Questions to Evaluate a Machine Learning Project

This set of questions can be used during a machine learning course that introduces technical people who did not previously have a background in ML, and helps them understand if their project is well designed according to machine learning standards.

Infrastructure

- What is the time to train your model? Can you use streaming data to reduce the amount of time to train? Can you split over multiple instances to reduce the time to train?
- What infrastructure are you using to train your models? Is it the lowest possible cost?
- Where are you storing your data; is that the best solution?
- What devops framework do you have to continuously integrate changes as you make them?
- Where are you developing your model, and is it the best choice for your scenario?

Conceptual

- Does the data that you're using reflect the real world?
- What actually impacts the real world prediction problem, and is that in your data set?

Data transformation

- Did you normalize your data?
- Did you randomly shuffle your data?
- Did you remove any outliers for your data?
- How did you handle missing or nonsensical values in your data?
- How are you handling any sequential elements of your data set?

Method

- Which model did you select, and why?
- How are you evaluating your model?
- Is your model overfitting, and what are you doing to counteract that?
- What are the limitations of your model, and what are its strong points?