A gentle introduction to ML via antibody-engineering

NCBI: Building Transparent ML/AI Solutions to Advance Biological Research Virtual Codeathon Feb. 26 – Mar 1, 2024

Tue 2/27: Workflow

Digital World Biology, LLC

Project Goals

Motivation

- ML/AI is hot
- Antibodies are important in therapeutics, diagnostics, and as reagents (spatial omics, flow cytometry, histology)
- Antibodies are used to teach protein purification, ELISAs, etc. in community college workforce education
- We get requests for ML modules

ML Education Challenges

- Vocabulary
- Methods, appropriateness
- Need infrastructure: data, tools, models
- Reproducing papers is hard/impossible
- Examples lack context (maps, apples, cars)
- Step by step examples are lacking
- **Teaching: sysadmin >> coding**

Can we?

- Identify a small number of key concepts (regression vs neural nets vs language models)
- Identify illustrative data sets, test cases
- Create infrastructure, libraries, install commands/scripts
- Documents steps and concepts taught
- Target different levels of experience

The Project

1. Immuneprofiling

Establish a jupyter notebook experience and underlying software

iReceptor dataset => comparative analyses

2. Machine learning

Establish a jupyter notebook experience and underlying software – AbLang2

SabDab dataset => characterization

3. Dictionary/glossary

Identify and define essential computation and biology terms related to 1 and 2

Docker: First attempt several issues