Foundation of Machine Learning

Day 3: Machine Learning Deployment

Sieu Tran

Agenda

Day 3 of the Foundation of Machine Learning course will focus on ML deployment

- 1 Recap of Day 1 and Day 2
- 2 Introduction to ML Deployment
- Continuation of Day 1 Coding Assignment
- 4 Continuation of Day 2 Coding Assignment
- 5 Group Discussion
- 6 Classification Problems
- 7 Day 3 Coding Assignment
- 8 Closing Words



Copyright © 2020 Accenture. All rights reserved

Recap



Day 1 and Day 2 Coding Assignments Revisit

Working Session 1



Case study:

• United States COVID-19 Cases and Deaths by State over Time

Data:

- Source: https://catalog.data.gov/dataset/united-states-covid-19-cases-and-deaths-by-state-over-time
- Descriptions: https://data.cdc.gov/Case-Surveillance/United-States-COVID-19-Cases-and-Deaths-by-State-o/9mfq-cb36/data

Problems:

- Perform some exploratory analysis:
 - Which states has the most deaths?
 - Which month has the most deaths?
 - Are there any correlations?
 - What does the rate of death look like over time?
- Can you perform dimensional reduction on this dataset?
- Can you predict the rate of death in the upcoming month?
- Consider setting threshold for high/medium/low death rate for classification problem
- Example hypotheses:
 - States near the sea are more effected
 - States that are politically conservative have higher death rates

Get started:

- Python users:
 - Install Anaconda
 - Install Python 3.7
 - Create a virtual environment:
 - conda create -n yourenvname python=x.x anaconda
 - Install requirements: pandas, numpy, and sklearn to start

OR

- Use Google Collab
- R users:
 - Install RStudios
 - Install requirements: *tidyverse* and *gbm* to start



Working Session 2



Case study:

• Facebook ad and consumer behavior

Data:

• Source: https://www.kaggle.com/datasets/loveall/clicks-conversion-tracking

Problems:

- Read in the data and perform some exploratory analysis
- Create a new column for rate of click (Click/Impression).
- Build a predictive (regression) model to predict the rate of click using a selection of other features
 - Can you perform the train/test/validation split?
 - Can you compute the three/four types of errors we discussed today?
 - Can you look at the residual plots and diagnose the problems your model might be facing?
 - How can you make the model better?

Get started:

- Python users:
 - Install Anaconda
 - Install Python 3.7
 - Create a virtual environment:
 - conda create -n yourenvname python=x.x anaconda
 - Install requirements: pandas, numpy, and sklearn to start

OR

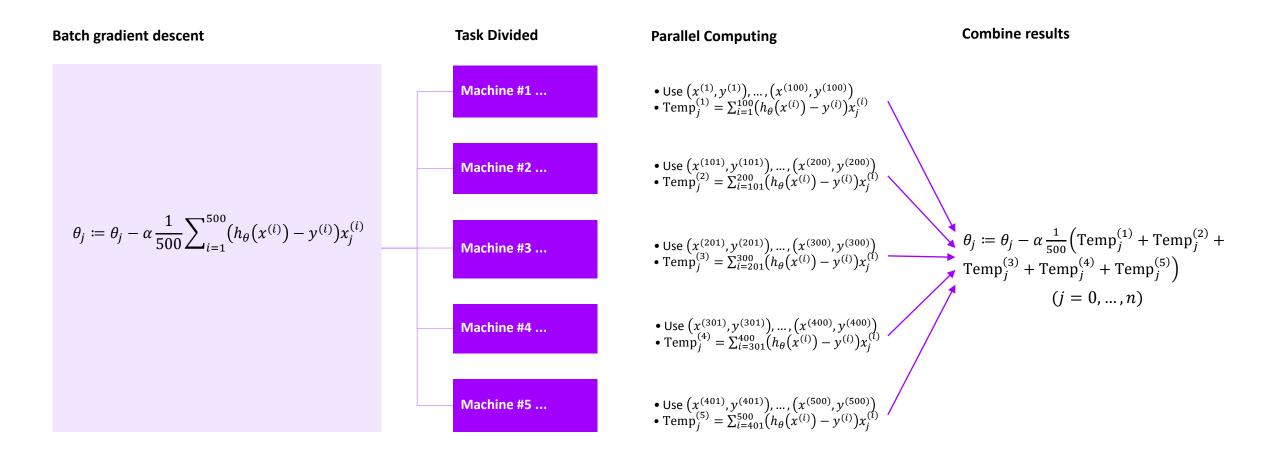
- Use Google Collab
- R users:
 - Install RStudios
 - Install requirements: *tidyverse* and *gbm* to start



Large Scale Machine Learning

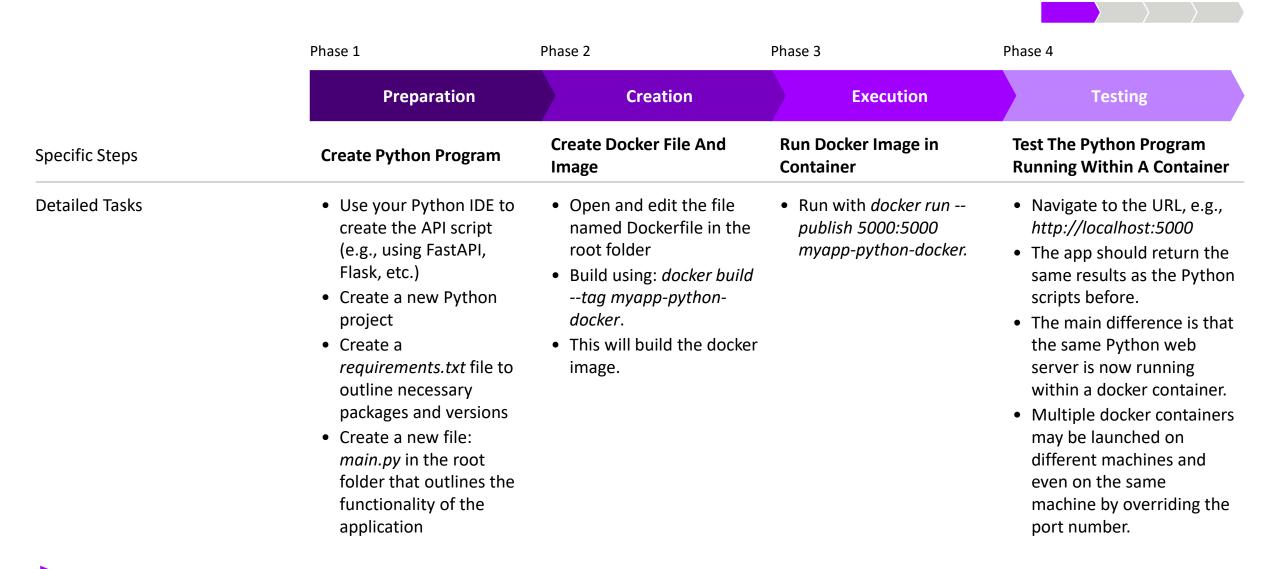
Map-reduce And Data Parallelism

Many learning algorithms can be expressed as computing sums of functions over the training set.



ML Deployment: Docker Example

Example Deployment With Docker



It's Coding Time!



Working Session 2



Case study:

• Deploy the models you have built from Day 1 and Day 2

Data:

- Source 1: https://catalog.data.gov/dataset/united-states-covid-19-cases-and-deaths-by-state-over-time
- Source 2: https://www.kaggle.com/datasets/loveall/clicks-conversion-tracking

Problems:

- Can you write them into functions instead of script?
 - Input: data
 - Output: predictions
- Can you clean up the code?
- Can you create the requirements.txt file?
- Can you check each package's versions?
- Can you collect the licenses for each packages and versions?

Get started:

- Python users:
 - Install Anaconda
 - Install Python 3.7
 - Create a virtual environment:
 - conda create -n yourenvname python=x.x anaconda
 - Install requirements: pandas, numpy, and sklearn to start

OR

- Use Google Collab
- R users:
 - Install RStudios
 - Install requirements: *tidyverse* and *gbm* to start



Closing Notes



Referenced Materials



- 1. Ng, Andrew. "Machine Learning By Prof. Andrew Ng", (2023), GitHub repository, https://github.com/vkosuri/CourseraMachineLearning.
- 2. Malik, Farhad. "Running Python In Docker Container". Medium, 03 Oct 2021, https://medium.com/fintechexplained/running-python-in-docker-container-58cda726d574.