Model Analysis

This exercise aims to help you practice the notions and skills you have acquired in the course (including labs).

Main steps:

1. Choose a tabular classification dataset.
    - For example, the Bank Marketing Dataset or the Rain in Australia Dataset.
    - You can find other datasets in the UCI ML Repository, Kaggle, Penn Machine Learning Benchmarks.
    - You are welcome to use your own dataset as well.
2. Train a classification model on the dataset (you can choose any training algorithm of your preference).
3. Analyze the model to answer the questions below.
4. Create and submit a PDF document reporting on your answers using descriptive text and explanatory data visualizations
    - You are not required to use Python, sklearn, or Altair for this assignment. You may use the languages and libraries of your choosing.

Questions to answer about the model

Dataset:
- What dataset did you choose?
- What is the class distribution of this dataset?
- Do any features have interesting distributions or interesting relationships with other features?
- Did you do any feature engineering or feature selection? If so, what informed those decisions?
- [Feel free to pursue additional questions and report on the results]

Error analysis:
- What is the overall performance of the model?
- What kind of errors does the model make? What are the types of instances the model is confused about? What is the potential reason for the confusion?
- Does the model make mistakes in areas of the decision space where it has high confidence?
- Does the model make errors near or far from the decision boundary?
- How do the errors distribute? Are there specific subsets (areas of the decision space) where the model makes a higher number of errors?
- [Feel free to pursue additional questions and report on the results]

Model logic:

- What drives the model decisions? What are the most relevant features? What impact do those features have on the model's behavior?
- Does the logic used by the model make sense intuitively?
- Does the logic of the model differ in different data subsets?
- [Feel free to pursue additional questions and report on the results]

[Optional] Comparative analysis: build one or more alternate models and compare them with the original one you developed and analyzed in the previous steps.
- How do the models differ in terms of errors?
- How do the models differ in terms of logic?