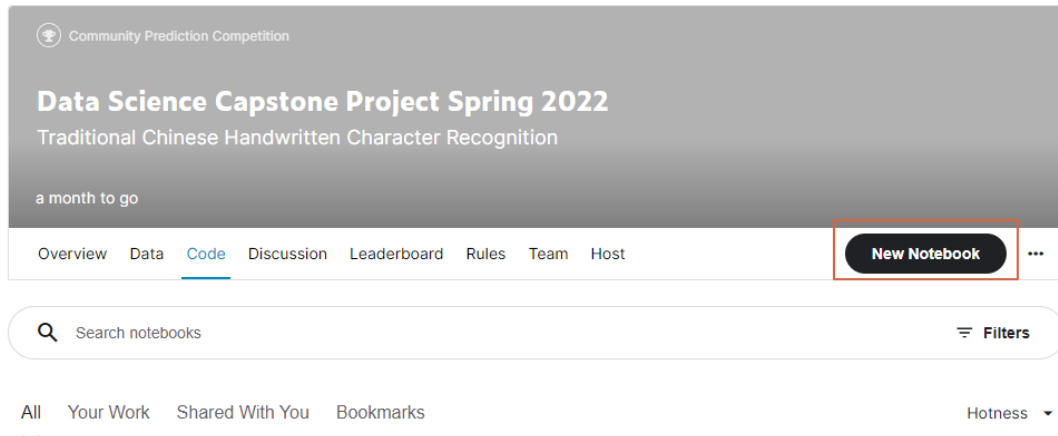


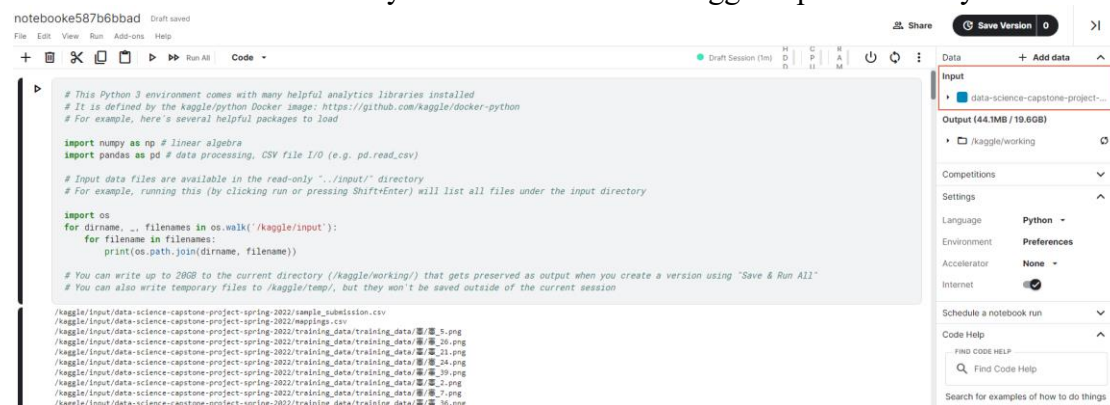
Kaggle tutorial

1. How to get the data?

Register [Kaggle](#) first and on the [competition website](#), click “New Notebook”:



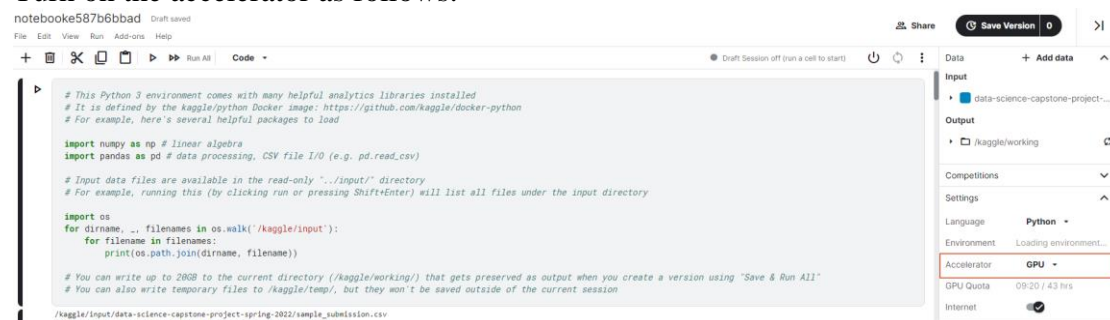
The data will be automatically downloaded to the “/kaggle/input” directory:



Notice that all the intermediate output will be stored into default location “/kaggle/working”. You can directly download the data in this directory by the GUI.

2. How to use GPU?

Turn on the accelerator as follows:

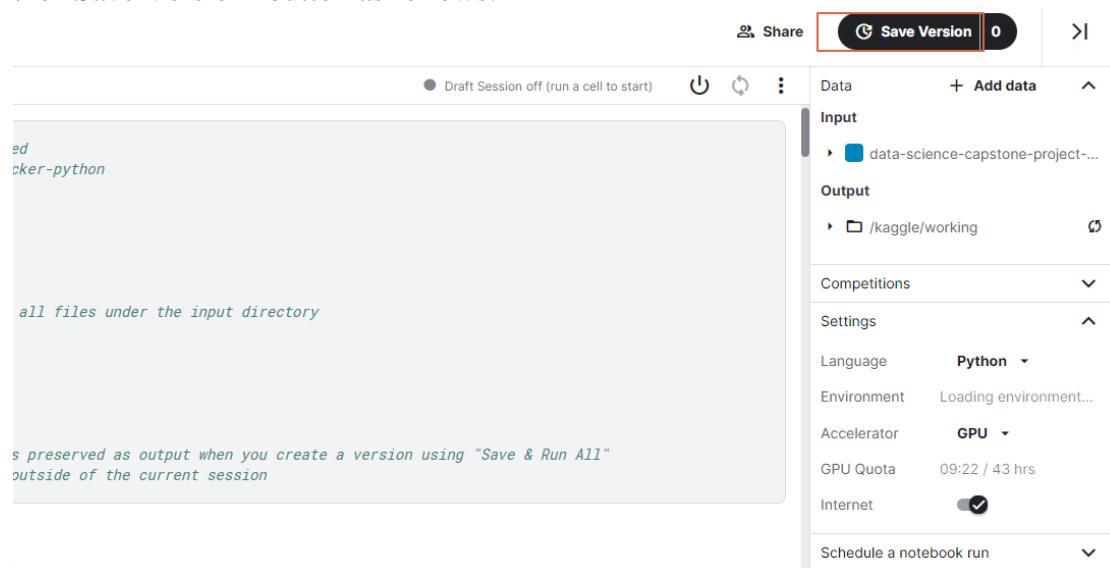


The Resources are listed below

- Kaggle GPU: 16G NVIDIA TESLA P100
- <https://www.kaggle.com/docs/efficient-gpu-usage>
- Limited to 30+ hrs/week depending on usage.
- Limited to 12hrs/run

3. Run the code in the background

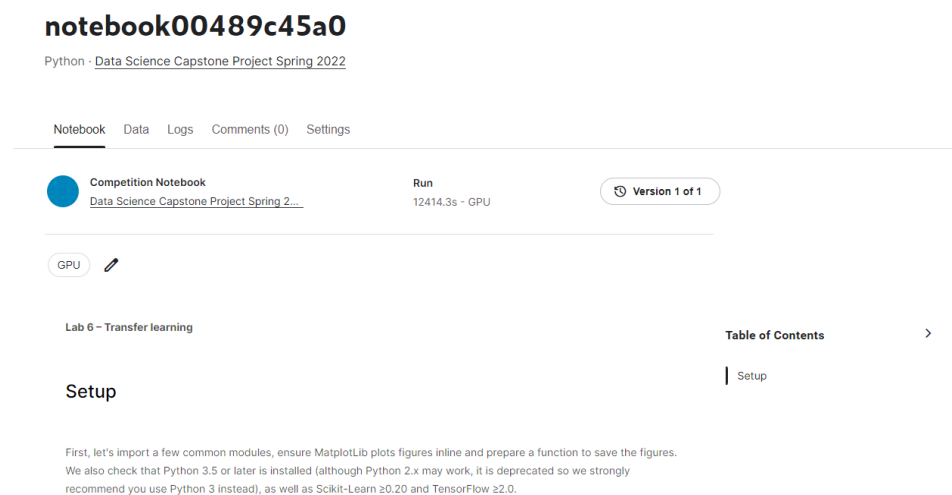
You can also run the code in the background with Kaggle. Firstly, make sure your code is bug-free, as any error in any code block would result in early stopping. Click the “Save Version” button as follows:



Notice that the output is limited to 20G and the max run time is limited to 12hrs. You can then see the results by clicking the version number:



The log will be shown in the Notebook panel



And your output can be accessed through the “Data” panel, where you can download your data. You can also save your model as a new dataset and import your new dataset into the notebook via “Add data” so that you can modify your code to load your checkpoint:

notebook00489c45a0


Python · Data Science Capstone Project Spring 2022

Notebook Data Logs Comments (0) Settings

Data

ev2_fine_tuning_0319.keras (56.21 MB) Submit ↓ >

Unable to show preview



Previews for binary data are not supported

Input (12.87 MB)

- Data Sources
 - Data Science Capstone ...
Output

 - ev2_fine_tuning_0319.keras
 - my_keras_model_0319.keras

Download All+ New Dataset+ New Notebook

Colab tutorial

1. How to get the data?

Register [Kaggle](#) first. In the “Data” panel of the [competition website](#), click Download all

Overview Data Code Discussion Leaderboard Rules Team Host My Submissions Submit Predictions ...


Data Explorer
12.87 MB

- test_data
- training_data
- mappings.csv**
- sample_submission.csv

mappings.csv (6.31 kB) ↓ ⌕
Detail Compact Column 2 of 2 columns ▼

About this file ✎

This file does not have a description yet.

Character	# Labels
800 unique values	
丁	0
三	1
上	2
不	3
世	4
丞	5
中	6
主	7
久	8
之	9
事	10
于	11

Summary

- 40.5k files
- 4 columns

Download All

Store your data into your google drive and get the sharable link. Execute the following commands

```
!pip install --upgrade --no-cache-dir gdown
!gdown --id aaaa # Replace aaaa with the string between d/ and /view in your link https://drive.google.com/file/d/aaaa/view?usp=sharing
!unzip -qq data-science-capstone-project-spring-2022.zip
```

Remember DO NOT store input data in your drive and load from there. The input/output is very slow (store at ./ instead). Your output data should be stored in your google drive so that it can be accessed next time.

2. For other tips in using Colab refer to the Colab tutorial at our course website <https://phonchi.github.io/nsysu-math604/schedule/>.

Tips

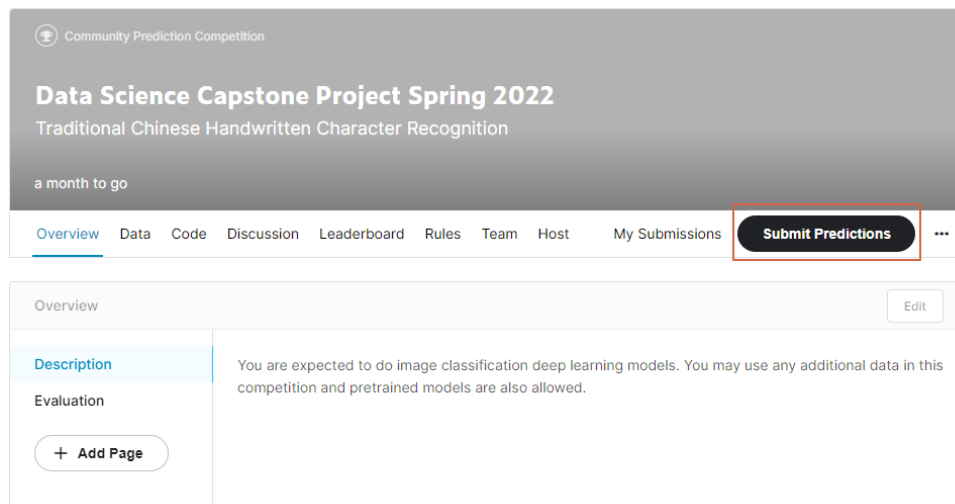
1. You can use the following code snippet to prepare your *submission.csv*:

```
test_dataset = image_dataset_from_directory(
    "test_data",
    shuffle=False,
    image_size=(50, 50),
    batch_size=1)

paths = test_dataset.file_paths
Ids = []
for x in paths:
    Ids.append(x.split("/")[-1])

df = pd.DataFrame()
df["Id"] = Ids
df["Predicted"] = predictions #It should be integer not float
df.to_csv("submission.csv", index = False)
```

Upload your submission to get an estimate of your score



2. Some tips that may help you to improve your accuracy
 - Data augmentation
 - Use transfer learning
 - Learning rate scheduling
 - Change optimizer
 - Think about how to deal with data that does not belong to the 800 common words
 - Use ensemble
 - Collect or generate more data