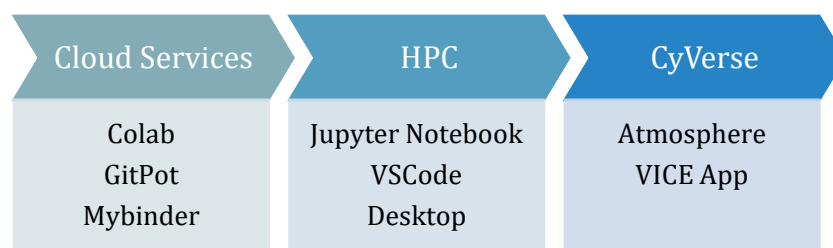# ADVANCED MACHINE LEARNING
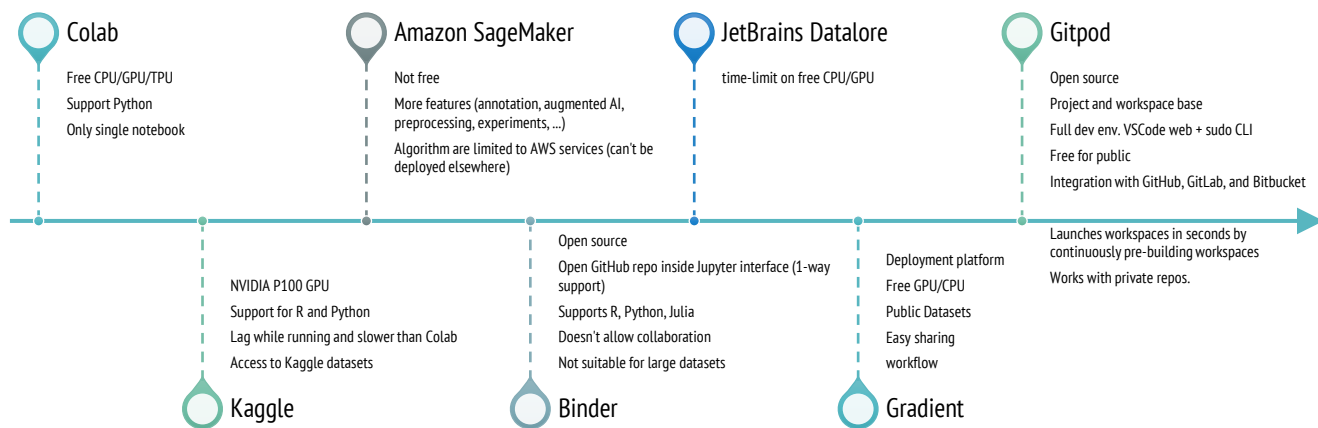## INTRODUCTION TO CLOUD HOSTED JUPYTER NOTEBOOKS

Date: 04/14/2022

Jupyter Notebook is a free and open-source tool that can be accessed through a web browser or directly through Jupyter, a Python program. It runs on several

Cloud Services	HPC	CyVerse
Colab	Jupyter Notebook	Atmosphere
GitPot	VSCode	VICE App
Mybinder	Desktop	

operating systems, including Linux, Windows, and Mac OS X. It supports code execution, inline figure plotting, Markdown (and LaTeX) text formatting, LaTeX equations, interactive widgets for displaying live code output in arbitrary programming languages (with language-specific backends) and controlling code execution with 2D code cells. Jupyter notebooks are excellent tools for data analytics, machine learning, and education in programming. They make it possible to create interactive documents that contain code, charts, and visualizations. However, sharing them is not always easy. You must first upload them to a remote server before distributing them. This can be time-consuming and inconvenient. Cloud-hosted notebooks, more commonly referred to as "cloud notebooks," are gaining traction with today's data scientists. They have numerous advantages over conventional notebook environments, which are frequently pre-installed on the user's machine. One of the primary advantages of cloud-hosted notebooks is improved accessibility. For instance, if you have internet access, you can access your profile at any time and from any location on the planet. This is not possible with traditional notebook environments, which must be installed on a personal computer. Additionally, users can save their work in the cloud for future reference or collaboration with peers or colleagues without having to download anything to their PC first. A cloud-hosted Jupyter Notebook is a server-based platform for conducting interactive data science and machine learning research. It streamlines the process of developing reproducible data analyses that can be shared online. Numerous variables play a role in determining which environment is optimal for you. The best cloud-hosted Jupyter notebooks balance security, functionality, and cost. You must create a prioritized list of the things that are most important to you. A notebook that is simple to use

is ideal for developing AI software. When choosing a platform for your Jupyter

Colab
Free CPU/GPU/TPU
Support Python
Only single notebook

Amazon SageMaker
Not free
More features (annotation, augmented AI, preprocessing, experiments, ...)
Algorithm are limited to AWS services (can't be deployed elsewhere)

JetBrains Datalore
time-limit on free CPU/GPU

Gitpod
Open source
Project and workspace base
Full dev env. VSCode web + sudo CLI
Free for public
Integration with GitHub, GitLab, and Bitbucket

Kaggle
NVIDIA P100 GPU
Support for R and Python
Lag while running and slower than Colab
Access to Kaggle datasets

Binder
Open source
Open GitHub repo inside Jupyter interface (1-way support)
Supports R, Python, Julia
Doesn't allow collaboration
Not suitable for large datasets

Gradient
Deployment platform
Free GPU/CPU
Public Datasets
Easy sharing workflow

Launches workspaces in seconds by continuously pre-building workspaces
Works with private repos.

Notebooks, the first consideration is whether you'll be utilizing GPU acceleration or CPU processing power. When performing large-scale computations, GPU acceleration is almost always preferable. On the other hand, if your computations are relatively simple, you may find that executing them on a single machine is sufficient. There must be considerations for security and privacy as well. Additionally, deep learning algorithms continue to require enormous amounts of data, which the majority of data scientists lack on their local workstation. Cloud notebooks enable us to conduct our analysis systematically, with well-designed services that cater to our needs and foster a reproducible environment in which we can share our work and conduct analysis with the least amount of backend effort.

In this workshop, we will explore how to exploit university computational resources such as HPC (Jupyter Notebook, Visual Studio Code), CyVerse (Atmosphere environment, VICE apps), and free-tier cloud hosting providers (GitHub Dev, Google Colab, Gitpod, Mybinder).