

CS5841/EE5841 Machine Learning

Lecture 8: MLOps and friends

Evan Lucas



Michigan Tech

Overview

- Course updates
- MLOps
 - Intro to concepts
- Practical
 - Source control
 - Sci-kit Learn
 - Experiment tracking
 - Building ML demos



Class updates

- Exam Friday!!
- Happy Valentine's Day



Happy Valentine's Day



state of the art for candy heart message generation in 2018 (GPT-2) [2]



state of the art for candy heart message generation in 2021 (GPT-3 DaVinci) [1]

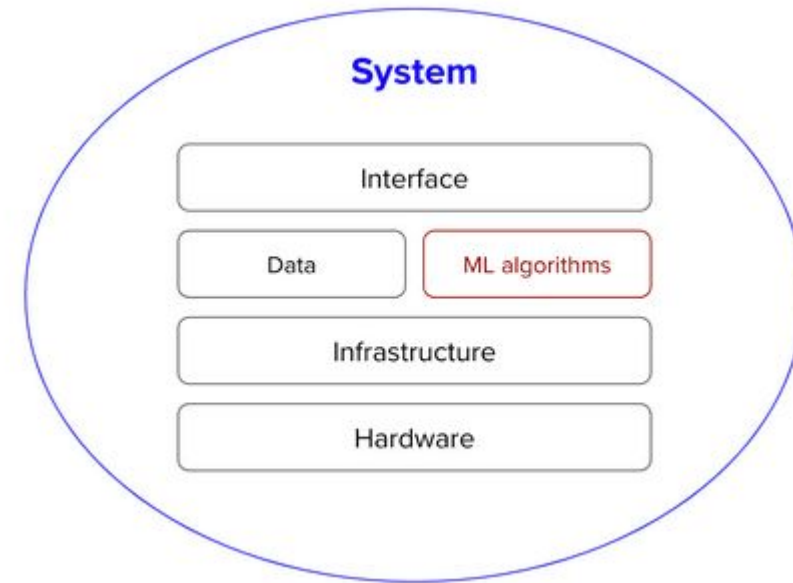
[1] <https://janellecshane.substack.com/p/okay-gpt-3-candy-hearts>

[2] <https://www.aiweirdness.com/candy-heart-messages-written-by-a-18-02-09/>



ML in production

- ML is a tiny part of the complete system
- ML isn't always the answer
 - DL isn't always the answer

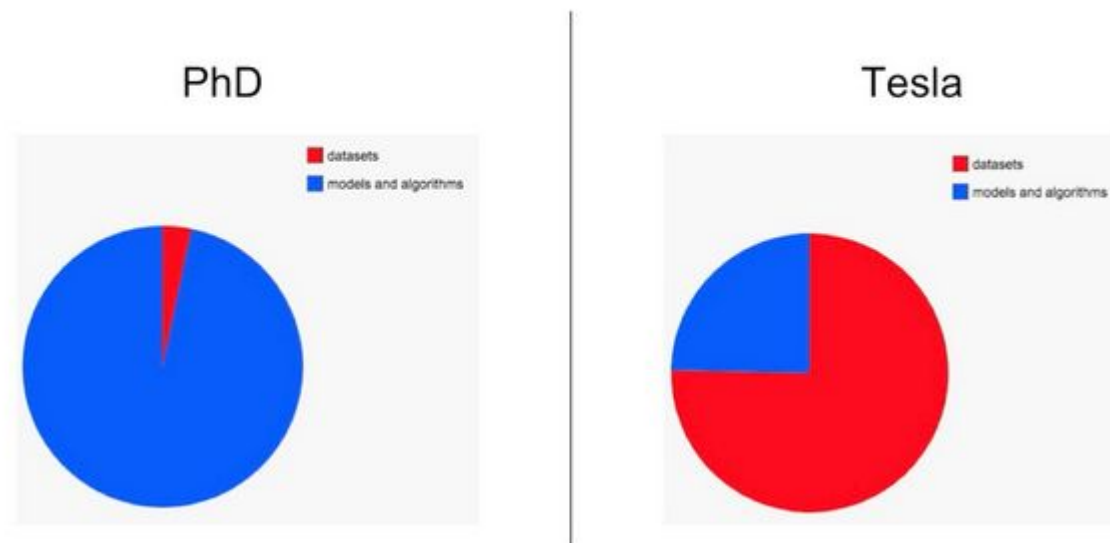


What takes the most time/money in real ML applications?



Data!!!

Amount of lost sleep over...



Typical MLOps Workflow*

1. Pick a metric
2. Collect data + labels
3. Engineer features
4. Train models
5. Realize labels are bad, go get new labels
6. Train model
7. Model performs poorly on one class -> collect more data for that class
8. Train model
9. Collect new data, realize model performs poorly due to data drift
10. Collect more recent data and labels
11. Train model
12. Deploy model
13. Realize that you picked the wrong metric
14. Pick a new metric
15. Collect data + labels

*crying, praying, and
daydreaming about \$\$\$ not
included



Source control

- git
 - Probably most popular modern source control management tool
- Github is a great resource
 - Learn to use the CLI - more powerful, faster
 - Website has almost everything the desktop app has
 - Integrates with VS Code
 - Students get freebies - free private repos, Copilot, etc.

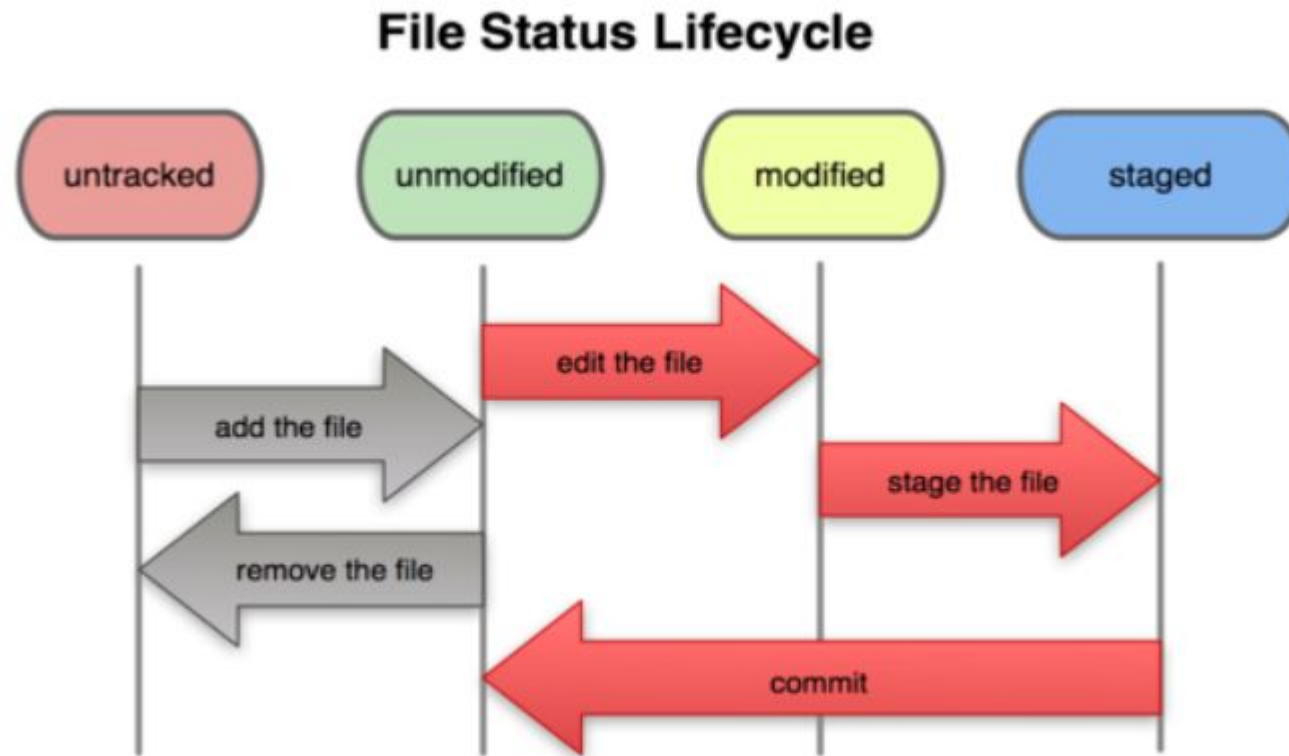


git

- Came out of Linux dev community to do version control on Linux kernel
 - Created by Linus Torvalds
- Intended for distributed development



Basic git workflow



Getting started with git

- Add SSH keys to your git profile
 - Check connection with `$ssh -T git@github.com`
- Install git (typically already installed for Linux/Mac)
- Configure git
 - `$git config --global user.name NAME`
 - `$git config --global user.email EMAIL`
 - without these, your commits won't be signed



Using git

Common commands:

- init - start a project
- add - add a file to the tracked files
- commit - commit a saved state of the tracked files
- push - send your commits to the central repository
- pull - get the commits last saved to the central repository
- checkout BRANCH - tell git which branch to modify
- merge - combine branches
- clone - copy a repository
- help COMMAND - get help for a command
- reset - unstages a file



Scikit-Learn

- Package of classic machine learning algorithms
- Not always the most efficient implementations...
- <https://scikit-learn.org/stable/index.html>
- Demo!



Experiment tracking

- ML requires iterative experimentation
- Tracking changes in parameters/data/etc. is critical
- One suggested tool (free for small teams and students) is Weights & Biases (<https://wandb.ai>)



Simple model demos

- For sharing simple demos of a model
 - Gradio - <https://www.gradio.app/>
 - Demo



Actual ML model integration

- Model inference on small inputs
 - Usually build an API for your model
 - FastAPI is common for Python projects
 - Full webpage with integrated model
 - Flask is a common option
- Model training or inference on large inputs
 - Containerize your model
 - Typically with Docker
 - Manage your containers with Kubernetes



Questions + Comments?



Slide sources and further reading

<https://stanford-cs329s.github.io/syllabus.html>

<https://huyenchip.com/machine-learning-systems-design/toc.html>

<https://fullstackdeeplearning.com/course/2022/>

