Summer Term 2019

Advanced Topics in Databases

Exercise 11

1. Tools Supporting Machine Learning
   Among the driving factors for the increased adoption of machine learning, we spoke in the lecture about the growing standardization in tools that are being used to support these applications.
   Please define for the following kinds of tools: What is their goal and what are their key features? What are some examples of them?

   - *Machine learning model management systems*
   - *In-database machine learning*
   - *Deep reinforcement learning frameworks*
   - *Deep learning frameworks*
   - *Tensor processing units*

2. Systems and ML

   In class we presented a canonical architecture for an AI system. Please discuss (a) the main components of the architecture, and (making reference to this architecture) (b) the cross-cutting concerns that motivate research at the intersection of data management and machine learning.

3. Basic Topics on Machine Learning for Data Management:
   In class we spoke about machine learning in production. In this context we highlighted guidelines for best practices and discussed about the ML-test score, according to industry practitioners (`https://developers.google.com/machine-learning/guides/rules-of-ml/`).
   Please answer the following questions considering such guidelines.

   (a) What is the difference between a *metric* and an *objective*? For learning how to build an index structure for range searches, what could be some metrics, and what could be some objective?

   (b) What is a *heuristic*? After giving this definition, explain rules #1, #3 and #7 from the rules of machine learning. For finding an item in a sorted list, give an example of 1 heuristic approach, and one case where it is not optimal.

   (c) Explain what is the *training-serving skew*? If you train a model to predict the runtime of a given operator, on a single-user system, over a chosen dataset and specific hardware; and you finally deploy the model on the same dataset and hardware, but varying number of users, what could be some causes of training-serving skew? If you train a model to forecast the arrival rate of queries to an e-commerce site during the months of September-October, and you then deploy it for the months of November-January, what could be some causes of training-serving skew?

Good Luck!