

APPLIED MACHINE LEARNING ASSIGNMENT 1

Assignment 1 (CA1: 40%)

The objective of the assignment is to help you gain a better understanding of machine learning tasks of classification and regression.

Guidelines

1. You are to work on the problem set individually.
2. In this assignment, you will solve typical machine learning tasks and present your results (via Jupyter notebook and ppt slides) on your solutions to the tasks.
3. Submit your Jupyter notebook (comprising your codes and comments), and ppt slides. You should submit using the assignment link under the Assignment folder on POLITEMall.
4. You are also required to fill up and submit the Declaration of Academic Integrity. This document is available on POLITEMall.
5. The normal SP's academic policies on Copyright and Plagiarism applies. Please note that you are to cite all sources.

Submission Details

Deadline: **1st of June, 2025, 23:59**

Submit through: **POLITEMall**

Late Submission

50% of the marks will be deducted for assignments that are received within ONE (1) calendar day after the submission deadline. No marks will be given thereafter.

Exceptions to this policy will be given to students with valid LOA on medical or compassionate grounds. Students in such cases will need to inform the lecturer as soon as reasonably possible. Students are not to assume on their own that their deadline has been extended.

APPLIED MACHINE LEARNING ASSIGNMENT 1

PART A: CLASSIFICATION (50 marks)**Background**

This is a dataset collected from a manufacturing factory. The dataset contains various information on different machines. From the perspective of quality control, the factory wants to build a machine learning model to predict whether the machine is working properly or faulty.

Dataset

You are to use the dataset provided: `factory_data.csv`

Tasks

1. Write the code to solve the prediction task. Normally you would be using scikit-learn, but if you prefer to work with your own implementation of learning algorithms, or some other toolkit, that is fine.
2. Write a report detailing your implementation, your experiments and analysis in the Jupyter notebook (along with your python code and comments). In particular, we would like to know:
 - How is your prediction task defined? And what is the meaning of the output variable?
 - How do you represent your data as features?
 - Did you process the features in any way?
 - Did you bring in any additional sources of data?
 - How did you select which learning algorithms to use?
 - Did you try to tune the hyperparameters of the learning algorithm, and in that case how?
 - How do you evaluate the quality of your system?
 - How well does your system compare to a simple baseline?
 - Can you say anything about the errors that the system makes? For a classification task, you may consider a confusion matrix.
 - Is it possible to say something about which features the model considers important? (Whether this is possible depends on the type of classifier you are using.)
3. Create a set of slides with the highlights of your Jupyter notebook report. Explain the entire machine learning process you went through, data exploration, data cleaning, feature engineering, and model building and evaluation. Write your conclusions.

Submission requirements

1. Submit a zip file containing all the project files (Jupyter notebook), all data sets used, and the slides (pptx or pdf).
2. Submit online via the Assignment link.

Evaluation criteria:

APPLIED MACHINE LEARNING ASSIGNMENT 1

Background research (exploratory data analysis)	10 marks
Application of suitable algorithms	10 marks
Suitable evaluation of algorithms	10 marks
Demo/Presentation	10 marks
Quality of report (Jupyter notebook), and explanation of results	10 marks

APPLIED MACHINE LEARNING ASSIGNMENT 1

PART B: REGRESSION (50 marks)**Background**

This dataset contains house sale prices for King County, which includes Seattle. It includes homes sold between May 2014 and May 2015.

Dataset

You are to use the dataset.

<https://www.kaggle.com/harlfoxem/housesalesprediction>

Tasks

1. Write the code to solve the prediction task. Normally you would be using scikit-learn, but if you prefer to work with your own implementation of learning algorithms, or some other toolkit, that is fine.
2. Write a report detailing your implementation, your experiments and analysis in the Jupyter notebook (along with your python code and comments). In particular, we would like to know:
 - How is your prediction task defined? And what is the meaning of the output variable?
 - How do you represent your data as features?
 - Did you process the features in any way?
 - Did you bring in any additional sources of data?
 - How did you select which learning algorithms to use?
 - Did you try to tune the hyperparameters of the learning algorithm, and in that case how?
 - How do you evaluate the quality of your system?
 - How well does your system compare to a simple baseline?
 - Can you say anything about the errors that the system makes?
 - Is it possible to something about which features the model considers important?
3. Create a set of slides with the highlights of your Jupyter notebook report. Explain the entire machine learning process you went through, data exploration, data cleaning, feature engineering, and model building and evaluation. Write your conclusions.

APPLIED MACHINE LEARNING ASSIGNMENT 1

Submission requirements

1. Submit a zip file containing all the project files (Jupyter notebook), all data sets used, and the slides (pptx or pdf).
2. Submit online via the Assignment link.

Evaluation criteria:

Background research (exploratory data analysis)	10 marks
Application of suitable algorithms	10 marks
Suitable evaluation of algorithms	10 marks
Demo/Presentation	10 marks
Quality of report (Jupyter notebook), and explanation of results	10 marks

— End of Assignment —