lighting or water temperature control. The environment could be based on a room of choice in a small flat. A more ambitious project might consider the aspects of the whole flat, but this is left to your choice.

The environmental parameters to be controlled could be ambient temperature, thermal conform and lighting using actuators such as cooling fans, heaters/boilers, blinds and dimmer switches. You might also consider other parameters such as TV or music volume control, and power down options for electronic devices and heating. Environmental parameters could be controlled based on monitoring sensors such as temperature, humidity, weather conditions, light levels, time of day, level of activity / motion of the user as well as mood and qualitative indicators, such as user preferences.



More details and the above figure can be found at: https://bstassen.wordpress.com/tag/ambient-assisted-living/

FLC Design

The FLC should be based on determining the inputs and outputs of the system, depending on what control behaviour(s) you decide the FLC should implement. ***Note*** *that, depending on the control behaviours you wish to implement, you can select to use a subset of the input sensors for example, so first think about the behaviour(s) the FLC should control*.

Design choices should be made to consider the type and number of fuzzy sets for the inputs and/or outputs of the FLC.

A set of suitable control rules should be defined, which can be experimented with to achieve a good control performance of the chosen behaviour(s).

The FLC should therefore implement the followings:

- Consideration of which Fuzzy Inference model to use: Mamdani or Sugeno (TSK) fuzzy models.

- Mapping the crisp input and output data into the designed fuzzy sets.
- Map input fuzzy sets into output fuzzy sets *(for Mamdani model)* based on a set of designed rules that capture the desired control behaviour of the robot.
- Employ appropriate inference operation (*rule implication*) that handles the way in which rules are activated and combined together (*composition and aggregation*).
- The outputs of the fuzzy inference engine will define a modified output fuzzy set (for Mamdani model) that specifies a possibility distribution of the control actions in relation to activated rules.

- Use an appropriate defuzzifier to convert the modified fuzzy outputs into nonfuzzy (crisp) control values that can then be used to set the actuation outputs.

Part 1 – Design and Implementation of the FLC

(35 Marks)

Design and implement a demonstrable FLC, which can be a simulated system programmed in Matlab, FuzzyLite or Juzzy, see links below:

- Matlab Fuzzy Logic Toolbox (http://uk.mathworks.com/videos/getting-started-with-fuzzy-logic-toolbox-part-1-68764.html,
- http://www-rohan.sdsu.edu/doc/matlab/toolbox/fuzzy/fuzzyt10.html)
- Fuzzylite (*http://www.fuzzylite.com*)
- Juzzy (*http://juzzy.wagnerweb.net*)

Provide suitable evidence of your implementation in the form of diagrams and screenshots of the different components.

(18 marks)

Discuss and justify your design decisions for the choice of fuzzy sets - membership functions, fuzzy rules, FLC inference mechanism selected, and defuzzification method that was chosen. Back up your explanations with evidence in the form of appropriate diagrams and screenshots.

(10 marks)

Perform analysis of the output behaviour of the controller showing the rules activation, controller output and control surface plots, demonstrating how the controller achieves the specified behaviours in relation to an operational scenario.

(7 marks)

Part 2 – Compare different optimization techniques on CEC'2005 functions.

(15 Marks)

Choose three functions from the CEC'2005 suite of benchmark functions available here:

http://www.cmap.polytechnique.fr/~nikolaus.hansen/Tech-Report-May-30-05.pdf

More details about the special session at CEC'2005 can be found here:

https://www.al-roomi.org/benchmarks/cec-database/cec-2005

Some of the links in these pages are broken, but **you will be able to download the MATLAB code for the functions if you click "Resources Database (Different Formats) [Download]"** on the last web page indicated above.

This part is to compare the performance of at least 3 different optimization techniques on the three functions you have chosen, for both D=2 and D=10, where D is the number of dimensions. If you want to challenge yourself, you may try higher dimensional spaces, for example D=100, but this is optional. As optimization techniques to compare in this part, you may choose Genetic Algorithms, Particle Swarm Optimization, Simulated Annealing, or other optimization methods available in the Global Optimization Toolbox or the Optimization Toolbox in MATLAB or developed as standalone programs by yourself.

To make the comparison meaningful you would have to run each optimization algorithm 15 times and report the average performance (including the standard deviation of the obtained results), as well as the best and the worst performance among the 15 runs. You may try to compare your results with results reported in the literature on the same functions.

In your report, you should include the description of the functions you have selected, the MATLAB code for those functions, the results obtained, and the parameters of the optimization algorithms used to obtain the reported results, any other MATLAB scripts or code used in your simulations, convergence graphs, etc.

Parts and Mark distribution for Task 2:

Part 1	**35**
Part 2	**15**

Submission Instructions:

Submission arrangement online via AULA/CUMoodle:
Submit before 18:00, late work will receive a mark of zero.
File types and method of recording: **Submit a Single Word file by putting your outputs to both tasks in this word file.**
Mark and Feedback date: **10/08/2024**
Mark and Feedback method: **given on each script.**

Marking and Feedback

How will my assignment be marked?

Your assignment will be marked by the module team.

How will I receive my grades and feedback?

Provisional marks will be released at 10/08/2024.

Feedback will be provided by the module team alongside grades release on each script, which was submitted to Aula.

Your provisional marks and feedback should be available within [2 weeks (11 working days)].

What will I be marked against?

Details of the marking criteria for this task can be found at the bottom of this assignment brief.

Assessed Module Learning Outcomes

The Learning Outcomes for this module align to the marking criteria which is provided above. Ensure you understand the marking criteria to ensure successful achievement of the assessment task. The following module learning outcomes are assessed in this task:

On completion of this module the student should be able to:

1. Apply supervised and unsupervised learning applications using Gaussian process emulators.
2. Apply Dirichlet processes for unsupervised learning applications.
3. Develop the knowledge and skills necessary to design, implement and apply the Graphical models to solve real world applications.
4. Evaluate the applications of fuzzy systems and their usage in hybrid intelligent systems, in combination with evolutionary computing and other machine learning methods.

Apply evolutionary computing methods to develop solutions for the real-world optimisation problems and appraise their advantages and limitations.

Assignment Support and Academic Integrity

If you have any questions about this assignment please see the Student Guidance on Coursework for more information.

Spelling, Punctuation, and Grammar:

You are expected to use effective, accurate, and appropriate language within this assessment task.

Academic Integrity:

The work you submit must be your own, or in the case of groupwork, that of your group. All sources of information need to be acknowledged and attributed; therefore, you must provide references for all sources of information and acknowledge any tools used in the production of your work, including Artificial Intelligence (AI). We use detection software and make routine checks for evidence of academic misconduct.

Definitions of academic misconduct, including plagiarism, self-plagiarism, and collusion can be found on the Student Portal. All cases of suspected academic misconduct are referred for investigation, the outcomes of which can have profound consequences to your studies. For more information on academic integrity please visit the Academic and Research Integrity section of the Student Portal.

Support for Students with Disabilities or Additional Needs:

If you have a disability, long-term health condition, specific learning difference, mental health diagnosis or symptoms and have discussed your support needs with health and wellbeing you may be able to access support that will help with your studies.

If you feel you may benefit from additional support, but have not disclosed a disability to the University, or have disclosed but are yet to discuss your support needs it is important to let us know so we can provide the right support for your circumstances. Visit the Student Portal to find out more.

Unable to Submit on Time?

The University wants you to do your best. However, we know that sometimes events happen which mean that you cannot submit your assessment by the deadline or sit a scheduled exam. If you think

this might be the case, guidance on understanding what counts as an extenuating circumstance, and how to apply is available on the Student Portal.

Administration of Assessment

Module Leader Name: Omid Chatrabgoun

Module Leader Email: ad8337@coventry.ac.uk

Assignment Category: Written

Attempt Type: Standard

Component Code: CW

Generic Marking Rubric

PG

Mark band	Outcome	Guidelines
90-100% Distinction	Meets learning outcomes	Distinction - Exceptional work with very high degree of rigour, creativity and critical/analytic skills. Mastery of knowledge and subject-specific theories with originality and autonomy. Demonstrates exceptional ability to analyse and apply concepts within the complexities and uncertainties of the subject/discipline. Innovative research with exceptional ability in the utilisation of research methodologies. Demonstrates, creativity, originality and outstanding problem-solving skills. Work completed with very high degree of accuracy, proficiency and autonomy. Exceptional communication and expression demonstrated throughout. Student evidences the full range of technical and/or artistic skills. Work pushes the boundaries of the discipline and may be strongly considered for external publication/dissemination/presentation.
80-89% Distinction		Distinction - Outstanding work with high degree of rigour, creativity and critical/analytic skills. Near mastery of knowledge and subject-specific theories with originality and autonomy. Demonstrates outstanding ability to analyse and apply concepts within the complexities and uncertainties of the subject/discipline. Innovative research with outstanding ability in the utilisation of research methodologies. Work consistently demonstrates creativity, originality and outstanding problem-solving skills. Work completed with high degree of accuracy, proficiency and autonomy. Outstanding communication and expression demonstrated throughout. Student demonstrates a very wide range of technical and/or artistic skills. With some amendments, the work may be considered for external publication/dissemination/presentation
70-79% Distinction		Distinction - Excellent work undertaken with rigour, creativity and critical/analytic skills. Excellent degree of knowledge and subject-specific theories with originality and autonomy demonstrated. The work exhibits excellent ability to analyse and apply concepts within the complexities and uncertainties of the subject/discipline. Innovative research with excellent ability in the utilisation of research methodologies. Work demonstrates creativity, originality and excellent problem-solving skills. Work completed with very consistent levels of accuracy, proficiency and autonomy. Excellent communication and expression demonstrated throughout. Student demonstrates a very wide range of technical and/or artistic skills.
60-69% Merit		Merit - Very good work often undertaken with rigour, creativity and critical/analytic skills. Very good degree of knowledge and subject-specific theories with some originality and autonomy demonstrated. The work often exhibits the ability to fully analyse and apply concepts within the complexities and uncertainties of the subject/discipline.

		Very good research evidence and shows very good ability in the utilisation of research methodologies. Work demonstrates creativity, originality and problem-solving skills. Work completed with very consistent levels of accuracy, proficiency and autonomy. Very good communication and expression demonstrated throughout. Student demonstrates a wide range of technical and/or artistic skills.
50-59% Pass		Pass - Good work undertaken with some creativity and critical/analytic skills. Demonstrates knowledge and subject-specific theories with some originality and autonomy demonstrated. The work exhibits the ability to analyse and apply concepts within the complexities and uncertainties of the subject/discipline. Good research and shows some ability in the utilisation of research methodologies. Work demonstrates problem-solving skills and is completed with some level of accuracy, proficiency and autonomy. Satisfactory communication and expression demonstrated throughout. Student demonstrates some of the technical and/or artistic skills.
40-49% Pass		Pass - Assessment demonstrates some advanced knowledge and understanding of the subject informed by current practice, scholarship and research. Work may be incomplete with some irrelevant material present. Sometimes demonstrates the ability to analyse and apply concepts within the complexities and uncertainties of the subject/discipline. Acceptable research with evidence of basic ability in the utilisation of research methodologies. Demonstrates some originality, creativity and problem-solving skills but often with inconsistencies. Expression and presentation sufficient for accuracy and proficiency. Sufficient communication and expression with professional skill set. Student demonstrates some technical and/or artistic skills.
30-39% Fail	Fails to achieve learning outcomes	Fail - Very limited understanding of relevant theories, concepts and issues with deficiencies in rigour and analysis. Some relevant material may be present but be informed from very limited sources. Fundamental errors and some misunderstanding likely to be present. Demonstrates limited ability to analyse and apply concepts within the complexities and uncertainties of the subject/discipline. Limited research scope and ability in the utilisation of research methodologies. Limited originality, creativity, and struggles with problem-solving skills. Expression and presentation insufficient for accuracy and proficiency. Insufficient communication and expression and with deficiencies in professional skill set. Student demonstrates deficiencies in the range of technical and/or artistic skills.
20-29% Fail -		Fail - Clear failure demonstrating little understanding of relevant theories, concepts, issues and only a vague knowledge of the area. Little relevant material may be present and informed from very limited sources. Serious and fundamental errors and virtually no evidence of relevant research. Fundamental errors and misunderstandings likely to be present. Little or no research with no evidence of utilisation of research methodologies. No originality, creativity, and struggles with problem-solving skills. Expression and presentation insufficient for accuracy and proficiency. Insufficient communication and expression and with serious deficiencies in professional skill set. Student has clear deficiencies in range of technical and/or artistic skills.

0-19% Fail		Fail - Clear failure demonstrating no understanding of relevant theories, concepts, issues and no understanding of area. Little or no relevant material may be present and informed from minimal sources. No evidence of ability in the utilisation of research methodologies. No evidence of originality, creativity, and problem-solving skills. Expression and presentation deficient for accuracy and proficiency. Insufficient communication and expression and with deficiencies in professional skill set. Student has clear deficiencies in range of technical and/or artistic skills.