Data Literacy: Exercise 12

Jessica Bader (5624582) and Laura Haege (4179267)

February 2021

1 EXAMple 1: Name and describe four high-level stages of the machine learning deployment workflow.

1.) Data Management: This is the phase where the data is prepared for use. First it must be collected. Often, collected data is not in a clean or usable format, so it must be organized. It may need additional augmentation, such as labeling. Some other steps here may be helpful, such as centering the data. By the end of this step, the data should be ready to be learned from.
2.) Model Learning: This is the part most commonly thought of as "machine learning". The model is chosen and trained. Hyperparameters are tuned to produce the best output.
3.) Model Verification: This is where the model is fully tested to provide the best guarantee that it will work possible given the specific circumstances. Requirements are evaluated (including technical, business driven, etc.). If possible, a formal proof of correctness or convergence comes here. The model is formally tested.
4.) Model Deployment: Now the model is integrated into the product. It is monitored and updated as needed throughout the rest of the products lifespan.

2 EXAMple 2: Name three potential security threats to a machine learning software system.

Data poisoning (training data injected to corrupt the model), model stealing (stealing model parameters), and model inversion (accessing sensitive training data through model queries).