

Gesture UI Project 1 2025

Due date : 21st March by 11pm.
Worth 35% of your overall grade.

1 Intro

In this assignment, you will take a data set that is supplied to you and use it to compare three classification algorithms and finally selecting one model that would be good enough for “deployment”.

You should train three classifier methods and compare the results e.g. SVM Vs Logistic Regression Vs kNN Vs Random Forest. Use cross-validation when suitable and try different kernels etc.

Your method section should set out how you cleaned the data, how you trained the classifiers and how you compared the results.

You must look at all problems yourself, such as imbalanced datasets, feature scaling, changing classes to a numeric value, determining if features are useful.

Important: Any random seeds must use your student ID number, (without the G and leading 0's). Example if G00432121 is your id number, then your seed must be 432121. If you do not do this, you will lose a lot of marks!

You can choose one of two datasets:

Dataset Option 1

The raw dataset comes from Human Activity Recognition Using Smartphones, <https://archive.ics.uci.edu/dataset/240/human+activity+recognition+using+smartphones> and has all the problems that would come with real-world data sets so you will have to do a lot of looking at the data to use it correctly. This is supplied as `activity.zip` on Moodle. Use the version I supplied, not the original.

The dataset takes many readings from the smartphone of the user, readings from the accelerometer and gyroscope, completing different activities. 30 volunteers of different ages completed the tasks - walking, walking-upstairs, walking-downstairs, sitting, standing, laying were the 6 tasks recorded. `features.txt` provided in the zip file contains a list of all features.

Dataset Option 2

The dataset provided comes with 70,000 images of different types of fashion. They are all png files.

For example, pictures of T-shirt/top, Trouser, Pullover, Dress, Coat, Sandal, Shirt, Sneaker, Bag, Ankleboot.

It is your job to develop a model that can distinguish these categories of clothes.

The Dataset/Tasks

- No matter which dataset you choose, investigate the data.
- Charts/Visualisations/Examine the data for missing information/imbalances etc.
- You should describe the process you used to make your choices for the model(s).
- Describe the chosen model and evaluate its performance properly.
- You should be able to summarise your results and processes sufficiently so that it fits into the restricted paper size.
- If I haven't even mentioned a method in class, I do not want you to use it in the project.

Resources

On the [Kaggle](#) page has a Python workbook that has code to visualise data. It may be worth reusing this to visualise the data.

Deliverables

The deliverable is a paper of at most 5 pages explaining your process and then the results. Some visualisations should be used too. The paper should be written in an academic style, i.e. in the third person and should reflect the project (no I or we), not the personal learning. You should not have snippets of code in your paper, it should explain the processes and the results - writing a function name is using code so do not do this. Say what was done and it should be language-agnostic, i.e. it does not refer to one particular language/library.

Submit the code you use to generate the results as part of your submission. The python workbook/py files, whatever method you go for.

2 Expectations

- Excellent grammar, spelling and technical writing are all essential skills that you must demonstrate.
- You should be able to explain your findings in a thorough but concise manner.
- You should show evidence of learning/interpretation from the primary sources.
- Don't insert code into the paper, all code goes in the code files **NOT** the paper.
- If you need to explain an algorithm, then do so with pseudo code.

3 Headings

The section headings are up to you and what makes most sense for your document. However, here are some suggestions

1. Introduction
 2. Methodology
 - (a) Data Pre-Processing, Labelling, Data Analysis and Visualisation, Other Techniques
 3. Experiments and results
 - (a) Classifier Models
 - (b) Results
 4. Conclusion
- or maybe
1. Introduction
 2. Data and Features
 3. Data Preprocessing
 4. Classifier Training
 5. Results Comparison
 6. Conclusion
 7. References

4 Requirements

- Upload in PDF only, you will be docked 10 % for all other formats.
- You will be docked 5% for each day you are late with submission, in accordance with ATU Marks and Standards.
- You will be docked 10% for every page over the page count limit of 5 pages.
- Plagiarised work will receive a mark of 0% for a first offence, see definition of plagiarism below.
- This is a solo project, not for groups to do!
- Some students may be selected to defend their work in the form as a viva as prescribed in Marks and Standards.

5 Marking

The assignment is marked primarily on the paper you submit, the notebook is to be submitted (and must have all outputs run) for verification of your results in the paper.

These are the categories that are marked

- Overall Document Quality - Writing Style, Presentation, Structure
- Data Analysis and Visualisations
- Proper Implementation of the 3 algorithms (each marked individually)
- Interpretation of numbers
- Overall Results and their Presentation

Each of these is marked on a rubric:

0	1	2	3	4	5
Not Answered/Completely	Poor	Weak	Average	Good	Excellent

Overall Results and Overall Document worth more than the other individual sections.

6 Suspected Cheating & Plagiarism

Student plagiarism occurs when a student presents the work of another as their own work, without appropriate acknowledgement. It can include:

- Presenting work which has been copied from the Internet, books, journals or other sources;
- Presenting work which paraphrases the writings of other authors, without acknowledgement;
- Presenting work which has been written by somebody else, such as another student or a family member;
- Presenting work which has been purchased from an Internet site or other source and submitting as own work;
- Presenting work which has been produced collaboratively as one's own individual work;
- Student plagiarism can also occur where a student submits the same piece of their own work for a number of different assignments;

Plagiarism by students can be deliberate, where a student intentionally attempts to pass off the work of another as their own work. It can also be accidental, where a student fails to use appropriate citation and referencing in their work or is unaware of what constitutes plagiarism.