Module 3 - Bonus Challenges

§M3: Logistic Regression

- Below are open-ended bonus challenges; solving them is not required but can help you better understand ML/AI in the context of engineering, and how to use them in practical cases.
- Bonus points earned in all homework assignments will be averaged (6 bonus points for each assignment) and then directly added to your final score to calculate your final letter grade.

Challenge 1.1. (Adapted from [1, 2]) As shown in Figure 1a, Assume a 3-member planar truss structure where each of the 3 links is grounded at one end and meets at a common joint at the other end. The common joint, free to move in the X-Y plane, is subjected to a load applied at an angle of 65. Each member of the truss is independently defined by two variables in Table 1. The detailed materials properties for each material type are listed in Table 2.

The performance of a 3-bar structure is evaluated to be good if it satisfy the constraints on the *mass*, and the x and y displacement of the loaded node, i.e., u_x and u_y, as shown in Equation 1.

$$mass \leq 3.5kg, \quad u_x \leq 1cm, \quad u_y \leq 1cm. \tag{1}$$

The goal of this task is to classify the performance of a design (either good or bad) based on the cross-sectional areas and material choices of its members. The attached file `'truss_6d.parquet'` contains a dataset with 10,000 designs. In this file, columns `'mat1'` to `'mat3'` represent the material choices for the three members, while columns `'area1'` to `'area3'` represent their respective cross-sectional areas. The column `'good'` indicates the performance of the design, with 0 for bad and 1 for good. Please complete this task according to the following requirements: (*6pts*)

1. Please use the built-in function `LogisticRegression` in `sklearn`.
2. Randomly split the dataset into 80% for training and 20% for testing, using the built-in function `train_test_split` in `sklearn`.Train your model on the training dataset and evaluate it on the testing set.

3. Submit your Jupyter notebook file with necessary comments, reporting the score for the classification.

4. Based on the trained model, please try to analyze the importance of each design variable in achieving a good 3-bar design.

Below are some useful tips:

1. `read_parquet` in `pandas` library is a useful tool to read the dataset file.

2. You could perform feature engineering and decide what representation to use for the design variables, rather than being constrained to the original input/output features.

3. To assess variable importance, explore resources on the interpretability and explainability of linear models and other machine learning techniques.

4. Refer to the documentation of `LogisticRegression` for guidance.

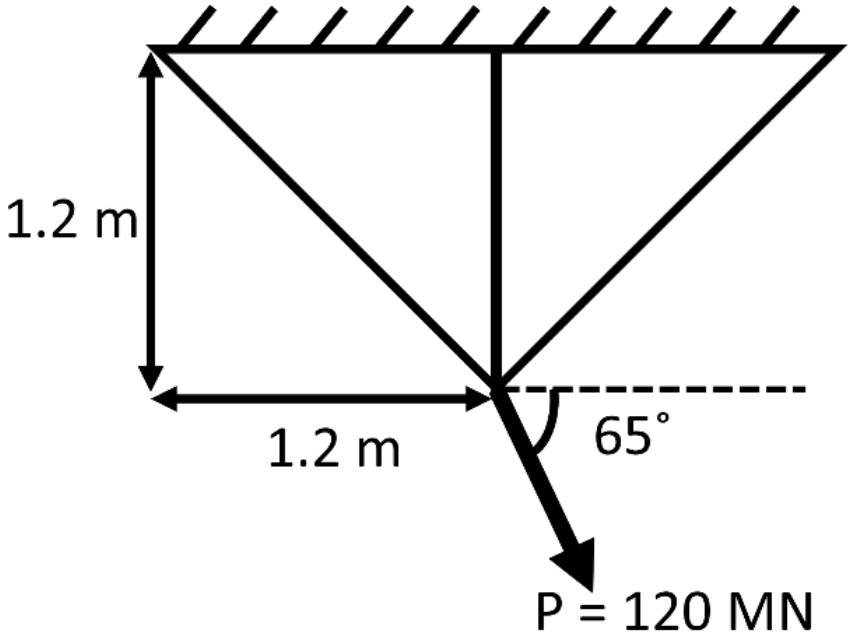

Figure 1: 3-bar truss problem schematic [2]

Table 1: Bar-wise Design Variables Description

Design variables	Notation	Range
Material type	`mat`	$\{1, 2, 3\}$
Area	`area`	$[1 \times 10^{-5}, 3 \times 10^{-3}]m^2$

Table 2: Materials Properties

Material type	Material name	Young's Modulus (GPa)	Mass Density (kg/m^3)
1	Aluminum7050	71.7	2800
2	Titanium	114	4430
3	Stainless Steel	200	8000

References

[1] Cyril Picard and Faez Ahmed. *Engineering Design Benchmark Problems for Classification Algorithms*. Version V1. 2024. DOI: 10.7910/DVN/ZRHXNY. URL: https://doi.org/10.7910/DVN/ZRHXNY.

[2] Conner Sharpe et al. "A comparative evaluation of supervised machine learning classification techniques for engineering design applications". In: *Journal of Mechanical Design* 141.12 (2019), p. 121404.