Instructions/Guides

Sunday, October 15, 2023 4:14 PM

Instructions to use Kaggle command line tool on remote math account

First, download an API key

On your laptop:

- Create an account at kaggle.com
- Click your user picture in the top right to open the menu
- Click 'Your Profile'
- Go to the 'Account' tab
- Scroll to 'API' and click the 'Create New Token' button
- An API key named 'kaggle.json' will be downloaded, most likely to the 'Downloads' folder on your laptop

Next, upload the API key

In the following instructions, type your LSU username (ex: mtige3) in place of $USER:

- Log on to the remote account via SSH (ex: ssh $USER@chaos8.math.lsu.edu)
- You will automatically be in your user account: /U1/accounts/$USER
- Make the .kaggle directory with this command: mkdir .kaggle
- This will create a folder with address /U1/accounts/$USER/.kaggle
- Open another terminal (or exit your SSH process) so you can use SFTP or SCP to upload the API key (I will give instructions to use SFTP)
- In your new terminal, use 'cd' to navigate to the folder containing the API key. Typically, you will just need to type 'cd Downloads'. Check to make sure 'kaggle.json' is there using 'ls'.
- Once you are in the right folder, log on to the remote account via SFTP (ex: sftp $USER@chaos8.math.lsu.edu)
- Change (remotely) into the .kaggle folder: cd .kaggle (this is inside SFTP)
- Upload the API key: put kaggle.json (still using SFTP)
- You may now exit SFTP: exit

Next, install the command line tool

- Go back to your SSH terminal, or log back in if you did not open a separate terminal to upload the file
- In the remote shell, install the kaggle command line tool: pip3 install kaggle
- Try to use the tool: kaggle
- If the command is not found, then do this
  - Type this command: nano .bashrc
  - This will open a file editor. Scroll to the bottom of the file
  - Add this line: export PATH=$USER/.local/bin/:$PATH
  - Save (ctrl+o) and exit (ctrl+x)
  - Reload the shell (exit, then log back in)
- Try again, now with the help flag: kaggle –h

You should see the full help output, which means the command line tool is working.

Instructions to download the dataset and generate a training/validationtesting data split

First, download the dataset

- Log on to the remote account via SSH (ex: ssh $USER@chaos8.math.lsu.edu)
- Navigate to your scratch directory: cd /scratch/$USER
- (Optional) You may want to make some directory here and change into it
- Use the command line tool to download the lettuce dataset: kaggle datasets download baronn/lettuce-npk-dataset
- You will get a file called lettuce-npk-dataset.zip. Unzip it: unzip lettuce-npk-dataset.zip
- You will now see a folder called FNNPK

Next, upload and execute the script

- Download the script called make_train_val.py from Teams onto your laptop
- Use SFTP to upload the script to the folder which contains the FNNPK folder
- Exit SFTP after uploading. Log in to your remote account once more through SSH.
- Change directories (cd) to the directory containing this FNNPK folder
- Execute the script: python3 make_train_val.py
- If you get an error saying you don't have the python package numpy installed, then install it: python3 –m pip install numpy. Then execute the script.
- If the script runs successfully, then you should now have a folder called something like lettuce-8270. Examine the structure of this folder by recursive listing its contents: ls –R lettuce-8270

Instructions to train a model

First, upload the script

- Download the script called train_1.py from Teams onto your laptop
- Use SFTP to upload the script to the folder which contains the lettuce-8270 folder
- Exit SFTP after uploading

Next, execute the script to train and save the model

- Log in to your remote account once more through SSH
- Change into the directory containing the lettuce-8270 folder (ex: cd /scratch/$USER /lettuce/)
- The model will be saved in a folder called models, so make this directory now: mkdir models
- Execute the training script: python3 train_1.py
- List the contents of the models folder: ls models. You should see a directory named model_2.keras (note: the reason that the saved model number appears as 2 whereas the script is labeled as 1 is because the version of the script that was uploaded to teams has this discrepancy. Look at the end of the the training script to see where the saved model name comes from, ex: less train_1.py, scroll to the end)

Instructions to evaluate a model on the testing data

First, upload the script

- Download the script called evaluate.py from Teams onto your laptop
- Use SFTP to upload the script to the folder which contains the lettuce-8270 folder
- Exit SFTP after uploading

Next, identify the model that you want to evaluate

- Log in to your remote account through SSH
- Change into the directory containing your models folder (ex: cd/scratch/$USER /lettuce)
- List the contents of the models folder: ls models
- If you have trained any models, they should appear here with names of the form model_$NAME.keras

Next, execute the script to evaluate the model

- If you want to evaluate the model saved as model_$NAME.keras, then do: python3 evaluate.py -m $NAME
- You will see the accuracy of prediction on the testing data set of 44 images in the output