

**American University of Sharjah**

**College of Engineering**

NGN 112 Introduction to Artificial Intelligence and Data Science

Sections 12 | Spring 2024

**Project announcement date:** 25/March/2024

**Project submission date:** 9/May/2024, Thursday, before 11:59 pm

**Start date of student presentations:** the last week of the semester, starting 13/May/2024

**Project rules:**

1. This is a team-based project, each project consists of 4 or 5 students. It is your responsibility to form a team.
2. The project entails working on a regression problem and submitting a detailed report (details below) and a PowerPoint presentation.
3. Your professor will email you the presentation orders. However, you need to be ready for presenting starting on the 13<sup>th</sup> of May.

**Submission:**

You are required to submit the following:

1. A report that contains all project requirements as specified below including:
  - a. Description of the dataset used
  - b. Python code
  - c. Graphical summaries
  - d. Numerical summaries
  - e. Regression results
2. PowerPoint slides to present your work to the class.
3. One student out of each team is required to make a submission on ilearn.

Your report must contain a cover page with course information, semester, professor and names and IDs of the team members.

## **The Python code:**

The Python code you submit in your report must be similar to what was introduced in class.

If you include code from other resources then you must write the source of the information as a comment in your code.

## **Project details:**

Your professor will assign to you datasets that you can use for your project. These include regression datasets.

Once you have access to the dataset, you are required to perform the following:

### **1. Descriptive tasks:**

- a. Write a description on the dataset using your own words
- b. Provide numerical summaries for the feature variables. This includes but is not limited to measures of center and spread.
- c. Provide graphical summaries of the feature variables. This includes but not limited to, box plots, histograms, pair plots and heat maps.
- d. Then comment on numerical and graphical summaries generated. In other words, what are your observations?

### **2. Preprocessing tasks:**

You are requested to repeat all the experiments in Section 4 below using the following normalization techniques:

- a. Normalize the feature variables using z-scores
- b. Normalize the feature variables using min-max

### **3. Data split into train and test sets:**

You are requested to repeat all the experiments in Section 4 below using three splits of the data. You can do that by fixing the `random_state` parameter to 1, 20 and 40. This will generate 3 different train and test sets. Then:

- a. Report the  $R^2$  and MSE for each test split, as described in Section 4.
- b. Report the average  $R^2$  and MSE for all test splits, as described in Section 4.

#### 4. Regression:

You need to use all of the following regression techniques:

- a. Linear regression
- b. SVM with linear kernel
- c. SVM with polynomial kernel
- d. SVM with RBF kernel
- e. Neural Networks

In summary, you need to carry out the experiments in the following manner:

**Loop** for both normalization techniques of Section 2 of the project details

**Loop** for each of the three data splits of Section 3 of the project details

**Loop** for each classifier or regressor of Section 4 of the project details

So the total number of experiments is:

$$2(\text{normalizations}) \times 3(\text{data splits}) \times 5(\text{regressors}) = 30 \text{ experiments}$$

Report the results of these experiments as 4 Tables; one for each of the three splits and one for the average of the three splits. Each Table shall have 5 rows (one for each regressor) and two columns (one for each normalization scheme). Each cell in the table will have two values; MSE and the R2 score. An example Table is shown below.

	Min-Max Normalization		Standard Scaling	
	MSE	R2-Score	MSE	R2-Score
Linear Regression				
Linear SVM				
Poly. SVM				
RBF SVM				
Neural Networks				

**Note on academic dishonesty:**

AUS is strict about plagiarism and academic dishonesty. The work that you submit must be developed by you and your team only. Otherwise, an academic dishonesty case will be reported to the dean's office and the penalty will range from receiving a zero in the project to getting an XF in the course.

Your professor will examine the code that you submitted and will ask you to explain it.