Machine Learning Engineering Bootcamp Capstone

Step 5: Data Wrangling

Summary

Time Estimate: 15 - 20 hours

Now, you'll apply some data wrangling techniques to your capstone dataset. As you're working in your Jupyter notebook, take notes documenting the data wrangling steps you took cleaning your dataset. Consider the following:

- What kind of cleaning steps did you perform?
- How did you deal with missing values, if any?
- Were there outliers? If so, how did you handle them?
- If your dataset is too large to work with, does it make sense to build your prototype on a smaller subset of the data?

Optional Step: Data Exploration

After you've obtained the dataset for your capstone project, cleaned, and wrangled it into a form that's ready for analysis, you will perform a preliminary exploration of the data. This exploratory data analysis (EDA) uses a combination of inferential statistics and data visualization to find interesting trends and identify significant features in the dataset. For example:

- Are there variables that are particularly significant in terms of explaining the answer to your project question?
- Are there strong correlations between pairs of independent variables or between an independent and dependent variable?

Project Submission

After completing the above, please submit a link to your Jupyter notebook in your Github repository that shows how you cleaned, wrangled, and (optionally) explored the data.

This step of your capstone project will be evaluated using this rubric.