Lab session 1

Machine Learning for Behavioral Data (CS-421)

February 24, 2021

Today

- Welcome!
- Quiz and survey
- Tutorial 1.1: Setting up your environment
- Tutorial 1.2: YouTube trending videos part 1

Team

Instructor

Tanja Käser
tanja.kaeser@epfl.ch

Head Assistants

Christian Giang
christian.giang@epfl.ch

Mirko Marras
mirko.marras@epfl.ch

Teaching Assistants

Yufan Ren
yufan.ren@epfl.ch

Paola Mejia
paola.mejia@epfl.ch

Lab Sessions

- Wednesday, 8:15-10:00
- Online (on Zoom): https://tinyurl.com/mlbd-lab
  - **Meeting ID**: 811 6702 9743
  - **Passcode**: mlbd-lab
- Content:
  - Discussion of (past) homework (questions, usual mistakes, etc.)
  - Interactive tutorials on new topics
  - Introduction/discussion of new homework
  - Project office hours (second part of the semester)

Lab Sessions

- GitHub: https://github.com/d-vet-ml4ed/mlbd
- Tutorials:
  - Will be pushed before the lab session
  - Solutions will be pushed after the lab session

Homework

- Individual
- **5** homeworks during the first part of the semester, one per week
- Homeworks will be graded (amounts for 20% of the grade)
- We will compute the grade based on the **4** best homeworks you submit
  - If you submit only 4 homeworks, they will be used to compute the average
  - If you submit all 5 homeworks, we will use the 4 best homeworks to compute the average
- Need to be submitted by next Tuesday 23:59 CET

Cheating

We will carefully analyze submissions to detect similarities in responses in addition to other types of correlations. If your submission is found to be suspicious according to these criteria, we will mark your homework and everyone’s homework in the same cluster and discuss the appropriate measures with the EPFL legal team.

Project

- Teams of 3 people
- We will provide data sets
- We will provide example research questions
- You will suggest an additional analysis/extension to the selected research question
- We will give feedback during the semester (see milestones)
- We will do project office hours (during lab sessions)
- You will do a short presentation in the last week of the semester
- Final project (Code + Report) delivered on June 11, 2021 23:59 CET

Tentative Syllabus – Project

Week	Lecture	Lab Sessions	Project
1	Introduction	Tutorial	
2	Data Handling	Tutorial + **Homework**	
3	Classical Models	Tutorial + **Homework**	
4	Model Selection & Evaluation	Tutorial + **Homework**	Presentation of data sets and research questions
5	Latent Variable Models	Tutorial + **Homework**	M1: Preferences on team members and data sets
6	Unsupervised Learning	Tutorial + **Homework** + PO	
7	Spring Break	Spring Break	Spring Break

PO = project office hours

Tentative Syllabus – Project

8	Recommender Systems	Tutorial + PO	M2: Research Questions and Exploratory Analysis
9	Neural Networks	Tutorial + PO	
10	Sequence Mining	Tutorial + PO	
11	Representation/ Feature Learning	Tutorial + PO	M3: Suggested Approach and Preliminary Results
12	Multimodal Analytics	Tutorial + PO	
13	Multimodal Analytics	Tutorial + PO	M4: Mature Approach and Results with Discussion
14	White Monday	PO	
15	Bias/Fairness		Project Presentations

PO = project office hours

Grading

TENTATIVE

- 20% **Homework** (4 out of 5 during the semester)
  - Individual
  - Consolidation and application of topics taught in lecture
- 40% **Project** (details follow in March)
  - Teams of 3 people
  - 30% Presentation, 70% Report
- 40% **Final Exam** (exam session)
  - Individually, at the laptop
  - Mix of practical questions and “interpretation”

Questions?

Quiz

Quiz

https://www.python.org/

https://github.com/

https://www.anaconda.com/products/individual

https://jupyter.org/

https://noto.epfl.ch/

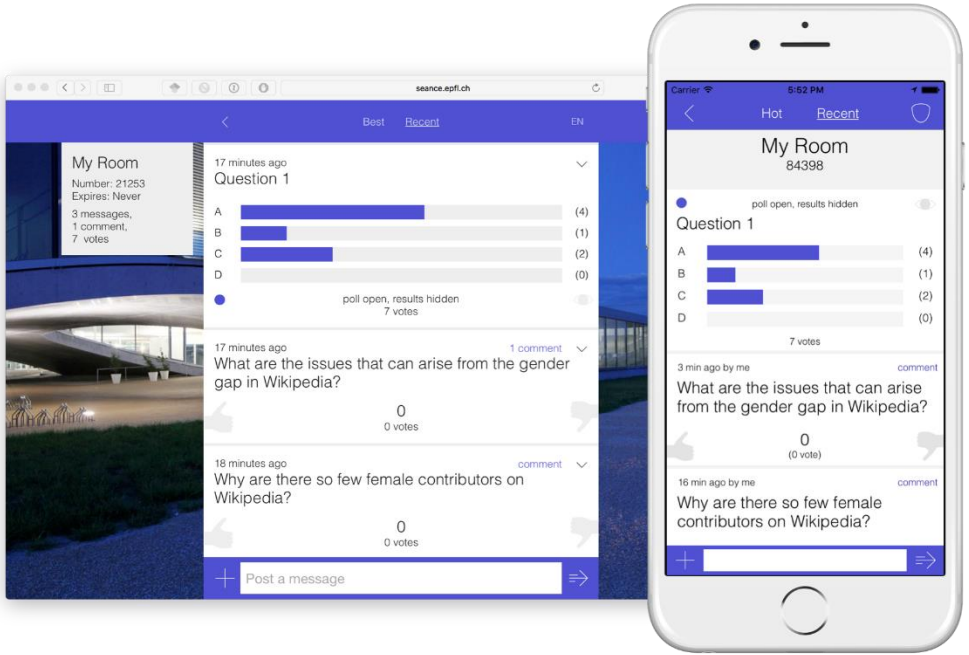
https://colab.research.google.com/

http://speakup.info/

SpeakUp

- **Android / iOS:**
  http://speakup.info/
- **Web App:**
  https://web.speakup.info/

- Room number: ???



Python

SpeakUp: How much do you know about Python?

A: It’s a family of nonvenomous snakes with 10 genera and 42 species.

B: I have heard about the programming language Python.

C: I have used Python a few times (e.g. for courses).

D: I use Python on a regular basis.

Jupyter

SpeakUp: How much do you know about Jupyter?

A: It’s the largest planet of our solar system.

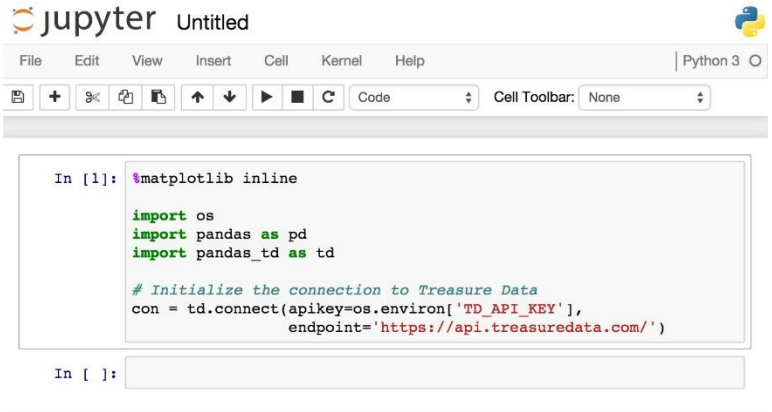
B: I have heard about Jupyter notebooks.

C: I have used Jupyter notebooks a few times (e.g. for courses).

D: I use Jupyter notebooks on a regular basis.

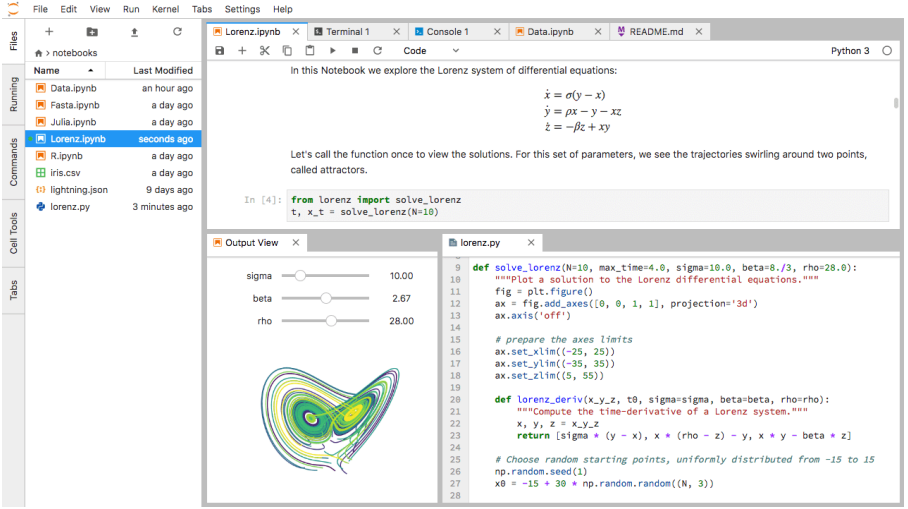
Jupyter

Jupyter notebook



Tutorial: https://www.dataquest.io/blog/jupyter-notebook-tutorial/

JupyterLab



Why JupyterLab: https://towardsdatascience.com/jupyterlab-a-next-gen-python-data-science-ide-562d216b023d

Anaconda (local env)

SpeakUp: How much do you know about Anaconda?

A: It’s the heaviest and one of the longest known snake species.

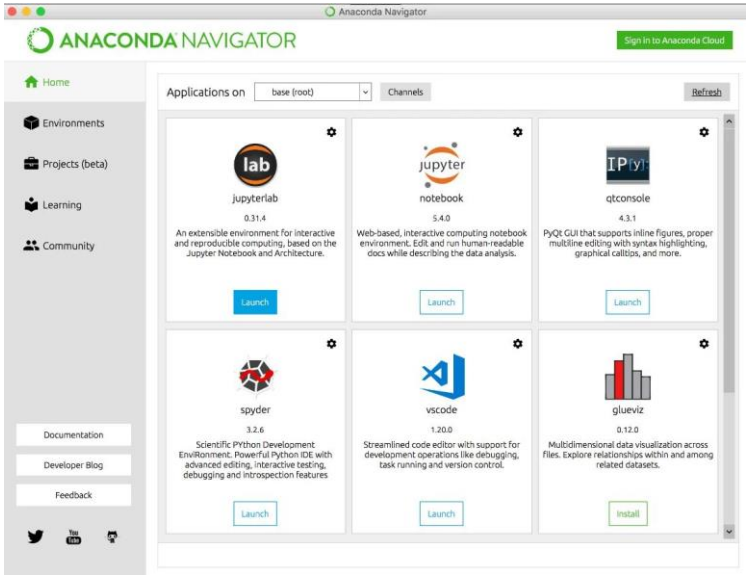
B: I have heard about Anaconda.

C: I have used Anaconda a few times.

D: I use Anaconda on a regular basis.

Anaconda (local env)

- You have the full control
- Works offline
- https://www.anaconda.com/products/individual



- **Tutorial**: https://www.edureka.co/blog/python-anaconda-tutorial/

Google Colab (online env)

SpeakUp: How much do you know about Colab?

A: It’s an abbreviation for an artist group from New York.

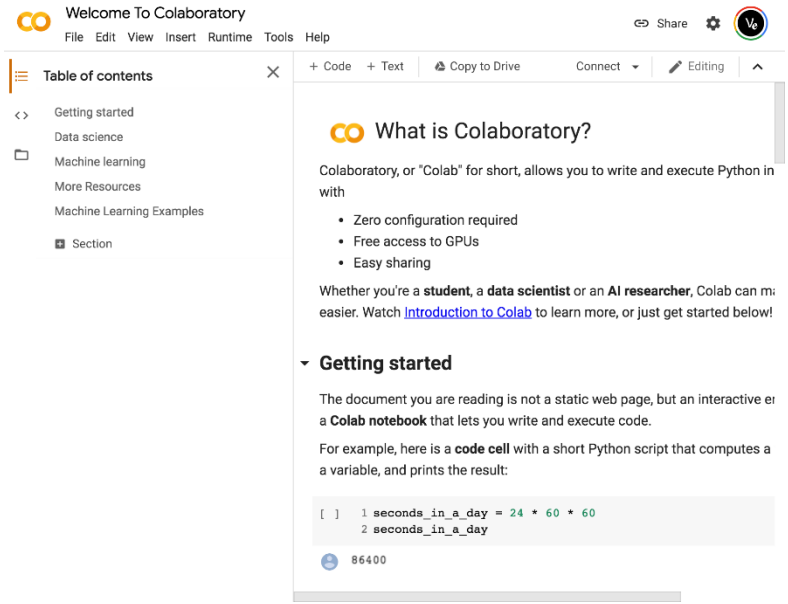
B: I have heard about Colab.

C: I have used Colab a few times.

D: I use Colab on a regular basis.

Google Colab (online env)

- Ready environment
- Uses Google's infrastructure
- Collaborative functionality
- Requires Google account
- https://colab.research.google.com/



- **Video**: https://www.youtube.com/watch?v=inN8seMm7UI

EPFL Noto (online env)

SpeakUp: How much do you know about Noto?

A: It’s a city in Sicily declared a UNESCO world heritage in 2002.

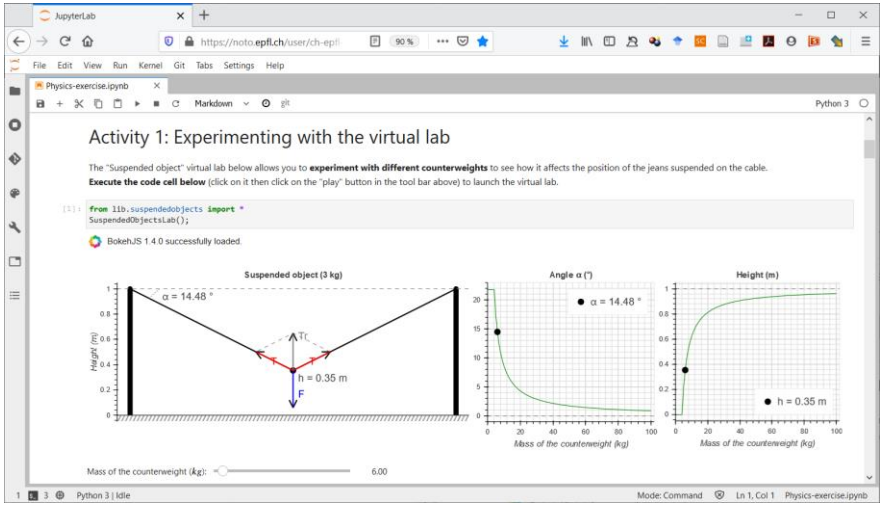
B: I have heard about Noto.

C: I have used Noto a few times.

D: I use Noto on a regular basis.

EPFL Noto (online env)

- Ready environment
- Login with your Gaspar
- https://noto.epfl.ch/



GitHub

SpeakUp: How much do you know about GitHub?

A: Git…..what?

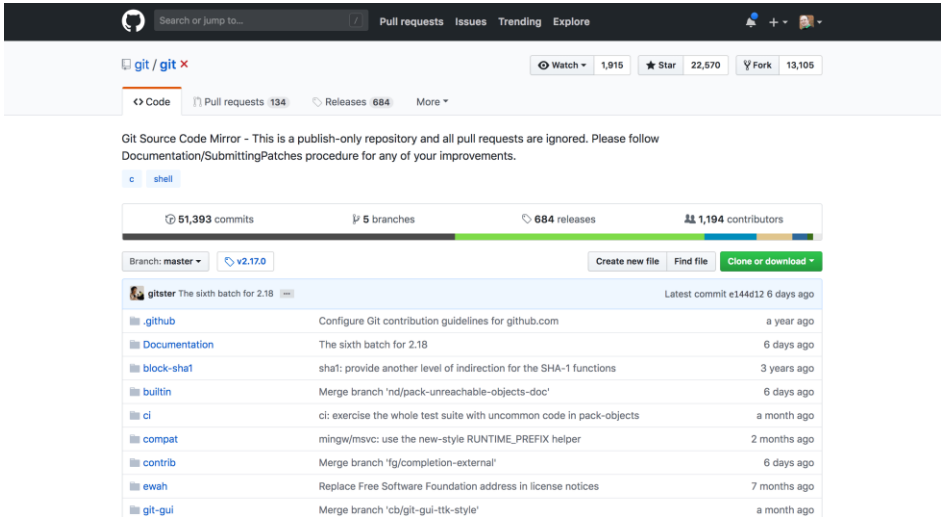
B: I have heard about GitHub.

C: I have used GitHub a few times.

D: I use GitHub on a regular basis.

GitHub

- Share files and code
- Version control
- The course repository:
  https://github.com/d-vet-ml4ed/mlbd



- **Tutorial**: https://www.edureka.co/blog/how-to-use-github/

Tutorial 1.1

- Set up an environment on which you can
  - Run Jupyter notebooks in Python
  - Connect to course repository: https://github.com/d-vet-ml4ed/mlbd
- We will use https://noto.epfl.ch/
  - But you are free to use whatever you want (e.g. Anaconda, Colab etc.)
  - It's your responsibility to have a working environment
- **Task**: Pull Tutorial 1.1 from GitHub and run it

Tutorial 1.1

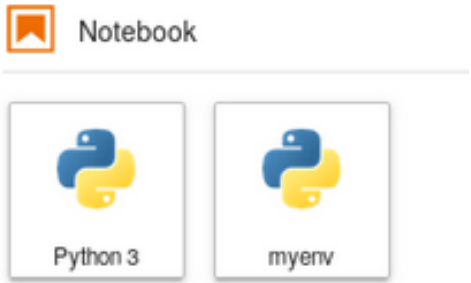
- Using Noto:
  - Go to https://noto.epfl.ch/
  - Login with your GASPAR
  - Go to Git -> Clone
  - Clone the course repository: https://github.com/d-vet-ml4ed/mlbd
  - Go through Tutorial 1_1.ipynb

Tutorial 1.2

- YouTube trending videos datasets for US and GB
- **Task**: Pull Tutorial 1.2 from GitHub and run it with the US dataset. Then do the same for the GB dataset and compare.

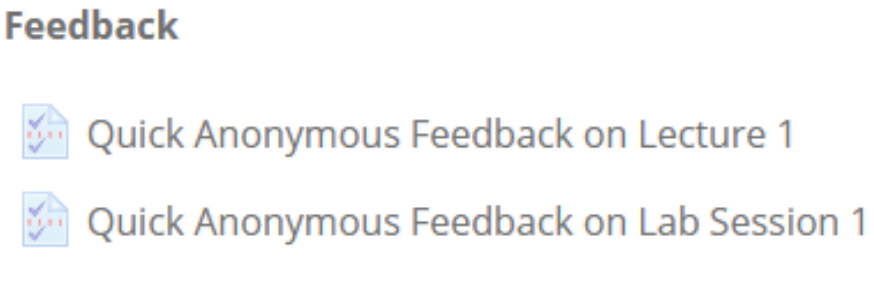
Tutorial 1.2

- Virtual environment:
  - https://janakiev.com/blog/jupyter-virtual-envs/
  - Create virtual environment: `python -m venv myenv`
  - Activate virtual environment: `source myenv/bin/activate`
  - add to Jupyter: `python -m ipykernel install --user --name=myenv`



Feedback

- New course -> please give feedback
- Very short feedback forms on Moodle every week (**anonymous**)
- Separate feedback for lecture and lab session



https://moodle.epfl.ch/mod/questionnaire/view.php?id=1130188

Questions?