Metadata

- Title: Wine Quality Analysis
- Authors: Rory Black, Gargee Jagtap, Ali Roth

Synopsis

source

This notebook contains an analysis and Bayesian Machine Learning methods to predict wine quality based on the physicochemical properties included in the dataset. This artifact was produced for DS6040 at the University of Virginia, taught by Dr. Don Brown in the Fall of 2022. The 3 contributors are students in the Residential Masters of Science in Data Science at the University's School of Data Science.

The Data for this project comes from wines in the northern region of Portugal and was downloaded from Kaggle on November 29, 2022. The files, along with more information on the data collection and variables included, can be found here: https://www.kaggle.com/datasets/rajyellow46/wine-quality. Please note that this data originates from the UCI Machine Learning Repository.

Installation

To run this code, you need to have python installed on your machine. Start by downloading the code file, and the 2 wine csv files: *WineQuality-RedWine.csv, and WineQuality-WhiteWine.csv*. Make sure all 3 of these files are in the same folder on your computer.

After doing the above, ensure that all the libraries at the top of the code file are installed on your machine. If not, install them by typing the following command in a code cell:

```
pip install *library*
```

The following libraries are needed:

- numpy
- pandas
- matplotlib
- pymc
- arviz
- sklearn
- seaborn

Once all the libraries have been installed, you can now run the notebook.

Bayesian Overview

This notebook implements four machine learning models using Hamiltonian Monte Carlo Sampling. The final model, *reduced_model* is used to make predictions on a test set pulled from that data in a 70/30 split. Following the predictions, the code runs a variational inference approach on the *reduced_model* to compare results to the ones obtained by sampling in interest of examining the use of this less accurate, computationally simpler approach.