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 and of Bayesian hyperparameter optimization for training neural networks on the benchmark datasets.
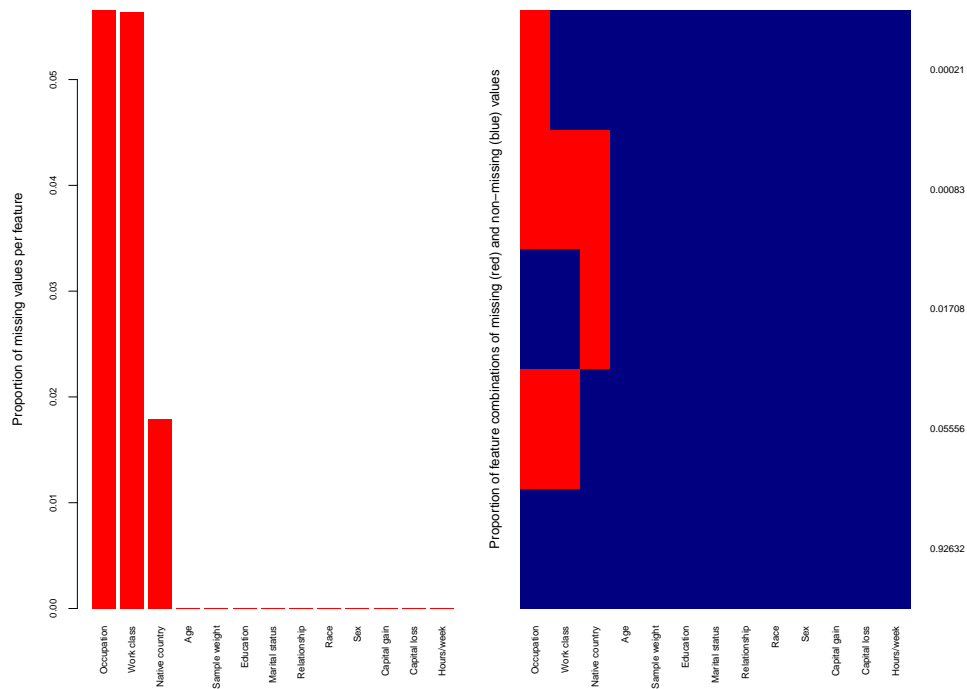
S1 Descriptive plots

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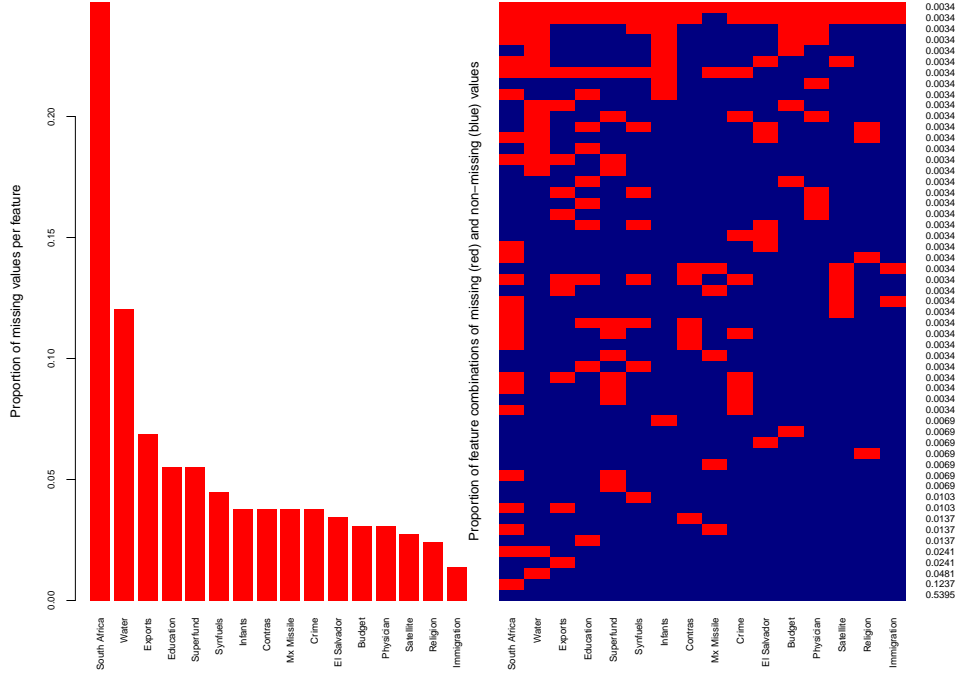
Figure 2: Histogram of proportion of missing values in each feature (Left) of CVRs training set and aggregation plot of all existing combinations of missing and non-missing values in the samples (Right).

Adult dataset

CVRs dataset