MISSING DATA IMPUTATION

FOR SUPERVISED CLASSIFICATION

Jason Poulos and Rafael Valle

University of California, Berkeley

Supplementary Material

This file contains descriptive plots of missing data patterns in the benchmark datasets, as well as a description and plots of Bayesian hyperparameter optimization for training neural networks on the benchmark datasets.
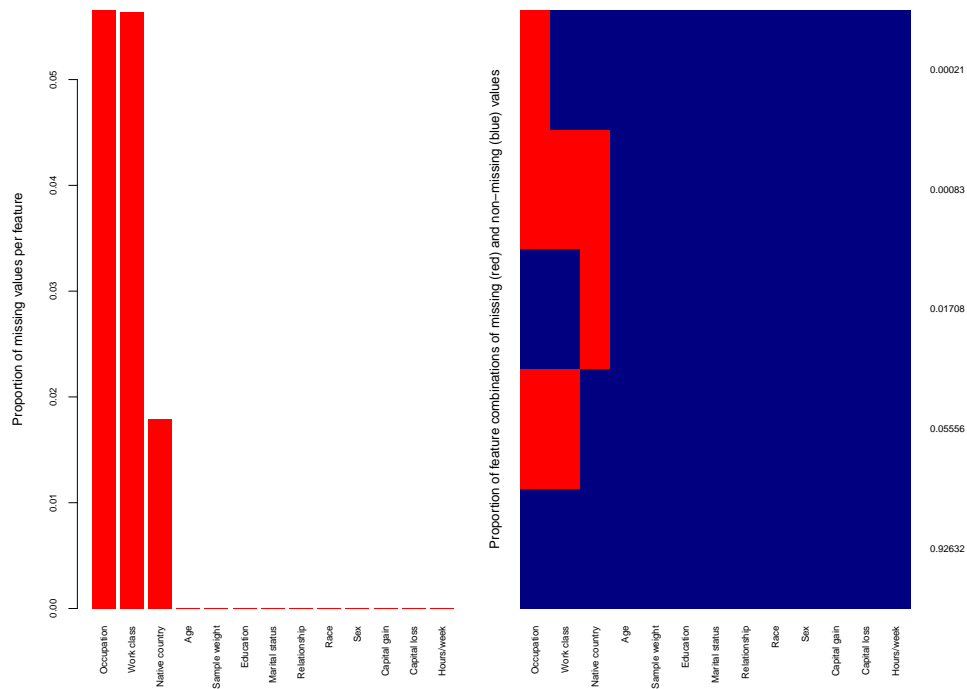
S1 Patterns of missing data

Figure 1: Histogram of proportion of missing values in each feature (Left) of Adult training set and aggregation plot of all existing combinations of missing and non-missing values in the samples (Right).
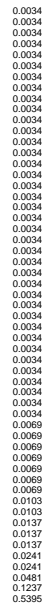
Figure 2: Histogram of proportion of missing values in each feature (Left) of CVRs training set and aggregation plot of all existing combinations of missing and non-missing values in the samples (Right).

S2   Bayesian hyperparameter optimization

The goal of Bayesian optimization is to choose a point in the hyperparameter space that appropriately balances information gain and exploitation. Figure 3 shows the exploration of hyperparameter space during Bayesian optimization for both Adult and CVRs datasets. Each circle represents a candidate neural network classifier trained on a differently imputed and perturbed dataset. More circles appear in the plot for CVRs simply due to the fact that the training set is smaller. We see that most of the candidate models use dropout and have an initial learning rate close to the maximum of 0.01. The plurality of candidate models appear to either have momentum (1) or not (0).

Adult dataset

CVRs dataset