UNIVERSITY OF MANNHEIM
Business School

Exercise 0

LLMs for the Economic and Social Sciences

Today

Overview

Introduction to:

- Python
- Package management
- Jupyter notebooks
- Version control
- GitHub
- BWUniCluster 3.0
- Google Colab
- Python refresher

Basics

Python

We will use Python throughout this course, and you will also need to use it for your projects.

https://www.python.org/

If you are not yet comfortable with Python, then you must invest time getting up to speed as soon as possible. An example tutorial can be found in the documentation.

Basics

Package management

Beyond basic Python, we will use additional packages e.g. NumPy.

When collaborating as a team, you should make sure you are using the same versions of packages, as functionality can change between versions. One option for handling this is Conda.

https://www.anaconda.com/download

Basics

Coding

When writing code, we often want more than just a text editor. There are many options with different features.

Jupyter notebooks enable interactive coding.
https://jupyter-notebook.readthedocs.io/en/latest/

There are also different computer programs that you can use called IDEs. These can help with highlighting errors in your code or making suggestions for code completion. Feel free to explore the options.

Version control

What is it? Why do we use it?

Version control enables the reversal of unwanted code changes and keeps track of progress.

It can be useful when coding alone, but it especially important when collaborating. Using version control means that multiple people can work on the code at the same time and then handle merging all the changes together.

A record of each version is maintained, meaning that you can also revert back to previous versions if needed.

Version control

GitHub

One of the most popular methods of implementing version control is using GitHub.

https://docs.github.com/en/get-started/start-your-journey/about-github-and-git

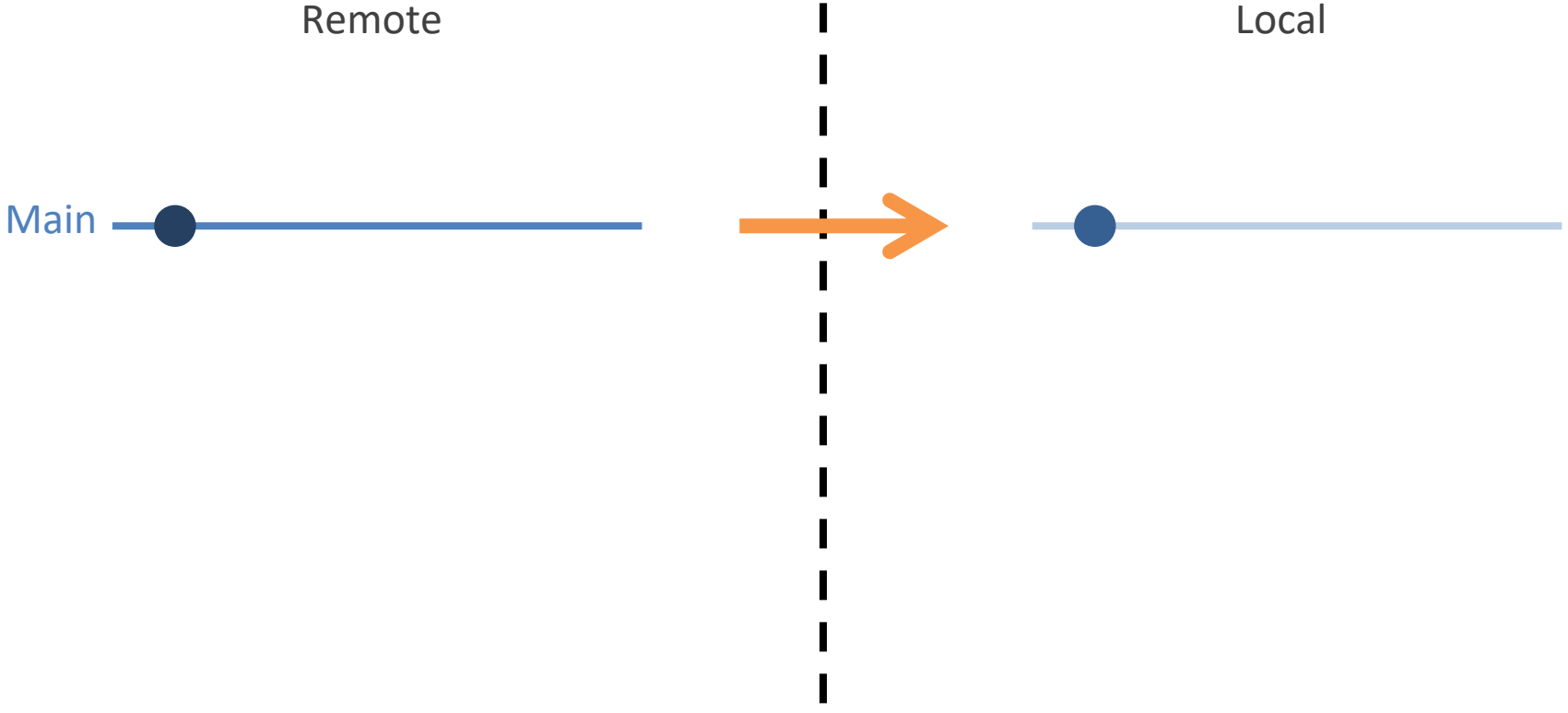
1. Create an account
2. Set up Git on your computer

Each code project is stored in a repository. There is a *remote* version of this online, but each user can also store a *local* version on their computer.

Now we will look at how to use GitHub theoretically, and then at the GitHub page for this course.

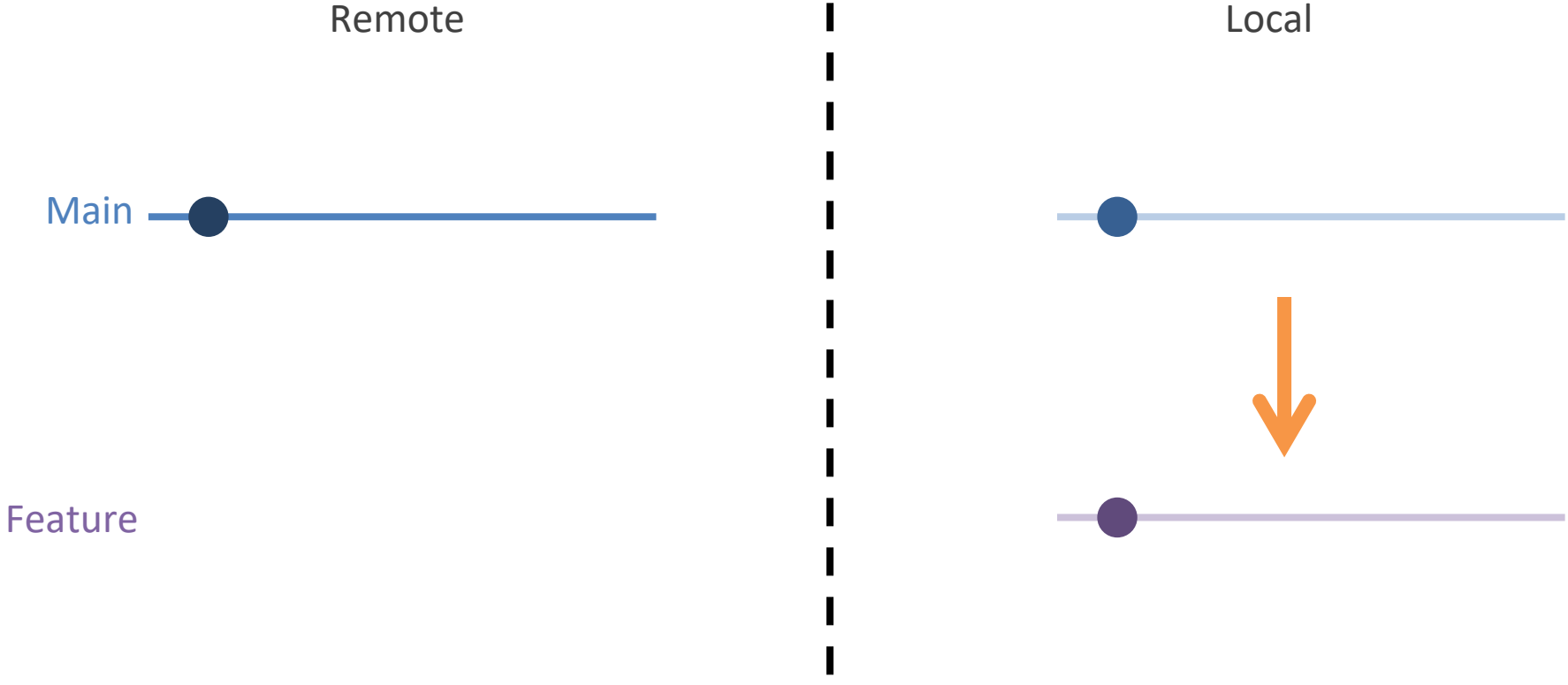
Version control

git pull



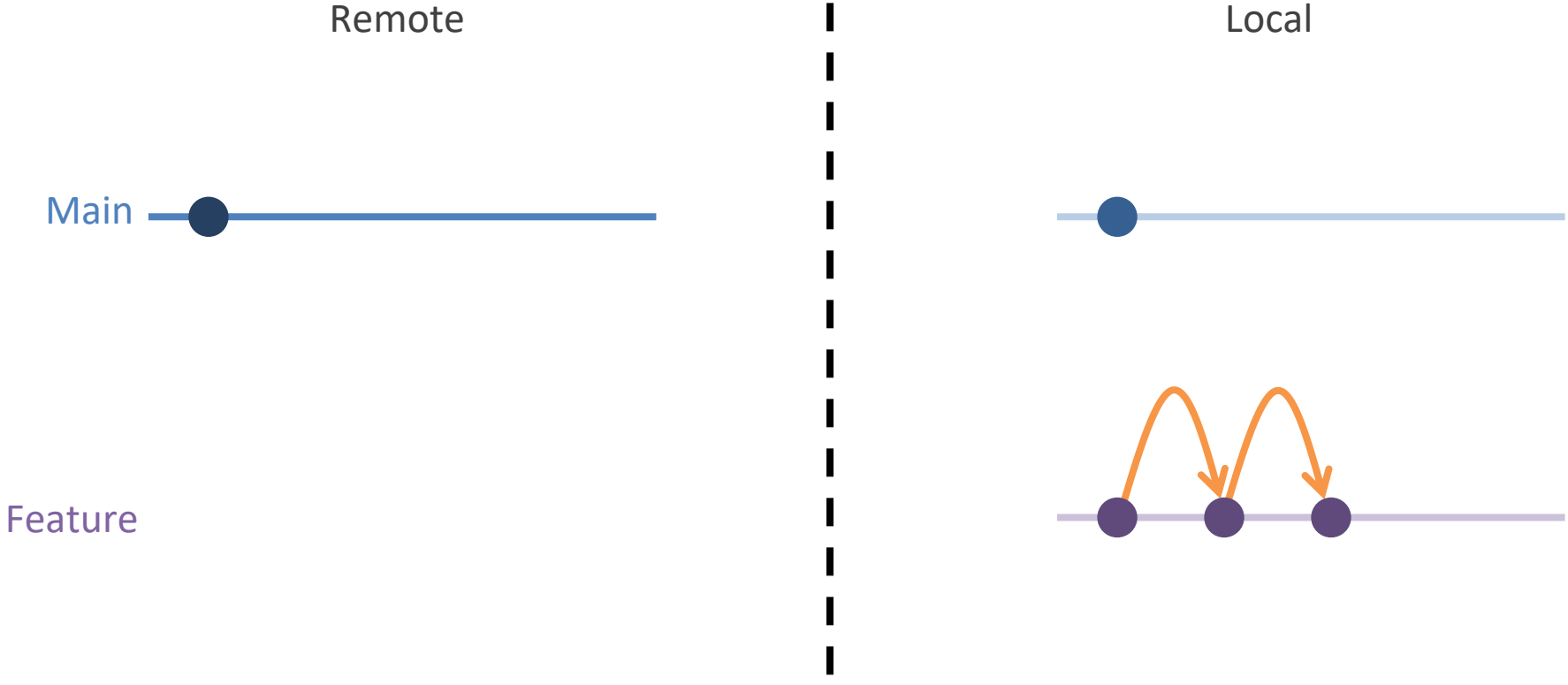
Version control

git branch new_branch_name



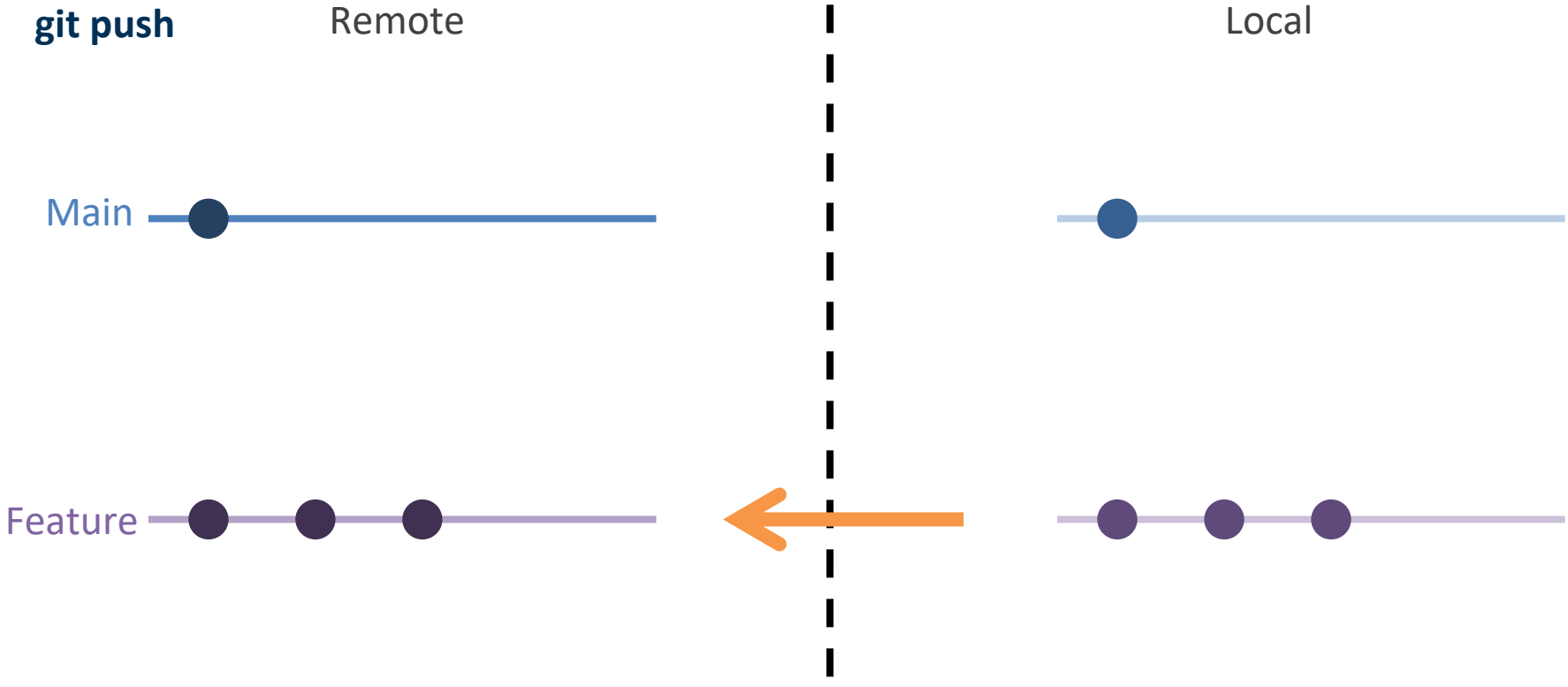
Version control

git commit –m “Added X”



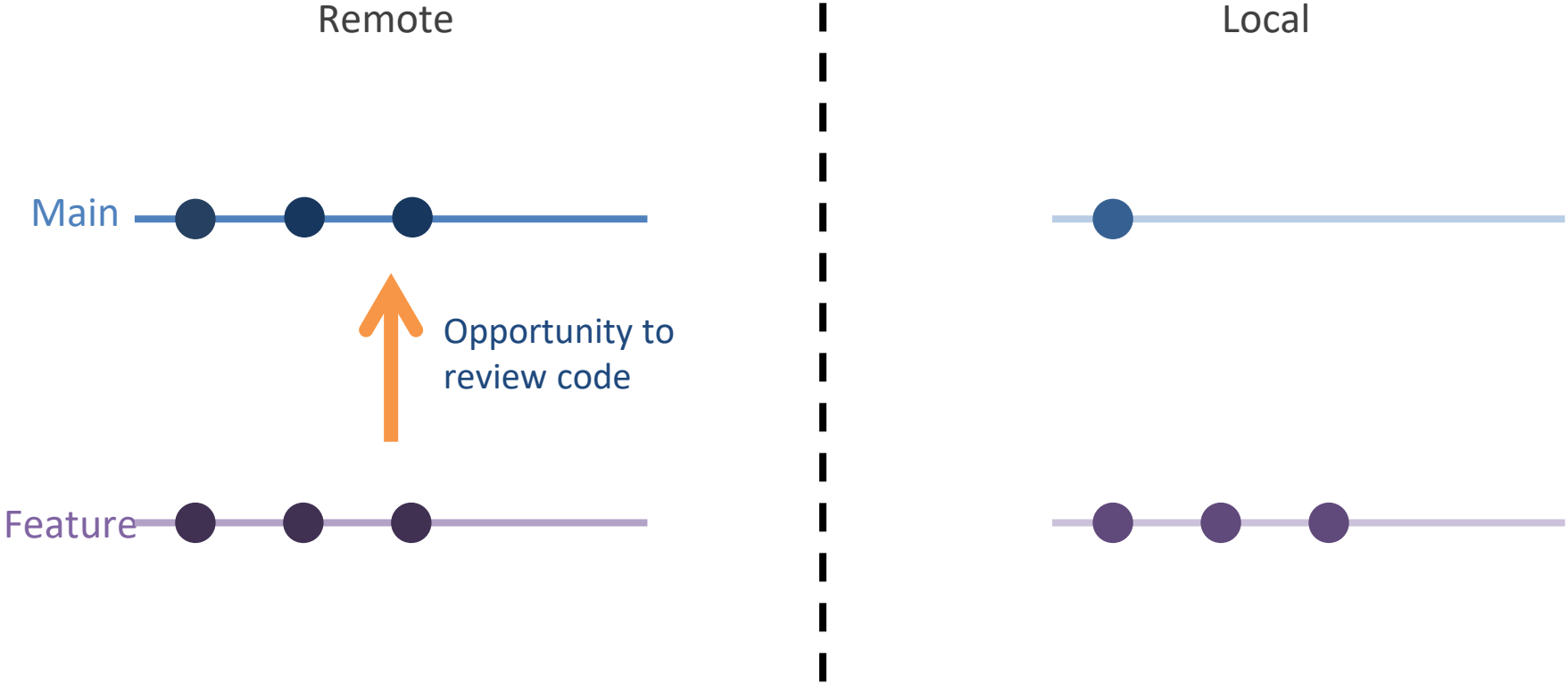
Version control

git add
git push



Version control

Pull request



Version control

GitHub demonstration

Feel free to follow along on your own computer.

GitHub course page

Computing resources

Why do we need additional resources?

LLMs range in size but generally require more computing resources than a standard laptop.

They can be accessed via APIs if hosted elsewhere or run on GPUs.

Additional computing resources could be accessing models via an API or hosting them yourself on GPUs.

Computing resources

BWUniCluster 3.0

https://wiki.bwhpc.de/e/BwUniCluster3.0/Getting_Started

High Performance Computing resource available through your membership of the university. After registration, you can request time using GPUs to run your experiments.

Getting access:

A. Request entitlement via the university - Request entitlement
B. Within 14 days of receiving entitlement, complete the questionnaire – Questionnaire
C. Register for an account on the cluster - Registration

Can be used via Slurm or a Jupyter notebook

Computing resources

Google Colab

https://colab.google/

Uses Jupyter Notebooks and enables access to the Gemini API and to GPUs

Practical

Text processing in Python

Now to 0_python_basics.ipynb…