IDSN 542: Machine Intelligence

Lab 5 Due: 10/5/25

Goal

You are going to find a dataset from Kaggle.com. Modify the code from the chapter 2 load data part 2 and 2b for the dataset that you choose

Setup

- Your Python file must begin with comments in the following format (replace the name and email with your actual information):

```
'''
Name
IDSN 542, Fall 2025
USC email
Lab 5
'''
```

Requirements

Choose a regression dataset from Kaggle.com. The dataset must have at least 1000 rows. You can find the regression datasets if you type in the word 'regression' in the search bar at Kaggle (https://www.kaggle.com/datasets).

Do any necessary data preparation. This includes:

- Taking care of any missing values
- Creating any new attributes that seem to make sense
- Converting any categorical data to numeric using OneHotEncoder or OrdinalEncodeer

Use the correlation function (corr()) to identify any linear relationships in your numerical attributes.

Write up a report as to what you did. This includes:

- Data preparation
- Screenshot of your linear correlation graph
- Analysis of what linear relationships exist in your data
- Why you used OrdinalEncoder or OneHotEndocer if you had categorical data
- If you created new attributes, why did you think that made sense

Submit your report and your Python code in a ZIP file and submit to Brightspace.