Anujit Nair

202418036

Assignment 2:

DVC Pipelines & Experiment Tracking

1. How does the DVC pipeline save time and reduce errors compared to running scripts manually?
   DVC automatically tracks dependencies between stages (data, code, parameters). When something changes, it only reruns the affected stages instead of the whole pipeline, saving time. It also ensures the correct order of execution, reducing human mistakes like using outdated data or models.

2. Which stages were automatically rerun when parameters changed?
   If preprocessing parameters change, DVC reruns preprocess -> train -> evaluate.
   If model parameters change, only train -> evaluate are rerun.
   If only evaluation code changes, then just evaluate is rerun.

3. How does DVC help in reproducing experiments and tracking metrics?
   DVC records data, code, parameters, and outputs for each experiment, making runs fully reproducible. It stores metrics (like accuracy, F1, hyperparameters) in JSON file. Commands like dvc exp show let you easily track and compare results across experiments.