Twitter Thread by Pratham

Pratham
@PrasoonPratham

Everyone tells you how important math is in machine learning.

But I believe that if you want to learn machine learning today, there are things more important than math which very few people will tell you about.

Here are 5 of them.

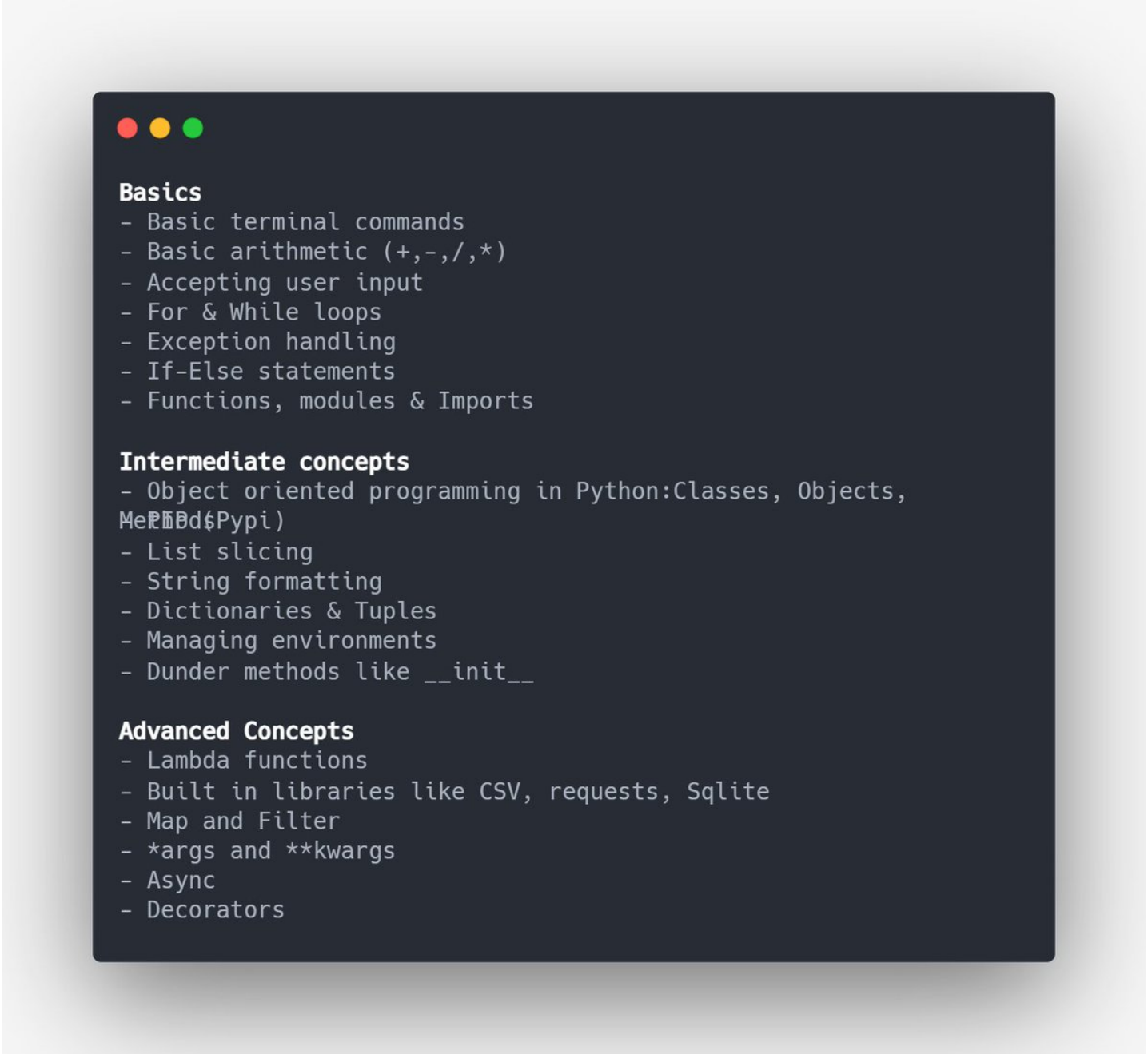
■ ■■

The field of machine learning has been very academic which is why there is so much emphasis on learning math for it.

But, today with the frameworks that we have, not knowing how to multiply matrices is far less of an issue than knowing how to use the tools in this list ■■

Yes, I know; you have R, Julia, and whatnot for machine learning but as of today, Python is the industry standard.

Get your data structures and algorithms right in Python because you will need them, here are some concepts worth learning ■■



Let's say you're building the next big GPT-3 powered app that's going to take over the world.

You add a new feature to your app but it breaks the entire thing, what do you do now?

Learn Git, the ultimate version control tool.

I use this cheat sheet for all my git needs ■■

Git Cheat Sheet

REBELLABS by ZEROTURNAROUND

Create a Repository

From scratch -- Create a new local repository
`$ git init [project name]`

Download from an existing repository
`$ git clone my_url`

Observe your Repository

List new or modified files not yet committed
`$ git status`

Show the changes to files not yet staged
`$ git diff`

Show the changes to staged files
`$ git diff --cached`

Show all staged and unstaged file changes
`$ git diff HEAD`

Show the changes between two commit ids
`$ git diff commit1 commit2`

List the change dates and authors for a file
`$ git blame [file]`

Show the file changes for a commit id and/or file
`$ git show [commit]:[file]`

Show full change history
`$ git log`

Show change history for file/directory including diffs
`$ git log -p [file/directory]`

Working with Branches

List all local branches
`$ git branch`

List all branches, local and remote
`$ git branch -av`

Switch to a branch, my_branch, and update working directory
`$ git checkout my_branch`

Create a new branch called new_branch
`$ git branch new_branch`

Delete the branch called my_branch
`$ git branch -d my_branch`

Merge branch_a into branch_b
`$ git checkout branch_b`
`$ git merge branch_a`

Tag the current commit
`$ git tag my_tag`

Make a change

Stages the file, ready for commit
`$ git add [file]`

Stage all changed files, ready for commit
`$ git add .`

Commit all staged files to versioned history
`$ git commit -m "commit message"`

Commit all your tracked files to versioned history
`$git commit -am "commit message"`

Unstages file, keeping the file changes
`$ git reset [file]`

Revert everything to the last commit
`$ git reset --hard`

Synchronize

Get the latest changes from origin (no merge)
`$ git fetch`

Fetch the latest changes from origin and merge
`$ git pull`

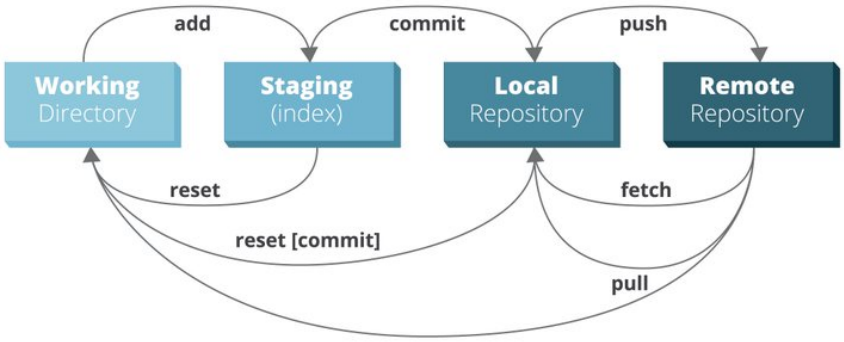
Fetch the latest changes from origin and rebase
`$ git pull --rebase`

Push local changes to the origin
`$ git push`

Finally!

When in doubt, use git help
`$ git command --help`

Or visit https://training.github.com/ for official GitHub training.



You will be constantly be working on multiple packages which specific python version requirements ■

Virtual environments like Anaconda or Poetry will make this process much easier for you.

Who likes to see broken dependencies while running code?

At some point in time, you will have to deploy that crazy machine learning app that you made, say hello to docker.

Docker essentially allows you to package your apps in a box that can run on any operating system, also very helpful in several other situations.

Learn docker ■

Storytelling is an important part of machine learning, explaining how your model works will make it easier for others to make changes and improve it.

Markdown and LaTeX allow you to present your text explanations beautifully, learn these markup languages.

I started working with these tools a lot more after I got my job this year and it was surprising how few people talk about them.