CSE 234
Data Systems for Machine Learning

Arun Kumar

Review 2
1. [8pts] Briefly describe two technical debts in ML applications that can be mitigated with better end-to-end governance.

2. [4pts] What is the main reason for MLFlow’s architecture to differ so much from TFX’s?

3. [8pts] Briefly explain two advantages of using a dedicated feature store in a production ML application.

4. [8pts] Briefly describe a component each of MLFlow and TFX that are roughly analogous to each other. How does that component help mitigate technical debt in ML applications?
5. [10pts] Briefly explain one major way in which ML-based applications differ from deterministic software applications and why that can cause technical debt for applications. Give a concrete real-world ML application example where that debt manifests itself.

6. [10pts] Briefly explain one major way in which feature engineering is a major source of technical debt with a concrete real-world ML application example. Clearly describe the prediction task and datasets. Briefly explain how you can mitigate this debt.
7. **[10pts]** Briefly describe one concrete real-world ML application at a Web company where online prediction monitoring is critical for the business. Clearly describe the prediction task and datasets. Briefly explain one way a monitoring failure can hurt the application.

8. **[10pts]** Briefly describe one concrete real-world ML application where federated ML is likely to be beneficial for both end users and the organization. Clearly describe the prediction task and datasets. Justify why it is beneficial.