NTHU Introduction to ML 2024
Lab 1
Grip Force Prediction using Regression

Jui-Yun Su

Po-Chih Kuo

Introduction

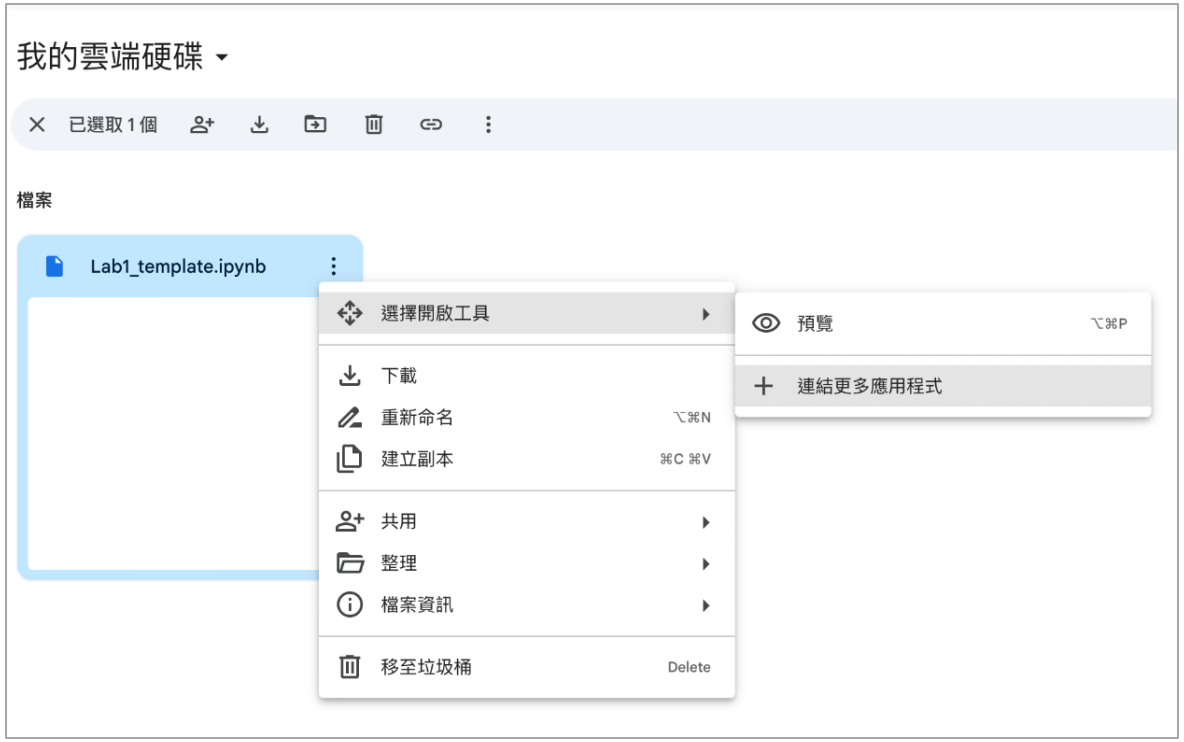
- Understanding human body performance is critical for various applications, from health monitoring to sports science. Factors such as age, gender, height, weight, body fat percentage, and blood pressure can all influence physical capabilities, including grip strength.
- In this lab, students will need to develop predictive models to accurately estimate one's grip force based on the body data.

Google Colab

- We use .ipynb file
- If you don’t have the environment to run .ipynb files, we recommend you to use Google Colab
- You can also use other tools to compile the .ipynb files, such as Jupyer Notebook or VScode

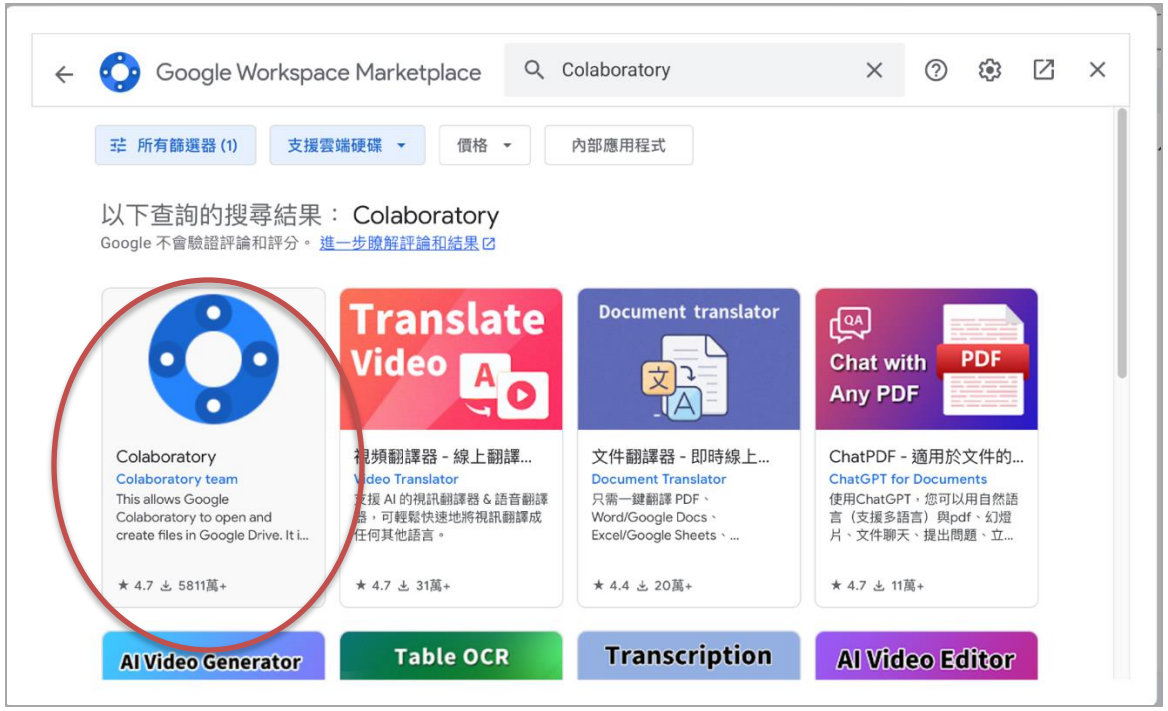
Google Colab

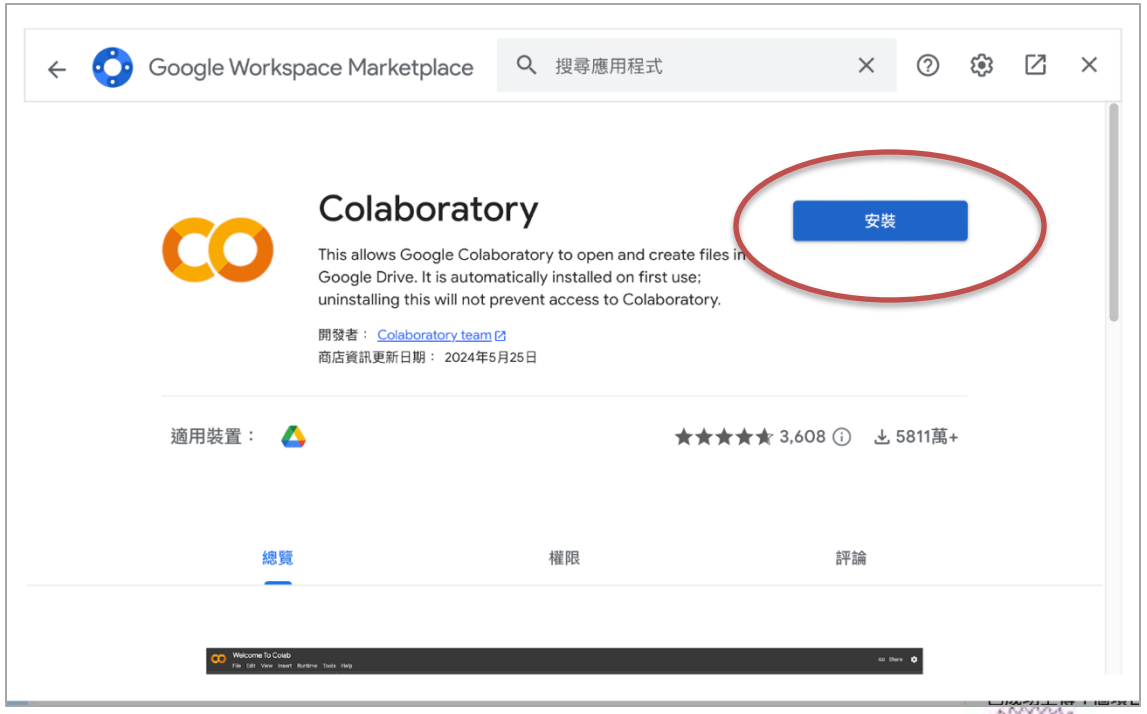
1. Upload the file to google drive
2. Right click -> Open with -> Connect more apps



Google Colab

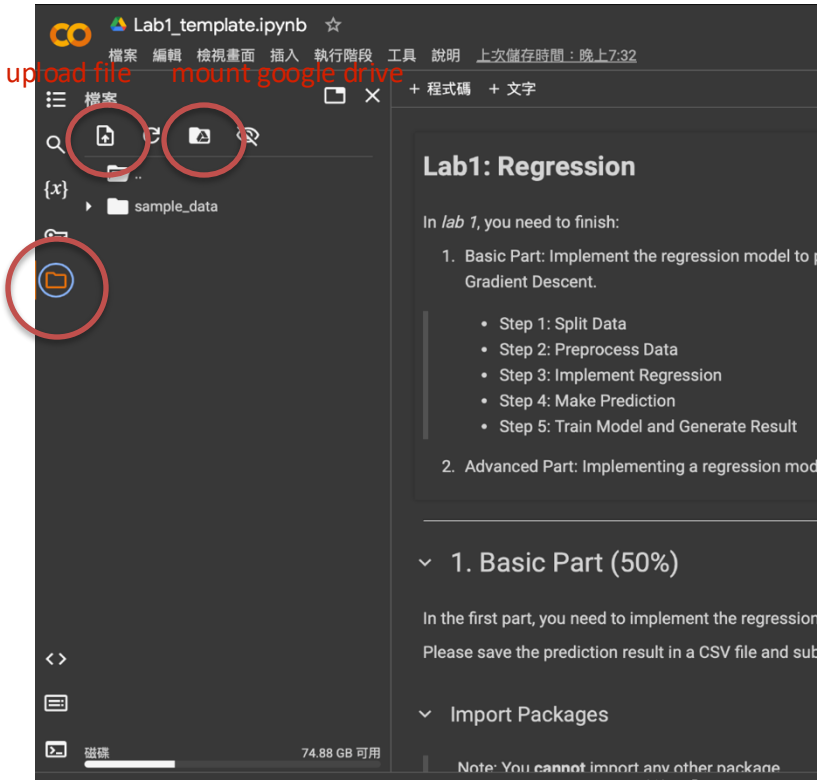
3. Search for “Colaboratory” , and install. Now you can use Colab to open the .ipynb file

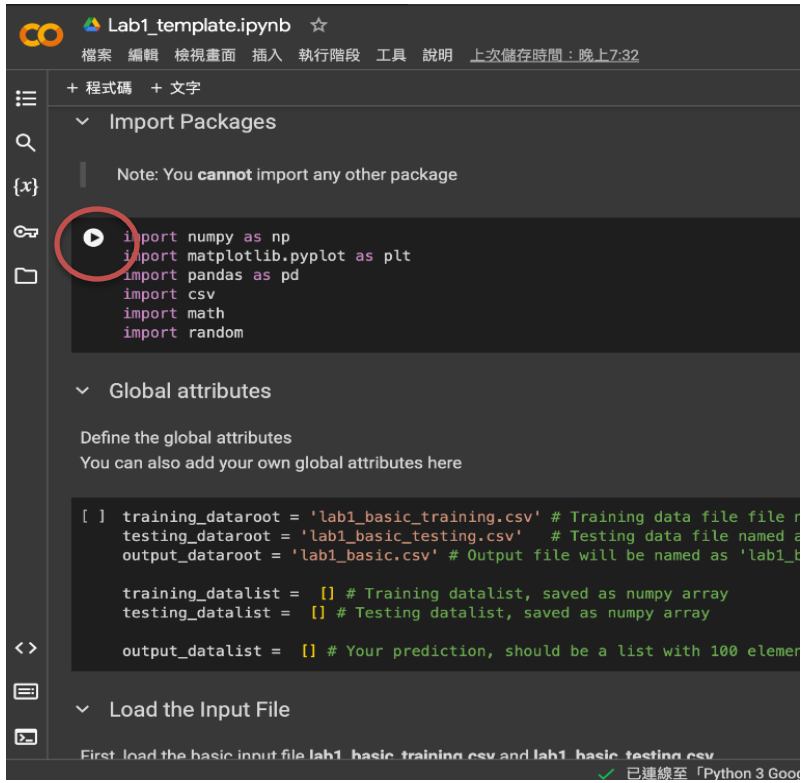




Google Colab

4. After opening the file, you can upload other files (e.g. datasets) so you can access them from your program. You can also mount your google drive
5. Click the “run” button to execute your code





Dataset

- Body measurement data
- Parts of data have been manipulated
  - Outliers and missing values(‘nan’) were added
  - Only in training dataset
- Attributes
  - age
  - gender
  - height
  - weight
  - bodyFat
  - diastolic
  - systolic
  - gripForce

Goal

- Predict grip force based on an individual's weight
- Combine other variables such as age, gender, and height to improve the prediction of grip force
- Implement a regression model to make these predictions, and preprocess/split the data for model training and testing.

Grading Policy

Item	Score
Basic Implementation	50%
Advanced Implementation	45%
Report	5%

Basic Implementation (50%)

- Given 10,000 subjects' body weight and grip force
- Implement gradient descent to build a regression model using body weight as the input variable to predict grip force
- Please use the provided file as your input and submit your predictions to Kaggle (link in Page 19)

Basic Grading Policy

- Baseline – 50%
  - Submit the answer (.csv) to Kaggle **ML2024-Lab1-BasicPart**
  - Get all if MAPE <= 25%

The Evaluation Metric

- MAPE (Mean absolute percentage error)

$$\text{MAPE} = \frac{100\%}{N} \sum_{i=1}^{N} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

- For example:
  - The value you predicted:
  - $\hat{y}$ = [592, 486, 538, 689, 752, 841, 491]
  - Ground Truth :
  - y = [491, 584, 541, 599, 615, 741, 512]
  - MAPE = 1/7 *0.928 = 0.1326 = 13.26%

The Evaluation Metric

- MAPE (Mean absolute percentage error):

$$\text{MAPE} = \frac{100\%}{N} \sum_{i=1}^{N} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

- For example:
  - The value you predicted:
  - $\hat{y}$ = [592, 486, 538, 689, 752, 841, 491]
  - Ground Truth :
  - y = [491, 584, 541, 599, 615, 741, 512]
  - MAPE = 1/7 *0.928 = 0.1326 = 13.26%
- We would evaluate your assignment by the average MAPE across all cities

HMI

Advanced Implementation (45%)

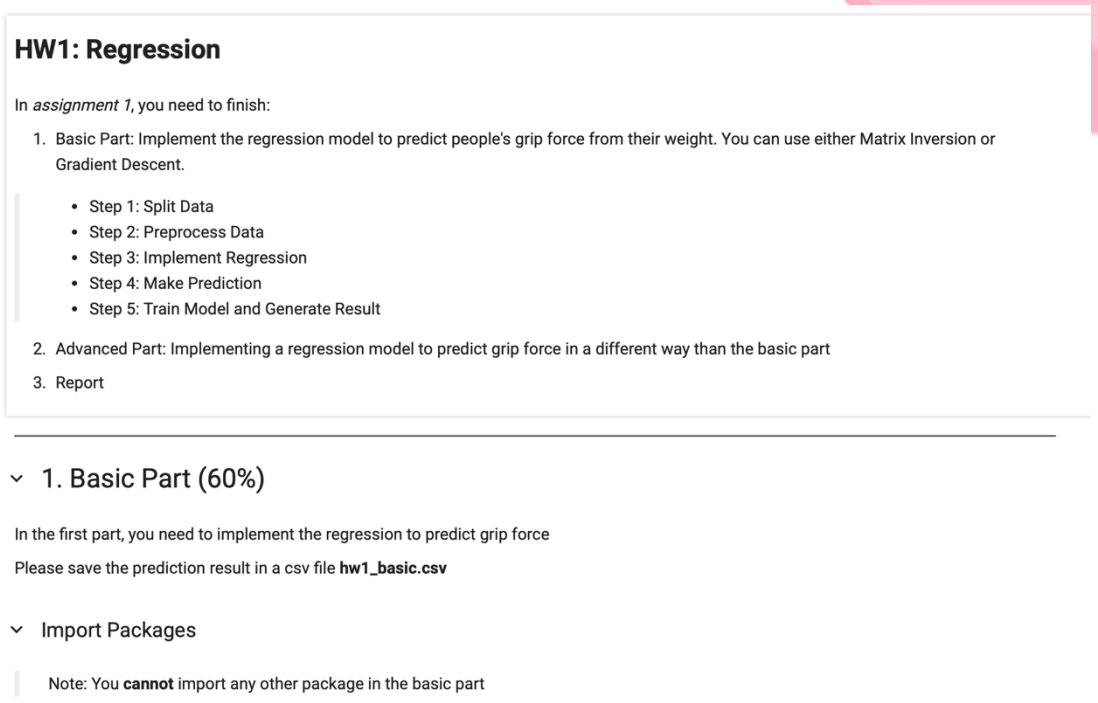
- Combine with other conditions or modify the approach from the basic model to improve your predictions
- You may complete this part in any way you prefer, including using matrix inverse, but do not import additional modules.
- Using only *body weight* (as done in the basic part) will not earn points
- Please use the provided file as your input and submit your predictions to Kaggle.

Advanced Grading Policy

- Baseline – 35% (There are 3 baselines to beat)
  - MAPE < 22% => 20% (1st baseline)
  - MAPE < 20% => 10% (2nd baseline)
  - MAPE < 17% => 5% (3rd baseline)
- Ranking –10%
  - Compete your MAPE with the whole class
- Submit the answer (.csv) to Kaggle **ML2024-Lab1-AdvancedPart**

Template

- You must use the given file “Lab1_template.ipynb” to build the model
- Except for the imported packages in the template, you may not use any additional packages in both the basic and advanced parts.



HW1: Regression

In *assignment 1*, you need to finish:

1. Basic Part: Implement the regression model to predict people's grip force from their weight. You can use either Matrix Inversion or Gradient Descent.
   - Step 1: Split Data
   - Step 2: Preprocess Data
   - Step 3: Implement Regression
   - Step 4: Make Prediction
   - Step 5: Train Model and Generate Result
2. Advanced Part: Implementing a regression model to predict grip force in a different way than the basic part
3. Report

---

1. Basic Part (60%)

In the first part, you need to implement the regression to predict grip force

Please save the prediction result in a csv file **hw1_basic.csv**

Import Packages

> Note: You **cannot** import any other package in the basic part

Basic Input File Format

- There will be two input files:

1. “lab1_basic_training.csv
    - Each row has weight and gripForce
    - Contains (10000 +1) rows

1. “lab1_basic_testing.csv”
    - Doesn’t have the gripForce column
    - Contains (100+1) rows

lab1_basic_training

weight	gripForce
55.4	32.8
53.6	39.4
78.4	52.6
50.6	33.1
4.8	26.6
58.3	44.6
53.0	32.3
67.8	52.3
81.9	37.6
76.0	62.4
70.7	50.0
66.0	49.0
59.6	43.4
77.3	49.8
48.6	32.8
72.9	55.0

lab1_basic_testing

weight
53.6
51.54
79.7
56.68
68.5
64.8
57.4
63.5
63.3
61.4
69.9
92.5
83.1
51.3
64.5
64.2

Advanced Input File Format

- There will be two input files:

1. “lab1_advanced_training.csv”
   - Each row has “age, gender, height, weight, body fat, diastolic, systolic, and gripForce”
   - Contains (10000 +1) rows

1. “lab1_advanced_testing.csv”
   - Doesn’t have the gripForce column
   - Contains (3000+1) rows

lab1_advanced_training

age	gender	height	weight	bodyFat	diastolic	systolic	gripForce
21.0	nan	169.2	55.4	23.6	67.0	115.0	32.8
31.0	F	162.7	53.6	29.3	78.0	142.0	39.4
41.0	M	176.0	78.4	22.8	93.0	153.0	52.6
49.0	F	156.1	50.6	16.3	95.0	147.0	33.1
49.0	F	155.6	4.8	23.6	61.0	100.0	26.6
23.0	M	172.5	58.3	6.5	75.0	118.0	44.6
21.0	F	160.6	53.0	23.4	68.0	116.0	32.3
40.0	M	172.0	67.8	20.7	84.0	132.0	52.3
50.0	M	170.7	81.9	29.5	87.0	144.0	37.6
45.0	M	180.0	76.0	29.0	84.0	134.0	62.4

lab1_advanced_testing

age	gender	height	weight	bodyFat	diastolic	systolic
39.0	M	180.1	73.8	28.1	80.0	133.0
27.0	M	174.3	79.6	20.5	60.0	136.0
36.0	F	168.6	59.1	31.4	74.0	106.0
49.0	M	163.0	78.8	37.5	73.0	123.0
22.0	F	167.9	66.7	33.1	87.0	136.0
21.0	M	183.0	80.1	17.6	72.0	138.0
21.0	F	162.6	64.3	36.0	62.0	117.0
35.0	M	170.7	74.1	21.7	86.0	145.0
28.0	F	157.9	52.3	29.8	84.0	118.0
62.0	M	169.0	70.7	30.5	96.0	146.0

Output File Format

- The prediction of both basic and advanced you submit to Kaggle must follow this format
- File names are not restricted
- Prediction for basic part should have 100+1 rows, and predictions for advanced part should have 3000+1 rows
- The first row is the header [“Id”, “gripForce”]
- Id starts from 0, and gripForce is the predicted answer
- Please make sure your model can correctly output this format of csv file

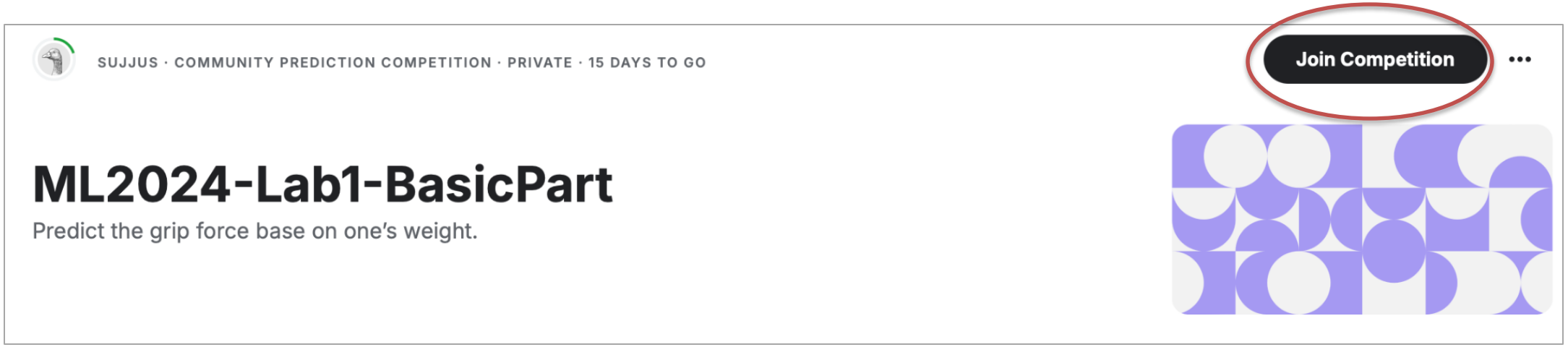
Id	gripForce
0	34
1	45
2	56
3	67
4	78
5	89
6	86
7	57
8	47
9	46
10	87

Kaggle

- We’ve created two competitions for Basic & Advanced part respectively.
- Basic part link: https://www.kaggle.com/t/94e9d46e5923438ebc4ef5b4d6bb3270
  - You can check if you pass the baseline directly.(50%)
- Advanced part link: https://www.kaggle.com/t/f833b5c488d14f8581ce76fe4ec68228
  - In advanced part, we split data into public & private two parts.
  - In public score, we have three baselines that you can also check if you get the points.(35%)
  - Private score is for ranking. The private score would be revealed after deadline.(10%)

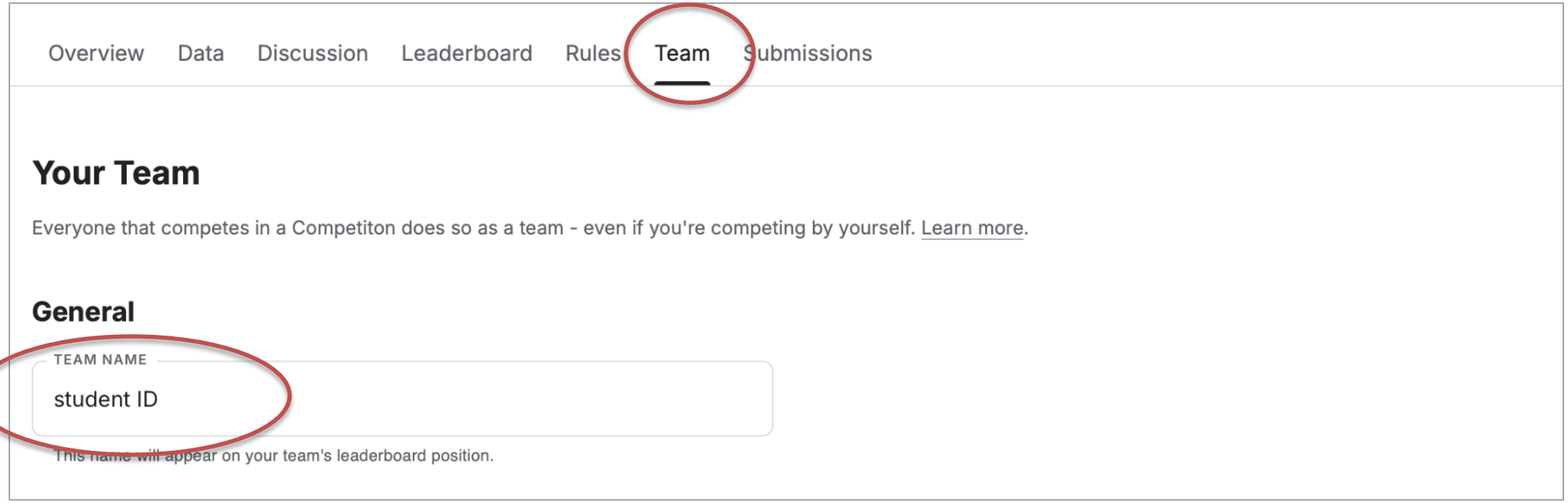
Kaggle

- Please register your account.
- Click the ‘Join competition’ button to join.



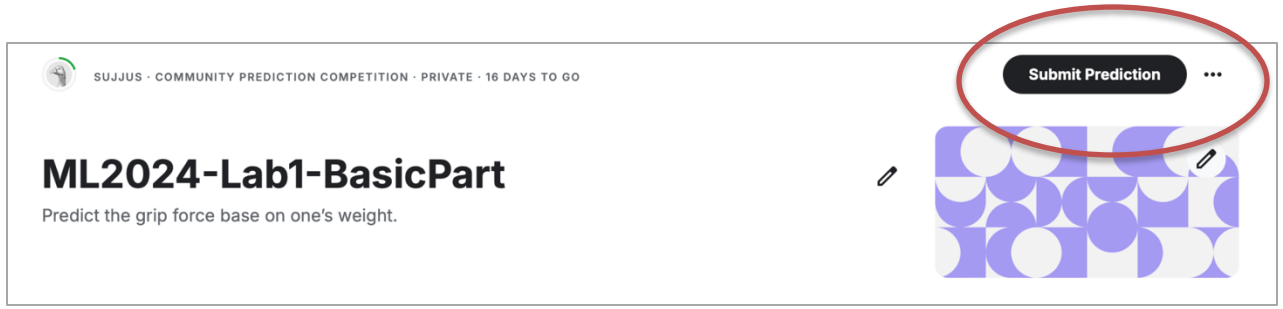
Kaggle

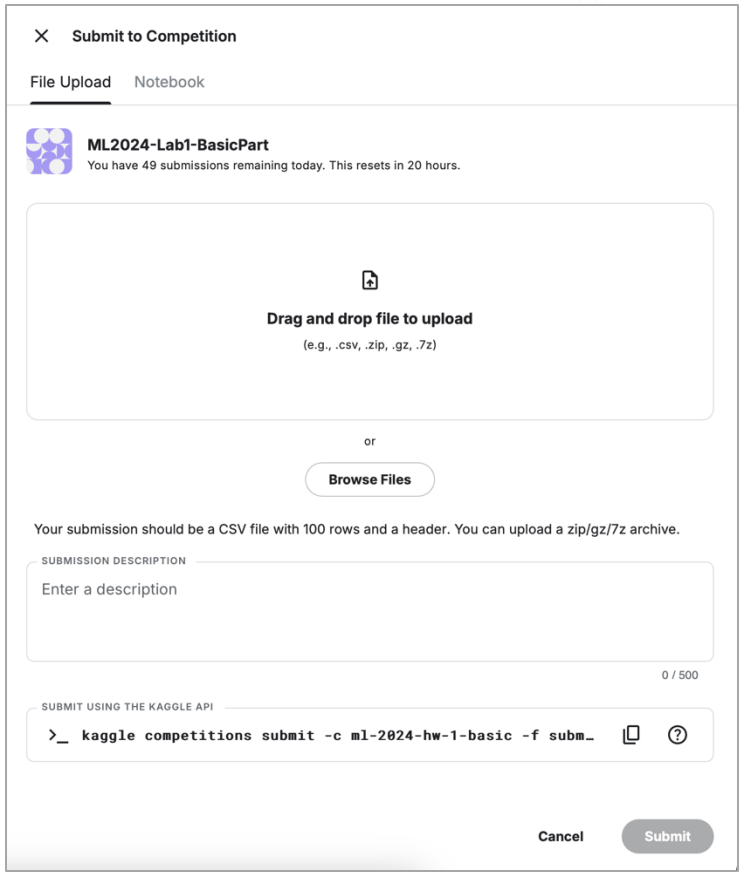
- After joining the competition, you should change your team name (each student is a team) to your **student ID.**



Kaggle

- Click the “Submit Prediction” button and upload your prediction file, then you will see your MAPE score

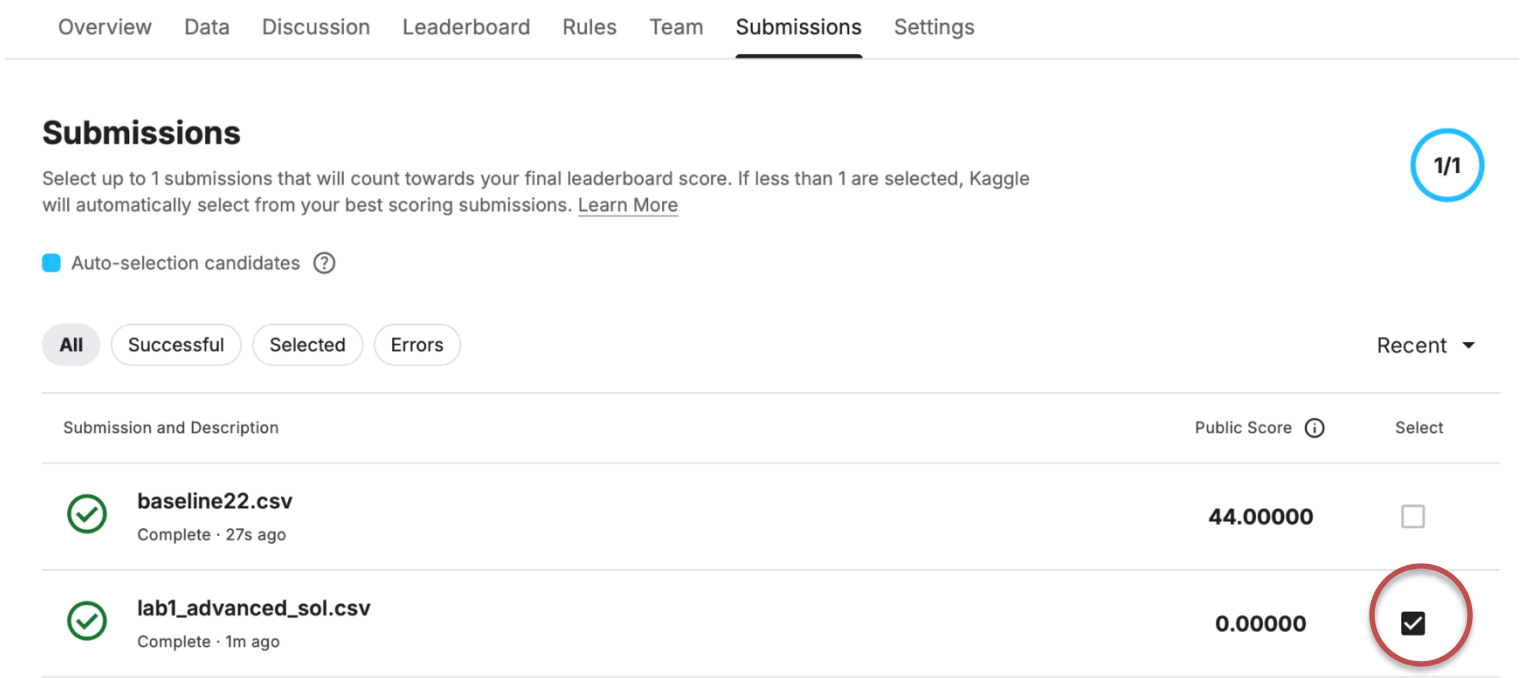




- You can submit 50 times per day.

Kaggle (optional)

- For advanced part, you can choose one of your predictions to compete with others
- If you don’t make a selection, the system will automatically choose the one with the best public score



Report

- Named as “**Lab1_report.pdf**”
- Write down your **regression equation** in basic part (1%)
- Briefly describe the **variables** and the **regression equation** you used in the advanced part (2%)
  - No point would be given for the advanced part if you do not clearly point out the difference between the basic part and the advanced part
- Briefly describe the difficulty you encountered (1%)
- Summarize how you solve the difficulty and your reflections (1%)
- No more than one page

Lab 1 Requirement

- Do it individually! Not as a team! (The team is for final project)
- Announce date: 2024/9/12
- Deadline: 2024/9/26 23:59 (Late submission is not allowed!)
- Submit the answers (csv) to corresponding Kaggle competition.
    - **ML2024-Lab1-BasicPart**
    - **ML2024-Lab1-AdvancedPart**
- Hand in following files to eeclass in the following format (Do not compressed!)
    - Lab1.ipynb
    - Lab1_report.pdf
- 5 Points would be deducted if your submission format is incorrect
- Lab 1 would be covered on the exam next time.

Penalty

- 0 points if any of the following conditions happened
  - Plagiarism
  - Late submission
  - Not using a template or importing any other packages in the basic part
  - No submission record on Kaggle
  - Your submission was not generated by your code

Questions?

- TA: Jui-Yun Su (ss113062515@gapp.nthu.edu.tw)
- Please do not ask for debugging.

When part of your code doesn't work so you replace it with something from the internet