

TASK 0

Ungraded Dummy Task

You are in the group 🧑‍🤝‍🧑 Amoebas consisting of 🧑 dsumner (dsumner@student.ethz.ch (mailto://dsumner@student.ethz.ch)), 🧑 lkasak (lkasak@student.ethz.ch (mailto://lkasak@student.ethz.ch)) and 🧑 tfang (tfang@student.ethz.ch (mailto://tfang@student.ethz.ch)).

1. READ THE TASK DESCRIPTION

2. SUBMIT SOLUTIONS

3. HAND IN FINAL SOLUTION

1. TASK DESCRIPTION

This is a dummy task to prepare you for the two mandatory tasks later in the semester. After this task, you should know how to:

- register for a task,
- form a team of up to three students,
- read the task and data description,
- download the provided data sets,
- make a submission with your predictions and source code,
- see how your submission scores with regards to the baselines and the other students, and
- hand in the task by both choosing which submission should be evaluated and writing an **individual** task description.

While this task is not mandatory and the problem itself is easy, we strongly recommend that you complete it as we will be **unable to extend** the deadlines for the mandatory tasks if you experience issues related to the points above.

TASK

This task is a trivial form of regression: Your goal is to predict a value y based on a vector x . While the exact relationship is usually not known, in this task, y is the mean of x . You may verify this on the provided training set. Your task is to make predictions for y on the provided test set.

DATA DESCRIPTION

[Download handout \(/static/task0_sl19dl.zip\)](/static/task0_sl19dl.zip)

In the handout for this project, you will find the the following files:

- **train.csv** - the training set
- **test.csv** - the test set (make predictions based on this file)
- **sample.zip** - a sample submission file in the correct format

Each line in train.csv is one data instance indexed by an Id. It consists of one double for y and 10 doubles for the vector x_1 - x_{10} :

```
Id,y,x1,x2,x3,x4,x5,x6,x7,x8,x9,x10
0,738.02,1764.05,400.15,978.73,2240.89,1867.55,-977.27,950.08,-151.35,-103.21,410.
...
```

The test set file (test.csv) has the same structure except that the column for y is omitted:

```
Id,x1,x2,x3,x4,x5,x6,x7,x8,x9,x10
10000,-483.79,1288.05,-129.87,-198.07,-334.48,-391.44,-612.40,-676.52,1327.22,-448
...
```

For your convenience, we further provide a sample submission file (which is zipped and has the correct structure as well as the correct CSV file name **submission.csv**):

```
,Id,y
0,10000,0.0
1,10001,1.0
2,10002,2.0
```

Note that, for each prediction, you need to include the same sample id (in the Id column) as specified in test.csv.

SUBMISSION FORMAT

For every data instance in the test set, submission files should contain three columns: (a row index without a column column), Id and y where y should be a double with your prediction.

The file should contain a header and have the following format:

```
,Id,y
0,10000,0.0
1,10001,1.0
2,10002,2.0
..
```

Please zip your submission file (as this will be needed for later projects) before submission and name the ZIP file as LastName_FirstName_Project0.zip. Please ensure that your submission file (within the ZIP) is named submission.csv - while this is not needed for this task, you will have to submit multiple CSV files for later projects. Thus, we need you to match the names exactly such that we can tell the files apart.

Please keep in mind that, as a group, you have a limited number of submissions as stated on the submissions page.

EVALUATION

The evaluation metric for this task is the **Root Mean Squared Error** which is the square root of the mean/average of the square of all of the error.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

How to compute it in Python:

```
from sklearn.metrics import mean_squared_error
RMSE = mean_squared_error(y, y_pred)**0.5
```

GRADING

This task is not mandatory to solve. However, in order to prepare you for the mandatory tasks, we provide you with the following information on how the subsequent tasks, will be evaluated.

We provide you with a dataset for which you have to make prediction or compute other quantities. For each of your submissions, a score is computed. When handing in the task, you need to select which of your submissions will get evaluated and provide a short description of your approach as set out in the project description. This has to be done **individually by each member** of the team. We will then compare your selected submission to our baseline. This project task is graded with either **pass or fail**. To pass the project, you need to achieve a better score than the baseline. Passing the projects is a mandatory part of the course. You will have to pass both Project 1 and Project 2 (unless you choose to do a paper presentation, in which case you only have to pass one of the two projects). We will consider the code and your solution description for the pass/fail decision. The following **non-binding** guidelines provides you with an idea of what is expected to pass the project. If you hand in a properly-written description, your source code is runnable and reproduces your predictions, and your submission performs better than the baseline, you can expect to have passed the assignment.

⚠ Make sure that you properly hand in the task, otherwise you may obtain zero points for this task.

FREQUENTLY ASKED QUESTIONS

🕒 WHICH PROGRAMMING LANGUAGE AM I SUPPOSED TO USE? WHAT TOOLS AM I ALLOWED TO USE?

You are free to choose any programming language and use any software library. However, **we strongly encourage you to use Python**. You can use publicly available code, but you should specify the source as a comment in your code.

🕒 IN WHAT FORMAT SHOULD I SUBMIT THE CODE?

You can submit it as a single file (main.py, etc.; you can compress multiple files into a .zip) having max. size of 1 MB. If you submit a zip, please make sure to name your main file as *main.py* (possibly with other extension corresponding to your chosen programming language).

🕒 IN WHAT FORMAT SHOULD I SUBMIT MY SUBMISSION FILES?

Please submit your submission files as one zip file named LastName_FirstName_ProjectX.zip. The zip should contain all csv files specified in the detailed project description for each respective project.

🕒 WILL YOU CHECK / RUN MY CODE?

We will check your code and compare it with other submissions. We also reserve the right to run your code. Please make sure that your code is runnable and your predictions are reproducible (fix the random seeds, etc.). Provide a readme if necessary (e.g., for installing additional libraries).

🕒 DO I NEED TO PROVIDE A REQUIREMENTS.TXT FILE?

Yes, please provide a requirements.txt file (or equivalent depending on programming language), to ensure that we can reproduce your results as easily as possible.

🕒 SHOULD I INCLUDE THE DATA IN THE SUBMISSION?

No. You can assume the data will be available under the path that you specify in your code.

🕒 CAN YOU HELP ME SOLVE THE TASK? CAN YOU GIVE ME A HINT?

As the tasks are a mandatory part of the class, **we cannot help you solve them**. However, feel free to ask general questions about the course material during or after the exercise sessions.

🕒 CAN YOU GIVE ME A DEADLINE EXTENSION?

⚠️ We do not grant any deadline extensions!