Sensor Guide

Folder Submission (70% of grade)

The folder submission is the collected mandatory assignments in the course. The students have had three mandatory assignments spread throughout the semester. They have been given feedback and a chance to update their submissions befor the final submission which will be graded.

How to Grade the Mandatory Assignments

There are two sides to all the assignments

- A technical side dealing with the tools
    - Have they used the appropriate tools for the job?
    - Have they used the tools correctly?
- How well are the tools used to acheive the intended goal?
    - Do the experiments follow a logical set of steps?
    - Do they give reasonable descriptions of their results?
    - Are they able to build on their findings?

A passing grade of E would require

- Have delivered experiments for all assignents, even though not matching the specified scope (number of datasets and techniques) i.e. at least one dataset and technique for all part of the assignments
- Both code and documentation is present for all assignments
- Should be able to use the tools to create working models, and that the models are tested on suitable data.
- Should be able to provide somewhat sensible description of what they have done why.

A normal grade of C would require

- The delivered experiments should match the scope specified in the assignment
- Both code and documentation is present for all experiments
- Should be able to use the tools to create working models, and that at least some of the relevant hypeparameters have been optimized. The models should have been tested on suitable data and with more than the metrics used for training.
- Should be able to give a somewhat sensible description of what they have done and why

The top grade of A would require

- The scope needs to be at least as big as specified, but likely more (dependent on the quality of work).
- Both code and documentation needs to be of high quality
- Should be able to use the tools to create working models, and that all the relevant hypeparameters have been optimized thoroughly. The models should have been tested on suitable data and with more than the metrics used for training.
- Should be able to give a good description of what they have done and why
- Should be able to use and describe tools beyond the specified curriculum.

Mandatory 1: Data Processing and Analysis

In this assignment you must find at least three datasets: one classification dataset, one regression dataset, and one clustering dataset (ideally unlabeled, but removing label from labeled dataset can also work).

With each dataset you must make the datasets ready for analysis. This can involve

- Filling in missing values
- Normalizing real values
- Encoding categorical values

You also must perform exploratory data analysis. The results from this analysis is typically what motivates the processing. This can involve

- Feature correlation investigations
- Feature distribution investigations
- Feature selection
- Feature engineering

You must divide the dataset into training, validation and test sets using a suitable sampling method (if your instances are grouped you must divide accordingly)

You must write a short report describing the datasets, the processing performed on them, and the resulting processed data. You need to describe the results you get in your analysis, and describe why you do the processing you choose to do. You should also attach/include the code you have used. You can choose to use a notebook instead of separate code and report.

Mandatory 2: Classification & Clustering

This assignment is split into two parts; one for classification and one for clustering.

Classification

Select at least one non-trivial classification dataset.

Describe the dataset

- what is purpose of the dataset?
- how is it structured?

The primary goal of the assignment is to find which classification technique, or possibly, combination of techniques, provide the best models for the dataset.

You must test at least three different classification techniques to test on each of the selected datasets

You must take steps to ensure that

- the data is preprocessed fairly and optimally for each of the techniques
  - this requires exploratory analysis
- that each technique is properly investigated i.e. that different hyperparameters are tested
  - manual approaches are accepted
  - automated approaches are suggested
- that the final comparison is done on previously unseen data
  - You should test you models thoroughly with several metrics and statistics

Clustering

This part of the assignment will be conducted in the same manner as the classification part, but this time using unlabeled data and clustering algorithms.

Delivery

You must deliver a report and your code. Remember to discuss your results and steps taken.

The delivery can be done with separate pdf and py files zipped togeter, or in a zipped notebook.

Mandatory 3: Neural Networks & Regression

This assignment is split into two parts; one for classification and one for clustering.

Regression

Select at least one non-trivial regression dataset.

Describe the dataset

- what is purpose of the dataset?
- how is it structured?

The primary goal of the assignment is to find which regression technique, or possibly, combination of techniques, provide the best models for the dataset.

You must test at least three different regression techniques to test on each of the selected datasets

You must take steps to ensure that

- the data is preprocessed fairly for each of the techniques
- that each techinique is properly investigated i.e. that different hyperparameters are tested
  - manual approaches are accepted
  - automated approaches are suggested
- that the final comparison is done on previously unseen data

Artificial Neural Networks for Classification

Select at least one non-trivial classification dataset.

Describe the dataset

- what is purpose of the dataset?
- how is it structured?

The primary goal of the assignment is to find what type of network architecture provide the best results for the dataset.

You must test at least two different architectures to test on each of the selected datasets

You must take steps to ensure that

- the data is preprocessed as necessary
- different hyperparameters are tested
  - manual approaches are accepted
  - automated approaches are suggested
- that the final comparison is done on previously unseen data

Delivery

You must deliver a report and your code. This can be done with separate pdf and py files zipped togeter, or in a zipped notebook.

Oral Exam (30%)

The oral exam will have two parts

- ~5 min about their folder submission
- ~15 min about the curriculum (lecture slides and additional course resources)

Your part in this exam process is to observe and evaluate, and to pose questions (if desired).