



UCL DEPARTMENT OF INFORMATION STUDIES

Coursework Assignment Brief

Assignment Title	Group Project - Regression Analysis
Component/Module	INSTG083/INSTM083 - Foundations of Machine Learning and Data Science
Assignment Code	GP
Set By	Luke Dickens
Moderated By	Antonis Bikakis

Learning outcomes to be assessed:

- Understanding, implementing and applying a selection of regression models, basis functions and fitting approaches to data. This includes content from the lectures, such as:
 - Linear regression
 - kNN regression
 - Linear Models (basis function regression)
 - Polynomial basis functions
 - Radial basis functions
 - S-shaped basis functions
 - Least squares/maximum-likelihood fitting
 - Regularised least-squares approximation
 - Bayesian regression
- Implementing any required components, and reusing code from tutorials, to train and test these models on data, including:
 - Reading data from file
 - Constructing data and design matrices as required
 - Defining appropriate basis functions
 - Implementing any models/algorithms not provided in tutorial code
 - Implementing an appropriate validation framework for model selection
 - Plotting results in a clear readable manner
- Reporting methodologies and results, and interpreting any findings in a way that demonstrates an understanding of the course materials on regression.



Submission requirements:

- For each group, the group representative should submit a single report of no more than 10 pages (Arial font, size 11), and a zip archive containing the group code.
- Code archives when unzipped, should have a main module called `group_project_evaluation.py`. When run, with the command:

```
python group_project_evaluation.py datafile.csv
```

this should load data from `datafile.csv` and recreate all results/figures in the report. Any results not plotted in a figure, should be printed clearly to screen.
- Each individual student should submit a short report (max 500 words) which reflects on what they have learnt during the assignment, including a breakdown of work among group members.
- Submission is through the moodle submission page for INSTG083 (<https://moodle.ucl.ac.uk/course/view.php?id=42291>).

Basic Assignment Description	The student group must implement and evaluate a selection of regression methods on the provided data. On completion, each group should submit one report describing the models used and the training, validation and testing methods applied, as well as a code archive (zip). When run, the code should recreate all computational processes, figures and results described in the report. Individual students should also submit a short report reflecting on what they learnt during the assignment, including a breakdown of work among group members.
Conditions	<p>Each group should specify a group representative who will submit the main report and code archive. Each machine-learning method evaluated should be described (briefly) in the report. Implementations can use native python code and the NumPy, SciPy, Matplotlib and Pandas libraries as well as any system libraries such as csv, sys, os. Students must not use any machine learning libraries within their implementations, e.g. scikit-learn.</p> <p>If students wish to investigate methods not covered in the lectures, then this is only possible so long as module leader feels the methods are appropriate. Therefore, students should seek the module leaders agreement under these circumstances. Any such methods must be clearly described and cited in the report, fully implemented in the allowed libraries, and the origin of any supporting code appropriately cited too.</p>



Marking and Rubric	Criteria Weighting
	<p>Reports will be assessed on their clarity, the appropriateness of methodologies & figures, and the relevance of the findings. Any models should be correctly and clearly described and the evaluation procedures (e.g. cross-validation) explained. Students do not need to re-explain the material covered in lectures, but should give enough information that an independent researcher could reconstruct their experiments, e.g. number of runs, parameters evaluated etc.</p> <p>Plots should be appropriately chosen, interpretable and clearly labelled, where appropriate, plotting lines or marks should be given a legend. All figures and tables should have clear captions, and salient features discussed within the body text.</p> <p>The report should conclude with a discussion of which general methods performed best on the data, what assumptions were made, and how certain the students are of their findings. Code will be assessed on whether it runs without errors, and should recreate all results/plots included in the report.</p> <p>Marks are broken down as follows. There are up to 70 marks assigned to the group submission (report and code). All group members will get the same mark for the group submission. The maximum marks for the group submission break down as follows:</p> <ul style="list-style-type: none">• 14 marks (20%) for description of Machine Learning methods used and evaluation procedures.• 28 marks (40%) for results and evaluation itself, including figures, tables and text interpreting and describing them.• 14 marks (20%) for the discussion and concluding remarks• 14 marks (20%) for the code (that it is clear, and produces the plots and results in the report). <p>The remaining 30 marks for each student will depend on the individual reports. The average total (group+individual) mark for group members will be equal to the percentage mark of the groups submission (with partial marks rounded up). If all students are seen to have contributed equal effort to the project, then the group submission mark (as a percentage) will be each students total mark. If some students are regarded as having contributed more than others, then marks from other group members will be reassigned to them.</p> <p><i>For instance, if a group of four gets 42/70 marks (60%) for the group submission, then the average individual component will be 18/30 marks (also 60%). If one member is thought to have contributed very little to the project, they could lose all 18 of these marks to the other group members. If the other members each take 6 additional marks, those members would have (42+18+6)/100 marks (66%) and the under-contributing member would get (42+18-18)/100 marks (42%).</i></p>



The assignment is worth % of the overall assessment for this course.

The assignment must be completed:

Date work set (provisional):

Date and time due in (provisional):

Standard lateness penalties will apply

Target date for return of marked work and full feedback (provisional):

(This provisional date is within 4 working weeks of submission, according to DIS policy.)

A detailed description of the assignment is: