Course software for Text Analysis with LLMs

You can do the practicals on various computing platforms. However, since the participants may use various types of computers and softwares, we recommend to use the UNIL JupyterLab to do the practicals.

- JupyterLab: Working on the cloud is convenient because the installation of the Python packages is already done and you will be working with a Jupyter Notebook style. Note, however, that the UNIL JupyterLab will only be active during the course and for one week following its completion, so in the long term you should use either your laptop or Curnagl. Access requires that you connect either via the eduroam Wi-Fi with your UNIL account or through the UNIL VPN. This point is especially crucial for researchers from the CHUV.

- Laptop: This is good if you want to work directly on your laptop, but you will need to install the required libraries on your laptop. Warning: We will give general instructions on how to install the libraries on your laptop but it is sometimes tricky to find the right library versions and we will not be able to help you with the installation. The installation should take about 15 minutes.

- Curnagl: This is efficient if you are used to work on a cluster or if you intend to use one in the future to work on large projects. If you have an account you can work on your /scratch folder or ask us to be part of the course project but please contact us at least a week before the course. If you do not have an account to access the UNIL cluster Curnagl, please contact us at least a week before the course so that we can give you a temporary account. The installation should take about 15 minutes. Note that it is also possible to use JupyterLab on Curnagl: see https://wiki.unil.ch/ci/books/high-performance-computing-hpc/page/jupyterlab-on-the-curnagl-cluster

If you choose to work on the UNIL JupyterLab, then you do not need to prepare anything since all the necessary libraries will already be installed on the UNIL JupyterLab. In all cases, you will have access to the UNIL JupyterLab.

Otherwise, if you prefer to work on your laptop or on Curnagl, please make sure you have a working installation before the day of the course as on the day we will be unable to provide any assistance with this.

If you have difficulties with the installation on Curnagl we can help you, so please contact us before the course at helpdesk@unil.ch with subject: DCSR ML course.

On the other hand, if you are unable to install the libraries on your laptop, we will unfortunately not be able to help you (there are too many particular cases), so you will need to use the UNIL Jupyter Lab during the course.

Before the course, we will send you all the files that are needed to do the practicals.

JupyterLab

Here are some instructions for using the UNIL JupyterLab to do the practicals.

Access requires that you connect either via the eduroam Wi-Fi with your UNIL account or through the UNIL VPN.

This point is especially crucial for researchers from the CHUV.

The webpage's link will be given during the course.

Enter the login and password corresponding to your UNIL credentials.

Fill in the form as shown in the lecture's slides.

We have already prepared your workspace, including the data and notebook.

Double click on "Transformers_with_Hugging_Face.ipynb"

Change the "ipykernel" (top right button "Python 3 ipykernel") to LLM

In the notebook, check that

platform = "jupyter"

To execute a command, click on "Run the selected cells and advance" (the right arrow), or SHIFT + RETURN.

When you have finished the practicals, select File / Log out.

Laptop

You may need to install development tools including a C and Fortran compiler (e.g. Xcode on Mac, gcc and gfortran on Linux, Visual Studio on Windows).

Please decide in which folder (or path) you want to do the practicals, go there and copie the notebook there:

```
cd THE_PATH_WHERE_I_DO_THE_PRACTICALS
```

In the notebook, set

platform = "laptop"

Here are some instructions for installing PyTorch and other libraries on your laptop. You need Python >= 3.8.

For Linux

We will use a terminal to install the libraries.

Let us create a virtual environment. Open your terminal and type:

```
python3 -m venv mlcourse


source mlcourse/bin/activate


pip3 install torch torchvision torchinfo transformers accelerate datasets sentencepiece pandas
scikit-learn matplotlib sacremoses notebook ipywidgets gdown wget
```

> You may need to choose the right library versions

To check that PyTorch was installed:

```
python3 -c "import torch; print(torch.__version__)"
```

There might be a warning message (see above) and the output should be something like "2.3.0".

You can terminate the current session:

```
deactivate


exit
```

TO DO THE PRACTICALS (today or another day):

You can use any Python IDE (e.g. Jupyter Notebook or PyCharm), but you need to launch it after activating the virtual environment. For example, for Jupyter Notebook:

```
source mlcourse/bin/activate


jupyter notebook
```

For Mac

We will use a terminal to install the libraries.

Let us create a virtual environment. Open your terminal and type:

```
python3 -m venv mlcourse


source mlcourse/bin/activate


pip3 install torch torchvision torchinfo transformers accelerate datasets sentencepiece pandas
scikit-learn matplotlib sacremoses notebook ipywidgets gdown wget
```

> You may need to choose the right library versions

To check that PyTorch was installed:

```
python3 -c "import torch; print(torch.__version__)"
```

There might be a warning message (see above) and the output should be something like "2.3.0".

You can terminate the current session:

```
deactivate


exit
```

TO DO THE PRACTICALS (today or another day):

You can use any Python IDE (e.g. Jupyter Notebook or PyCharm), but you need to launch it after activating the virtual environment. For example, for Jupyter Notebook:

```
source mlcourse/bin/activate


jupyter notebook
```

For Windows

If you do not have Python installed, you can use either Conda:
https://docs.conda.io/en/latest/miniconda.html (see the instructions here:
https://conda.io/projects/conda/en/latest/user-guide/install/windows.html) or Python official
installer: https://www.python.org/downloads/windows/

We will use a terminal to install the libraries.

Let us create a virtual environment. Open your terminal and type:

```
python3 -m venv mlcourse


source mlcourse/bin/activate


pip3 install torch torchvision torchinfo transformers accelerate datasets sentencepiece pandas
scikit-learn matplotlib sacremoses notebook ipywidgets gdown wget
```

> You may need to choose the right library versions

To check that PyTorch was installed:

```
python3 -c "import torch; print(torch.__version__)"
```

There might be a warning message (see above) and the output should be something like "2.3.0".

You can terminate the current session:

```
deactivate
```

TO DO THE PRACTICALS (today or another day):

You can use any Python IDE (e.g. Jupyter Notebook or PyCharm), but you need to launch it after activating the virtual environment. For example, for Jupyter Notebook:

```
mlcourse\Scripts\activate.bat


jupyter notebook
```

Curnagl

For the practicals, it will be convenient to be able to copy/paste text from a web page to the terminal on Curnagl. So please make sure you can do it before the course. You also need to make sure that your terminal has a X server.

For Mac users, download and install XQuartz (X server): https://www.xquartz.org/

For Windows users, download and install MobaXterm terminal (which includes a X server). Click on the "Installer edition" button on the following webpage: https://mobaxterm.mobatek.net/download-home-edition.html

For Linux users, you do not need to install anything.

Here are some instructions for installing PyTorch and other libraries on the UNIL cluster called Curnagl. Open a terminal on your laptop and type (if you are located outside the UNIL you will need to activate the UNIL VPN):

```
ssh -Y < my unil username >@curnagl.dcsr.unil.ch
```

Here and in what follows we added the brackets < > to emphasize the username, but you should not write them in the command. Enter your UNIL password.

For Windows users with the MobaXterm terminal: Launch MobaXterm, click on Start local terminal and type the command ssh -Y < my unil username >@curnagl.dcsr.unil.ch. Enter your UNIL password. Then you should be on Curnagl. Alternatively, launch MobaXterm, click on the session icon and then click on the SSH icon. Fill in: remote host = curnagl.dcsr.unil.ch, specify username = < my unil username >. Finally, click ok, enter your password. If you have the question "do you want to save password ?" Say No if your are not sure. Then you should be on Curnagl.

See also the documentation: https://wiki.unil.ch/ci/books/high-performance-computing-hpc/page/ssh-connection-to-dcsr-cluster

You can do the practicals in your /scratch directory or on the course group "cours_hpc" if you have asked us in advanced:

```
cd /scratch/< my unil username >


or


cd /work/TRAINING/UNIL/CTR/rfabbret/cours_hpc
mkdir < my unil username >
cd < my unil username >
```

Clone the following git repos:

```
git clone https://git.dcsr.unil.ch/ML-Courses/llm_course.git
```

Let us install libraries from the interactive partition:

```
Sinteractive -m 10G -G 1

module load python/3.11.7

python -m venv mlcourse

source mlcourse/bin/activate

pip install -r llm_course/requirements_gpu.txt --extra-index-url
https://download.pytorch.org/whl/cu128
```

To check that PyTorch was installed:

```
python3 -c "import torch; print(torch.__version__)"
```

There might be a warning message (see above) and the output should be something like "2.10.0".

You can terminate the current session:

```
deactivate

exit
```

TO DO THE PRACTICALS (today or another day):

```
ssh -Y < my unil username >@curnagl.dcsr.unil.ch

cd /scratch/< my unil username >

or

cd /work/TRAINING/UNIL/CTR/rfabbret/cours_hpc/< my unil username >
```

You can do the practicals on the interactive partition:

```
Sinteractive -m 10G -G 1

module load python/3.11.7

source mlcourse/bin/activate
```

```
python
```

In the practical code (i.e. the Python code in the html file), you will need to set the paths as follows:

platform = "curnagl"

During the practicals, if you receive an error message "Disk quota exceeded", you will need to make some space in your home directory. For example, by deleting .cache.

Révision #26
Créé 19 juillet 2024 06:07:52 par Philippe Jacquet
Mis à jour 24 février 2026 17:40:14 par Philippe Jacquet