STT 890 — Intro to MLOps — MLFlow "In-Class"

Due: Friday, September 26, 2025 11:59:59 PM

All submissions should be uploaded to the designated D2L dropbox.

Omeiza Olumoye

This work will involve model registries and versioning in MLFlow. It will use the dataset reg2, which a target y and 2 potential predictor variables x1 and x2. Use scikit-learn for the linear regression (for MLFlow, this is the model flavor) (so you will have to split up the data).

1. Read through the quickstart and model registry tutorials linked in this directory.
2. Build 3 models:
    a. Try a linear regression model using x1 only to predict y. Look at how well it does. Call this model model_1.
    b. Try a linear regression model using x2 only to predict y. Look at how well it does. Call this model model_2.
    c. Finally, use x1 and x2 to predict y. Compare the errors and R^2 values to the previous model. How do they compare?
3. To do the version control with MLFlow, follow these steps
    a. Create a repository in your Git for the models
    b. Put the models there.
    c. Register the models in MLFlow (you can use the above names or
4. Start and view the tracking server for the models.

Turn in a pdf of your notebook, along with a screen shot of the tracking server and the requirements.txt file.

1. Read through the quickstart and model registry tutorials linked in this directory. Done

2. Build 3 models:

   a. Try a linear regression model using x1 only to predict y. Look at how well it does. Call this model model_1. model_x1: y $\sim$ 0.0556x1 + 6.699

   b. Try a linear regression model using x2 only to predict y. Look at how well it does. Call this model model_2. model_x2: y $\sim$ 0.1332x2 + 6.8825

   c. Finally, use x1 and x2 to predict y. Compare the errors and R^2 values to the previous model. How do they compare? model_x1_x2: y $\sim$ 1.8603x1 + 0.9175x2 + 0.4954

	model_x1	model_x2	model_x1_x2
MAE	0.9189	0.9902	0.1936
MSE	1.2476	1.3742	0.0605
RMSE	1.1169	1.1723	0.2459
R^2	-197.6360	-8.5352	0.9469

Table 1: Errors and R^2

From Table 1, the third column which has the least values in all the errors signifying its superiority over models model_x1 and model_x2. model_x1_x2. As regards the R^2 value, the third model has almost perfect prediction, whereas the models model_x1 and model_x2 have an interpretability worse than the baseline model that barely predicts the mean. In the nutshell, the model model_x1_x2 is a much better model for this particular dataset.

3. To do the version control with MLFlow, follow these steps

   a. Create a repository in your Git for the models. Done

   b. Put the models there. Done

   c. Register the models in MLFlow (you can use the above names or ...) Done

4. Start and view the tracking server for the models. Done

Turn in a pdf of your notebook, along with a screen shot of the tracking server and the requirements.txt file. The files are tracking_server_shot_1of4.png $\cdots$ tracking_server_shot_4of4.png; and requirement.txt will be pushed to the github repository as well.