AI Paridigm: Neural AI

Image classification: MNIST dataset

COMP 741/841 Week 2 - Spring 2024

Agenda

- Reading assignment: Gains and perils of neural networks
- Lab 1
  - Development tools
    - git and GitHub, conda, Python virtual environments, Jupyter Notebooks
    - VS Code
  - Building an image classifier
- Due next week

Reading Assignment

- Main takeaways
- New concepts and terms
- Questions and issues of interest

Development Tools - Git and GitHub

- git: version control system to track changes to files in a repo
- GitHub: Microsoft-owned hosting service for storing and managing remote git repos
- Cloning repos
  - From GitHub org of the course: https://github.com/2024-spring-comp-741-841/
  - Get invitation link from Discord

Development Tools - Git

Basic local coammands to track changes

```
```

git status
git add file
git status
git commit –m 'meaningful message'
git status
git log
```
```

Networking commands

```
git clone <url remote repo>  <local repo name>
git push origin main
```

Development Tools – Python Virtual Environment

What is a Python virtual environment?

- Isolated run-time environment that has:
  - its own Python installation and
  - installed packages as needed by the project.
- Package and environment management tools:
  - conda, venv, pipenv, poetry, …

Why use a virtual environment?

- To manage versions and dependencies of packages by isolating them into the same environment
- To avoid installing packages globally (root directory) or user home directory

Development Tools – conda

- Creates and manages virtual environments
- Used for Python and other programming languages,
- Runs on Windows, MacOS, Linux
- Verifies that current environment installations do not conflict with new installations
- Installs Python instances
- Bundled with Anaconda or Miniconda

Development Tools – anaconda and miniconda3

- Anaconda platform - contains
  - Distribution of Python and R
  - Lots of packages automatically installed (~250) and more (~1,500)
  - GUI, CLI, …, and **conda**
- Miniconda – "bootstrap" version of Anaconda contains:
  - Distribution of Python
  - **conda**, and a small set of packages that **conda needs**
- We get **conda** from **miniconda3** installation

Development Tools – Jupyter Project

Provides tools for **interactive computing** across many programming languages.

- Jupyter **kernel**
  - Run-time environment of a Jupyter notebook in a specific programming langauge
    (Python, R)
  - **ipykernel**
    - Provides Python kernel (IPython) for a Jupyter notebook
- Jupyter Server
  - Backend (or core services and API) to a Jupyter application
  - Allows running Jupyter notebooks through web-based interfaces

Development Tools – Jupyter Project

- Jupyter Stack – layered services for
  - Users of Jupyter applications
  - DevOps of Jupyter applications (deploy and serve Jupyter apps to others)
  - Contributors to the Jupyter Server library
- Jupyter notebook (formerly IPython notebook)
  - Web application that creates and runs **interactive computational documents**
  - Computational document (or notebook)
    - Made of **cells**
    - Cells contains code, text (written in Markdown), visualizations, and more

Development Tools – Jupyter Noteboook and JupyterLab

- Managed by a kernel (run-time environment)
  - **ipykernel** runs selected Python-based notebooks (file extension **.ipynb**
  - Notebook cells can be run in any order
  - Variables are saved and managed by the kernel
- JupyterLab – newer user interface than the classic notebook
  - Available in GitHub coding platform Codespace
- Amazon SageMaker Studio Lab (SMSL)
  - Web-based interactive development environment for notebooks, code ,and data
  - Based on JupyterLab

Image Classification with MNIST

- What is MNIST?
  - Dataset of handwritten digits
  - Occasionally considered the "Hello World" of machine learning
- What is image classification in context of AI?
  - Neural network can identify and classify images it hasn't 'seen' before
  - Neural network can't distinguish between multiple objects in image
    - What happens if a model trained on MNIST is given an image with two digits?
- What are some real-world applicaitonsfor image classification?

Lab 1 – Getting Started

- Join COMP 741/841 GitHub organization
  - 2024-spring-comp-741-841
  - Home for remote repositories we create in this class
- GitHub Classroom roster has your Family name, Given name
  - When first "invited" to get access to a new repo, select your full name from the roster
  - A copy of the Lab 1 repo is created for your access only.

Due Next Week and Required Learning Materials

See **Week 2** module in Canvas.