IEOR 8100
Reinforcement Learning

Introductory presentation

Course Staff

- Instructor: Shipra Agrawal
  - Office hours: Wednesday 2:00pm-3:00pm Mudd 423

  (starting next week)
- TA
  - Robin Tang (PhD candidate, IEOR)

  Office hours 12:30 to 1:30 pm on Friday. Mudd 301 (computer lab)

ieor8100.github.io/rl/

Communication: *Piazza*

- No emails (unless absolutely necessary)
- Post questions on Piazza
  - Sign up for piazza
  - Can post publicly/privately/anonymously
- Announcements will be made on Piazza

Course requirements

- 4 lab assignments
- One paper presentation
- One research project

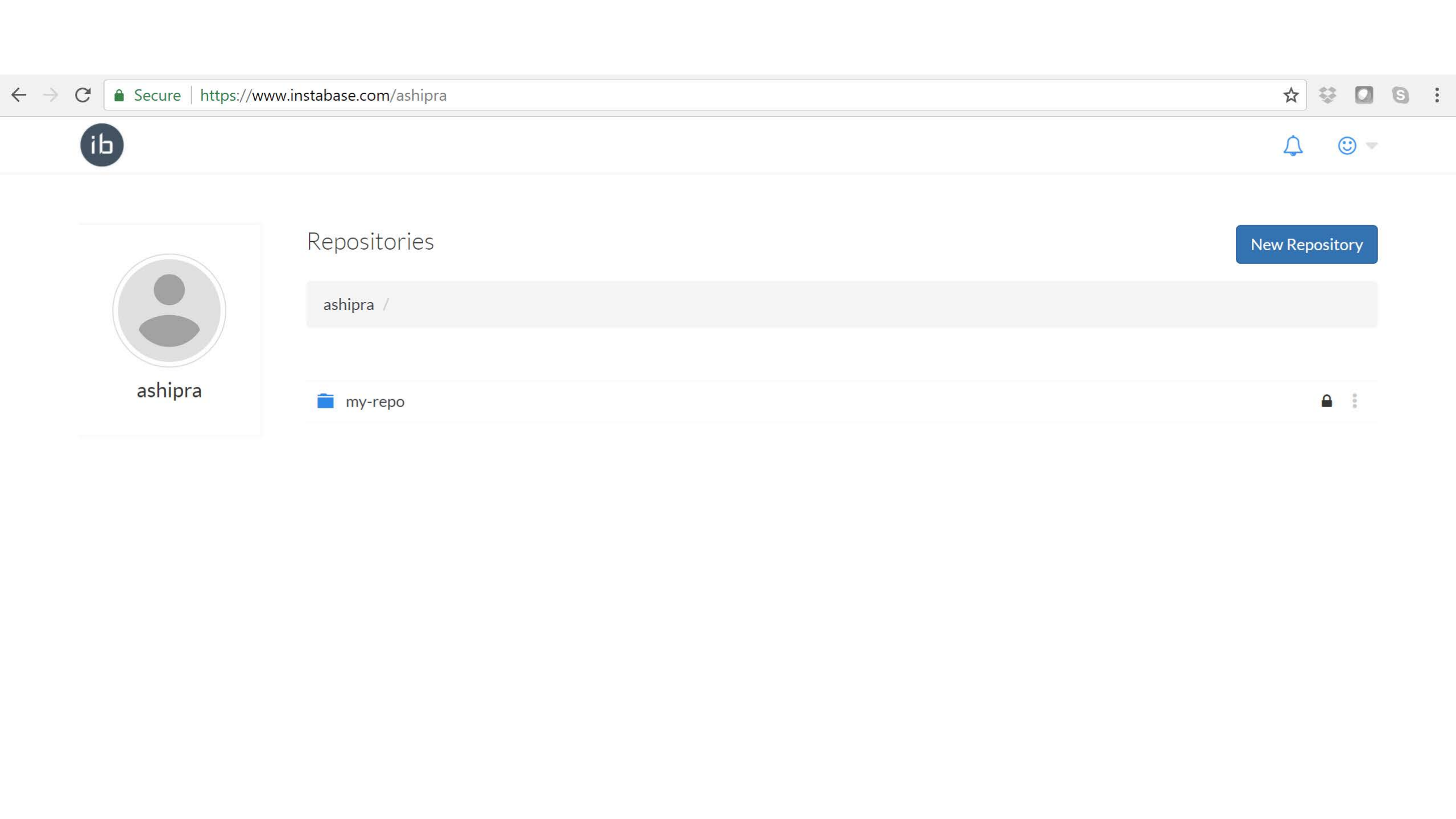
Research papers, project ideas and presentation schedule will be posted in two weeks.

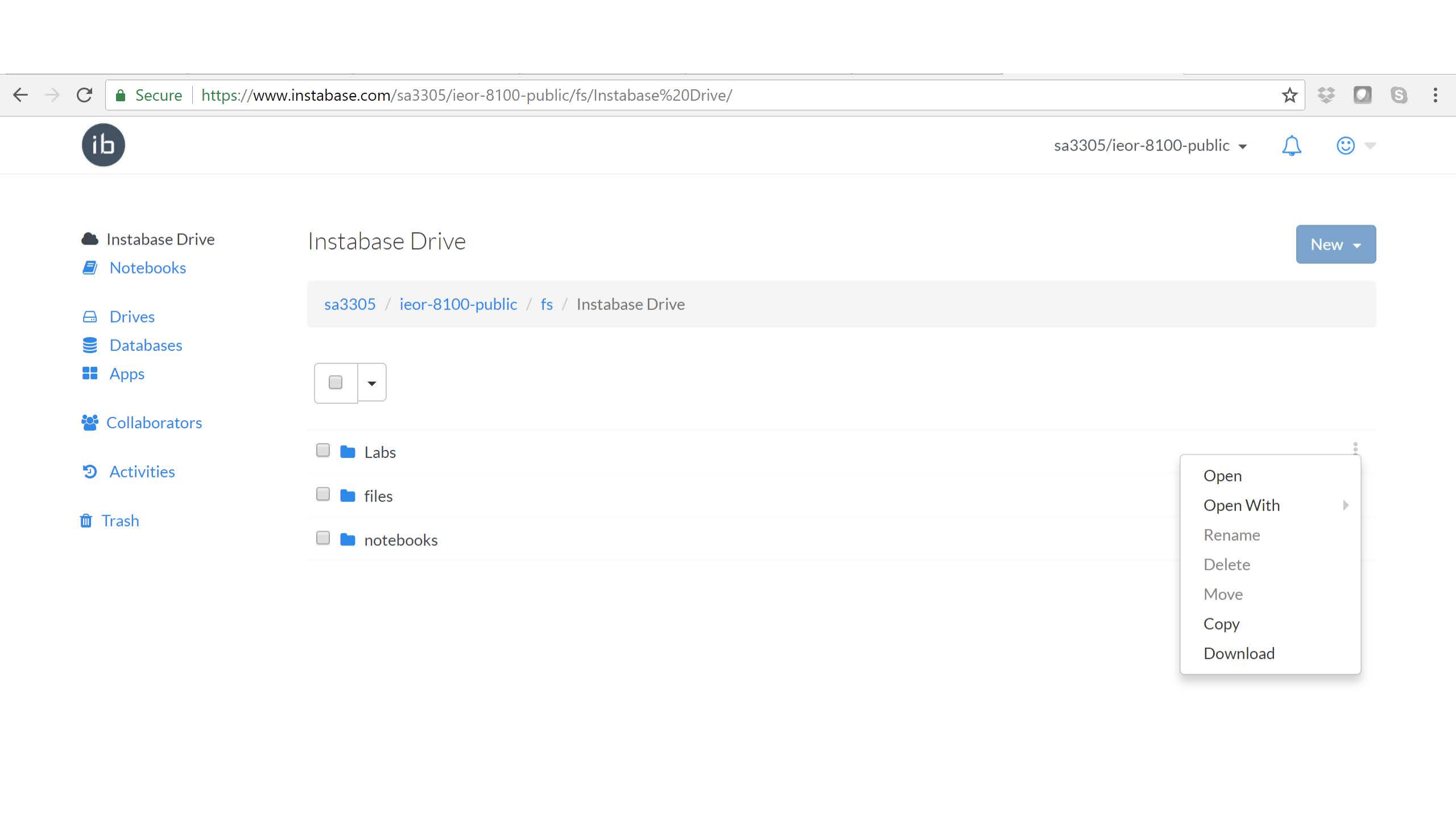
Assignments: *Instabase*

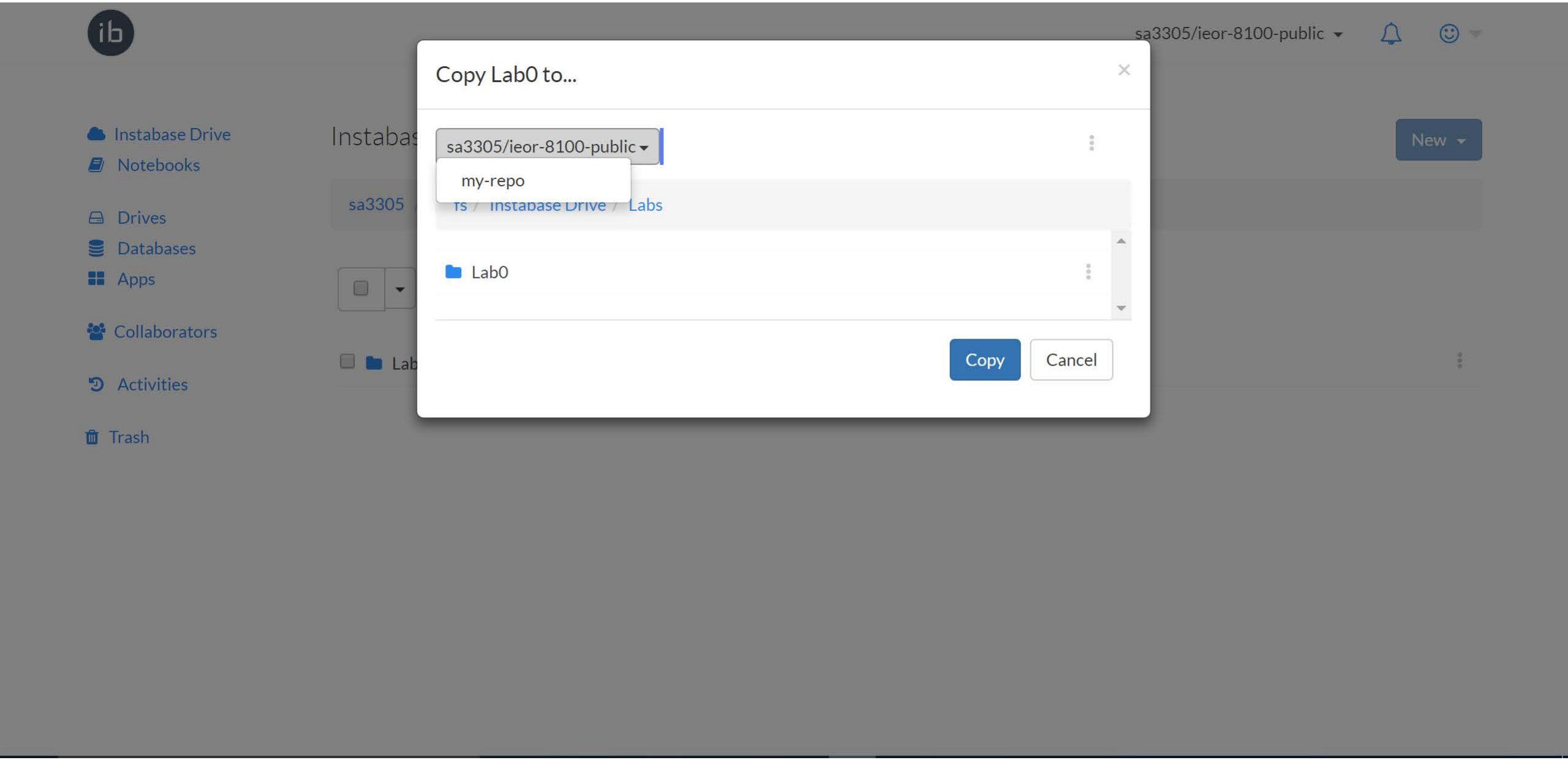
- All software is pre-installed (Python+Jupyter+tensorflow+openAI gym)
- Create account using username as UNI, email as UNI@columbia.edu
  - Signup instructions https://ieor8100.github.io/rl/cloudPlatform.html
  - Signup requires a token.
- Access the lab assignments (Jupyter notebooks with skeleton code)

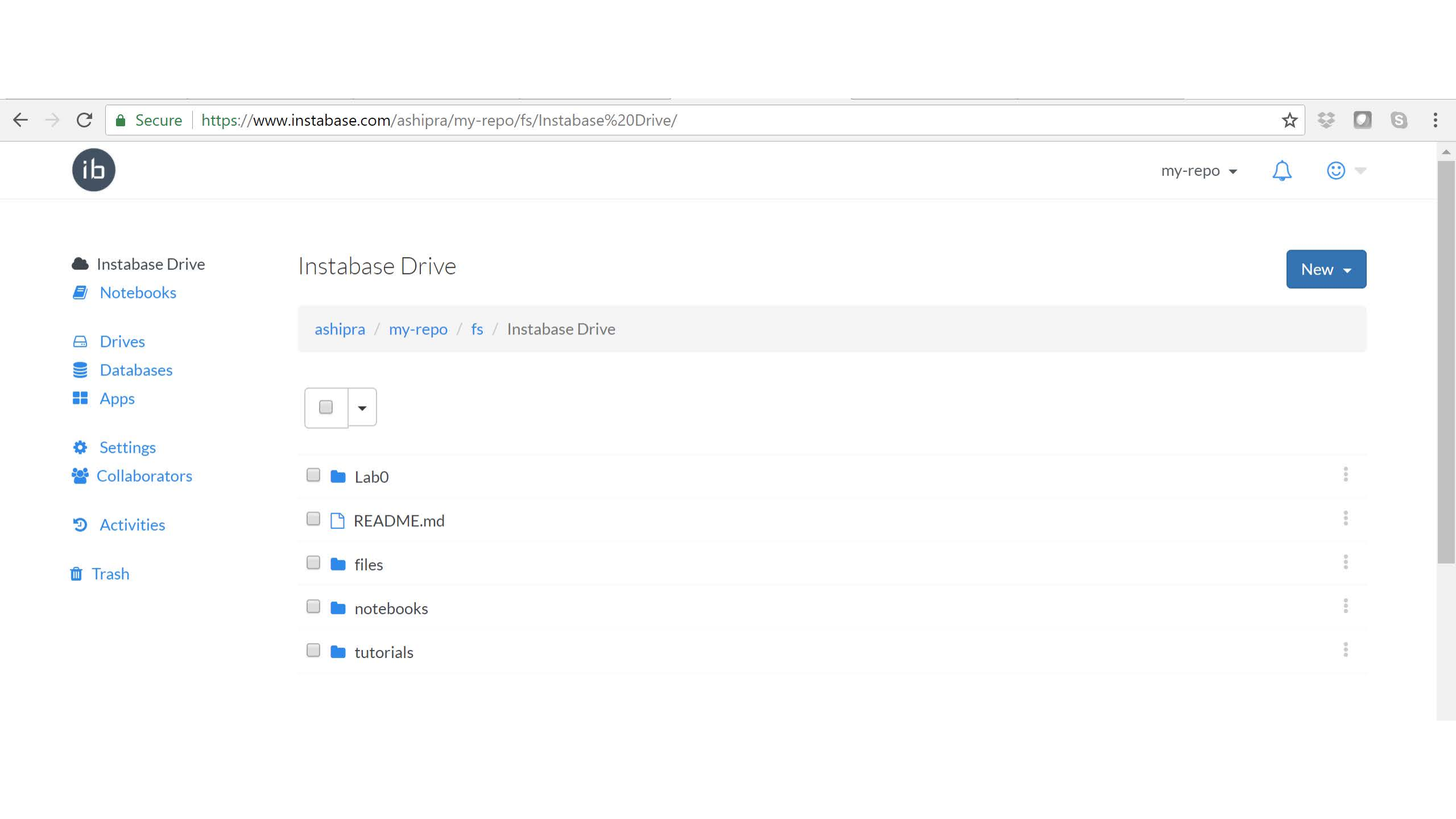
https://www.instabase.com/sa3305/ieor-8100-public

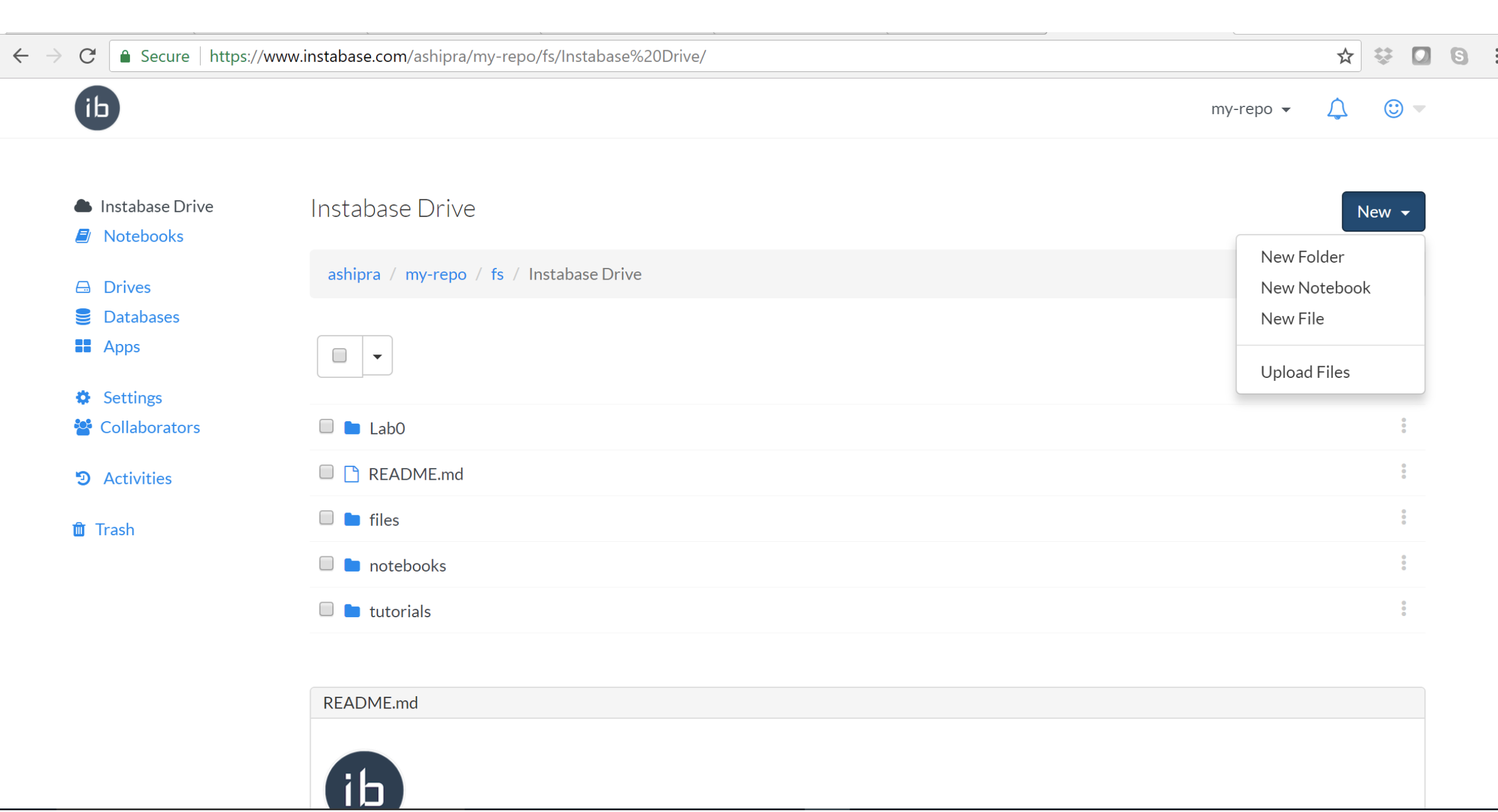
  - Login and copy the required folder to your repository (my-repo)
  - Change/write code as required
  - Submit using the submit link


Secure
https://www.instabase.com/ashipra
ashipra
Repositories
New Repository
ashipra /
my-repo


Secure
https://www.instabase.com/sa3305/ieor-8100-public/fs/Instabase%20Drive/
sa3305/ieor-8100-public
Instabase Drive
Notebooks
Drives
Databases
Apps
Collaborators
Activities
Trash
Instabase Drive
New
sa3305 / ieor-8100-public / fs / Instabase Drive
Labs
files
notebooks
Open
Open With
Rename
Delete
Move
Copy
Download


sa3305/ieor-8100-public
Instabase Drive
Notebooks
Drives
Databases
Apps
Collaborators
Activities
Trash
New
sa3305
Copy Lab0 to...
sa3305/ieor-8100-public
my-repo
fs / Instabase Drive / Labs
Lab0
Copy
Cancel


Secure
https://www.instabase.com/ashipra/my-repo/fs/Instabase%20Drive/
my-repo
Instabase Drive
Notebooks
Drives
Databases
Apps
Settings
Collaborators
Activities
Trash
Instabase Drive
New
ashipra / my-repo / fs / Instabase Drive
Lab0
README.md
files
notebooks
tutorials


Secure
https://www.instabase.com/ashipra/my-repo/fs/Instabase%20Drive/
my-repo
Instabase Drive
Notebooks
Drives
Databases
Apps
Settings
Collaborators
Activities
Trash
Instabase Drive
New
New Folder
New Notebook
New File
Upload Files
ashipra / my-repo / fs / Instabase Drive
Lab0
README.md
files
notebooks
tutorials
README.md

TO DO

1. Sign up for Piazza
2. Signup for Instabase
   - Instructions on the website https://ieor8100.github.io/rl/cloudPlatform.html
   - Important use your UNI as username, email UNI@columbia.edu
   - contact us on Piazza if you have any difficulty signing up
3. Access Lab0, submit it as trial. (will not be graded)
4. Get your computer ready for offline implementation:

   Software installation instructions posted on the website https://ieor8100.github.io/rl/installation.html

Course Introduction

Reinforcement Learning

- Agent interacts and learns from a stochastic environment
- Science of sequential decision making
- Many faces of reinforcement learning
  - Reward systems (Neuro-science)
  - Classical/Operant Conditioning (Psychology)
  - Optimal control (Engineering)
  - Dynamic Programming (Operations Research)

Characteristics of Reinforcement Learning

- Sequential/online decisions
- No supervisor, only reward *signals*
- Feedback is delayed
- Actions effect observations (non i.i.d. training examples)

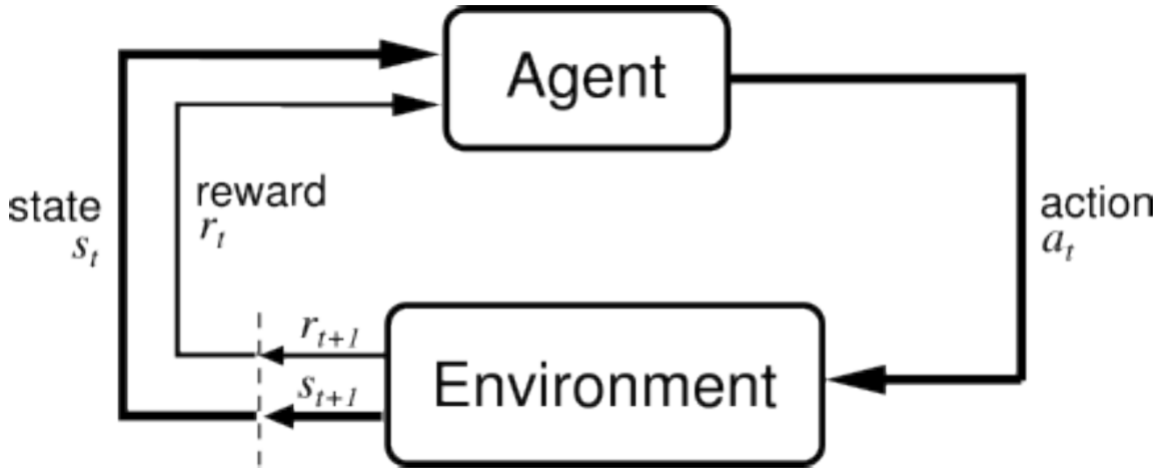
Examples

- Automated vehicle control/robotics
  - An unmanned helicopter learning to fly and perform stunts
- Game playing
  - Playing backgammon, Atari breakout, Tetris, Tic Tac Toe
- Medical treatment planning
  - Planning a sequence of treatments based on the effect of past treatments
- Chat bots
  - Agent figuring out how to make a conversation
  - Dialogue generation, natural language processing

Modeling foundation: MDP

- Markov Decision process: model for sequential dec
  - Past information is captured by **state**
  - Agent takes an action, observes **new state and reward** generated from a stochastic model
  - Objective is some aggregate function of the individual rewards

Sequential decisions
Reward signals
(partial labels)
Delayed feedback
Actions effect observations



Reinforcement learning

- Reinforcement learning ≡ MDP with unknown stochastic model
- Agent observes samples : rewards, state transition
- Learn a good strategy (policy) for the MDP
  - Implicitly or explicitly learn the model dynamically from observations

The algorithm design problem

Design a strategy for taking actions sequentially, after observing current state

- Generate good reward
- Generate informative sample observations

and converge to optimal strategy

Challenges:

- Complex combination of learning and optimization
- There may be a tradeoff between reward and information
- Scale: large number of states, need to use structure

Course Goals

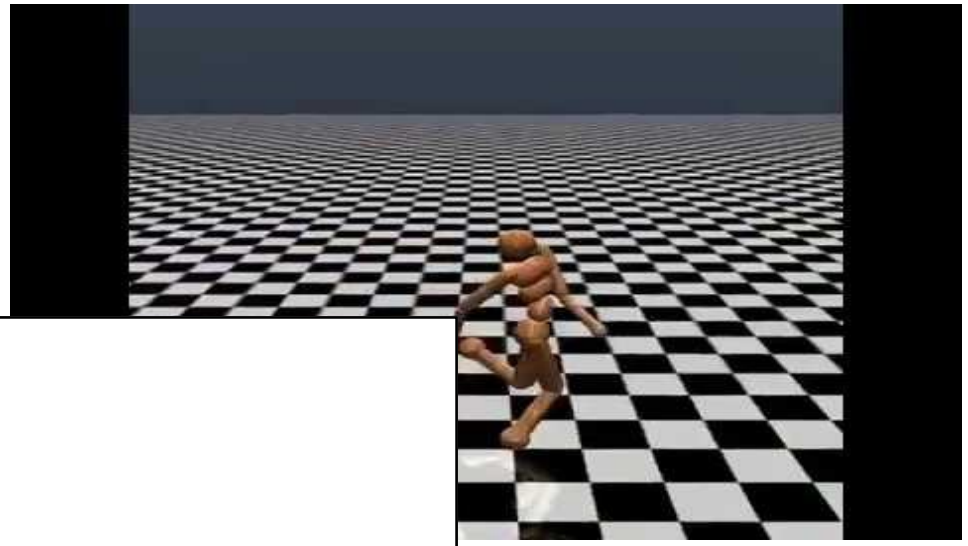
- Rigorous understanding of the MDP foundation:
  - Stochastic structure, algorithm design, convergence
- Conceptual understanding of recent algorithms for reinforcement learning
  - Mathematical insights into design principles
  - Occasional convergence results
- Ability to implement RL algorithms using some popular software platforms and simulators
  - Utilize Deep learning with tensorflow
  - OpenAI gym
- Ability to understand recent research papers
- New research!

Topics (tentative)

- **Introduction to MDP**: value-iteration, policy iteration, Q-value-iteration
- **Q-learning**: Tabular, function approximation
- **Deep Q-networks**: architecture, backpropagation, experience replay
- **Policy gradient methods**: Function approximation, Natural policy gradient, Trust region policy optimization, Actor critic methods,
- **Model-based RL**
- Further challenges: Exploration vs. exploitation, Adversarial training, Generalization, Multi-agent RL

Open AI gym
https://gym.openai.com/envs/

- Simulated environments for testing reinforcement learning algorithms

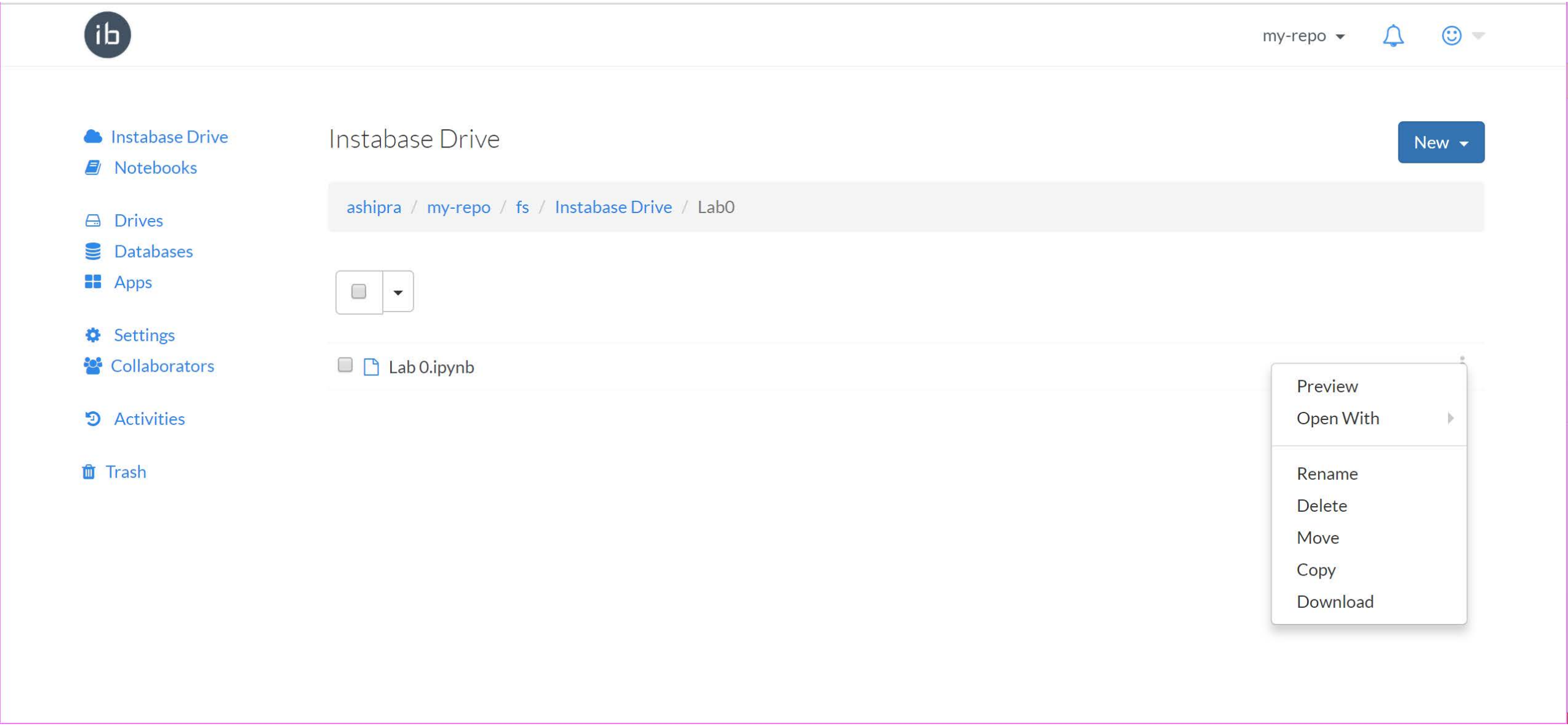


manoid-v1

Breakout-v0

Lab0

- Lab0 is a trial assignment (Jupyter notebook with skeleton code)
- Play with OpenAI gym environments, make changes to python code, plot the performance of random strategies.
- Submit using the link at the end



Open the notebook with Jupyter

Submit by clicking on the link at the bottom (After making any changes you want. You can submit multiple times. The version will be updated every time you submit.

Secure | https://www.instabase.com/user/ashipra-nb/notebooks/ashipra/my-repo/fs/Instabase%20Drive/Lab0/Lab%200.ipynb

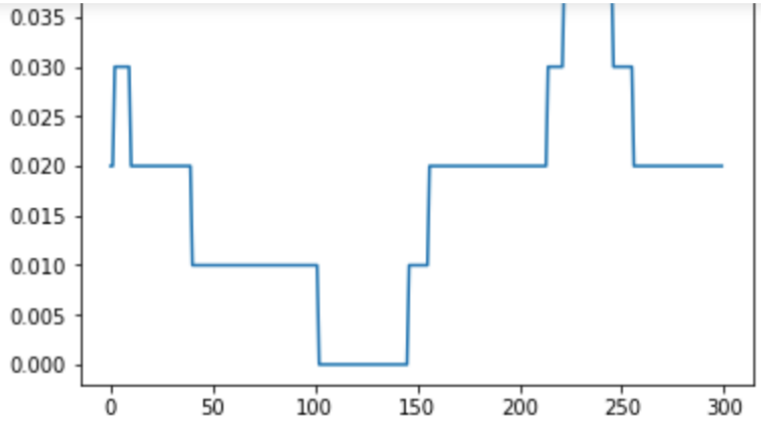
jupyter notebook

Lab 0 (autosaved)

File Edit View Insert Cell Kernel Widgets Help

Not Trusted | Python 3

Run | Markdown



In your programming assignments, you will be given a skeleton of code like above. You will be asked to make changes to the code to achieve specific tasks and then submit the notebook as your submission for the assignment. Make a submission using below. This will not be graded but will help us ensure that everything is set up correctly in your instabase account.

Submit it using the following link


ima...
021 3 1