Machine Learning Engineering Bootcamp

Capstone: Scale Your Prototype with Large-Scale Data

Step 8

Learning Objective

- Ensure that the ML/DL prototype created can scale to large volumes of data.
- Analyze and explain the trade-offs required to scale an ML algorithm.

Criteria	Meets Expectations	Weightage
Time Estimates	15 - 20 Hours	
Completion	❏ The code is updated to GitHub. (1 point)	1 point
Process and understanding	❏ The submission shows that the student understands how to **scale** a machine learning or a deep learning model. (2 points) ❏ The scaled prototype can handle the complete dataset that the student has collected (even if the student has only used a sample, so far) and is capable of handling all of the data that a real-world version of the application would need to handle. (2 points) ❏ The submission demonstrates that the student made well-thought-out decisions about scaling their prototype: ❏ Choice of tools/libraries: scikit-learn, SparkML, TensorFlow, Keras, and PyTorch, etc. (2 points)	8 points

	❏ Choice of machine learning/deep learning technique (2 points)	
Presentation	❏ Well-documented GitHub repository and code. The Jupyter notebooks for the code provide step-by-step documentation that's easy to follow. (1 point)	1 point

Excellence:

1. *The student has designed their scaled prototype to work with web-scale data involving billions of data points.*
2. *The code is particularly clean and elegant.*