Missing Data Project Results

Anindro Bhattacharya

Winter 2022

1 NeuMiss Architecture

1.1 Real Missing Data

The NeuMiss architecture was trained on a diabetes dataset containing missing values with the following hyperparameters:

- depth = 9
- # of epochs = 100
- batch size = 10
- learning rate = 0.01/8

	Log Loss (Binary Cross Entropy Loss)
Neumann with residual connections	10.673
Neumann without residual connections	28.960

Table 1: Log loss for NeuMiss applied to diabetes dataset containing missing values

1.2 Masking a Complete Dataset

The NeuMiss architecture was trained on a diabetes dataset containing no missing values with the following hyperparameters:

- depth = 9
- # of epochs = 100
- batch size = 10
- learning rate = 0.01/4

Random Forest Classifier was trained with `n_informative=4`, `n_redundant=0`, `max_depth=8` using `sklearn`.

	Log Loss (Binary Cross Entropy Loss)	AUROC
Neumann with residual connections	5.589	0.5
Neumann without residual connections	38.149	0.536
Random Forest Classifier	31.169	0.5

Table 2: Log loss and AUROC for NeuMiss and Random Forest Classifier applied to diabetes dataset without any data amputation

	MCAR	MAR	MNAR
Neumann with residual connections	0.457	0.588	0.597
Neumann without residual connections	0.506	0.503	0.707

Table 3: AUROC for NeuMiss applied to diabetes dataset masked with various mechanisms of missingness

1.3 Estimating $(\Sigma_{obs(m)})^{-1}$

The actual covariance matrix is given below:

$$(\Sigma_{obs(m)})^{-1} = \begin{bmatrix} -0.074430585 & -0.4932546 & -0.16943908 & 0.281695 \\ -0.2888431 & -0.3939461 & 0.14461023 & 0.4538434 \\ 0.19937307 & -0.23638344 & -0.36582643 & 0.48601776 \\ 0.47908944 & -0.08196759 & -0.26269215 & -0.36598015 \end{bmatrix} \tag{1}$$

The following is the estimated covariance matrix with MCAR masking:

$$(\Sigma_{obs(m)})^{-1} \approx \begin{bmatrix} 0.11353505 & -0.27601248 & -0.24196988 & 0.013374865 \\ 0.46621287 & 0.41319656 & 0.2962613 & -0.4033689 \\ 0.45569807 & -0.010969877 & 0.06865537 & -0.058603525 \\ -0.18620259 & -0.46399462 & -0.13635439 & 0.39681786 \end{bmatrix} \tag{2}$$

The following is the estimated covariance matrix with MAR masking:

$$(\Sigma_{obs(m)})^{-1} \approx \begin{bmatrix} 0.23384035 & 0.14557779 & -0.4143983 & -0.011400878 \\ -0.09307128 & 0.32880175 & -0.30733913 & -0.17595828 \\ -0.22125101 & -0.3588832 & 0.22725868 & 0.1586389 \\ 0.3361441 & 0.03262663 & -0.08088577 & -0.1738466 \end{bmatrix} \tag{3}$$

The following is the estimated covariance matrix with MNAR masking:

$$(\Sigma_{obs(m)})^{-1} \approx \begin{bmatrix} -0.39501578 & -0.304408 & 0.39646292 & -0.46904248 \\ -0.23762369 & -0.0014175177 & 0.4798187 & -0.36039358 \\ 0.08218175 & 0.4626212 & 0.009880781 & -0.41792983 \\ 0.4573294 & 0.47776115 & -0.30522943 & -0.015576959 \end{bmatrix} \tag{4}$$

	MCAR	MAR	MNAR
MSE	0.245	0.160	0.241

Table 4: Mean squared error (MSE) of estimated $(\Sigma_{obs(m)})^{-1}$ with MCAR, MAR and MNAR masking

2 Doubly Robust Estimators

The goal of a doubly robust estimator is to estimate $E(Y)$ where Y is a scalar outcome which is missing some subjects. $\mathbf{V}$ is the set of always observed baseline variables and Δ is the missingness indicator (i.e., Y is missing if $\Delta = 0$ and Y is observed if $\Delta = 1$.

With the estimators implemented, three assumptions are made:

- Y is MAR, with a missingness rate of 0.4
- $P(\Delta = 1|Y, \mathbf{V}) = P(\Delta = 1|\mathbf{V}) \equiv \pi(\mathbf{V}) > 0$
- $\mu = E(Y) = E\{E(Y|\mathbf{V})\}$

2.1 Naive Mean Calculation

This is when $E(Y)$ is calculated using the following equation:

$$\mu = \frac{1}{n}\sum_i Y_i \tag{5}$$

2.2 Horvitz-Thompson Estimator

First, a propensity score model($\pi(\mathbf{V})$) is fit using logistic regression to estimate the likelihood of Y being missing given $\mathbf{V}$. Then, the mean is estimated using the following equation:

$$\hat{\mu}_{HT} = \frac{1}{n}\sum_i \frac{\Delta_i Y_i}{\pi(\mathbf{V}_i; \hat{\alpha})} \tag{6}$$

where $\hat{\alpha}$ is the maximum likelihood estimator of α.

2.3 Outcome Regression Estimator

First, a model $\Psi\{s(\mathbf{V}; \beta)\}$ is fit using linear regression for $E(Y|\Delta = 1, \mathbf{V})$, where $s(\mathbf{V}; \beta)$ is a linear regression function and Ψ^{-1} is a known link function (in this case, the identity function was used). Then, the mean is estimated using the following equation:

$$\hat{\mu}_{OR} = \frac{1}{n}\sum_i \Psi\{s(\mathbf{V}; \beta)\} \tag{7}$$

2.4 The Doubly Robust Estimator

Note, the same notations are used in this section as in the Horvitz-Thompson and Outcome Regression estimators. First, $E(Y|\Delta = 1, \mathbf{V})$ is modeled as $e(\mathbf{V}; \beta, \phi) = \Psi\{s(\mathbf{V}; \beta) + \phi\pi^{-1}(\mathbf{V}; \hat{\alpha})\}$, where $\phi = Y - s(\mathbf{V}_i; \beta)$. Then, the mean is estimated using the following equation:

$$\hat{\mu}_{dr} = \Psi\{s(\mathbf{V}_i; \hat{\beta}) + \phi\pi^{-1}(\mathbf{V}_i; \hat{\alpha})\} \tag{8}$$

2.5 Comparing Estimators

The naively calculated mean on the complete data is 0.03116.

Estimator	$\hat{E(Y)}$	Bias
Horvitz-Thompson	0.07259	-0.04143
Outcome Regression	0.03214	-0.00098
Doubly Robust	0.03043	0.00072

Table 5: Estimates of $E(Y)$ with correctly specified π and s models

Models are incorrectly specified by fitting them on the original dataset masked with MCAR missingness and missingness rate of 0.8.

Estimator	$\hat{E(Y)}$	Bias
Horvitz-Thompson	0.05721	-0.02605
Doubly Robust	0.03043	0.00072

Table 6: Estimates of $E(Y)$ with incorrectly specified π and correctly specified s models

Estimator	$\hat{E(Y)}$	Bias
Outcome Regression	0.00970	0.02145
Doubly Robust	0.06804	-0.01688

Table 7: Estimates of $E(Y)$ with correctly specified π and incorrectly specified s models

Estimator	$\hat{E(Y)}$	Bias
Doubly Robust	0.06804	-0.01688

Table 8: Estimates of $E(Y)$ with incorrectly specified π and incorrectly specified s models

3 Using Doubly Robust Estimators on Imputed Datasets

3.1 Method

The performance of each estimator was gauged by comparing the estimated means of the datasets. With the exception of the naive calculation and MICE, the estimators employed (i.e., NeuMiss and Doubly Robust) are designed to only predict the values of one column (the outcome Y). To account for this, an iterative process was used. In each iteration, a different feature was treated as the outcome Y (the number of iterations equals to the number of features). The calculated means from each iteration were then averaged to estimate the average of all the estimated values.

3.1.1 Estimators

- **Naive Calculation:** Calculate the mean of all the values in the umasked data.
- **NeuMiss (no DR):** Same as NeuMiss (Bayes' predictor)
- **MICE (no DR):** Mean of dataset imputed with MICE algorithm
- **NeuMiss (DR):** Doubly robust estimator with outcome regression model fit on data imputed using NeuMiss
- **MICE (DR):** Doubly robust estimator with outcome regression model fit on data imputed using MICE

3.2 Results

Using a doubly robust estimator reduces the bias with NeuMiss only when the samples to features ratio is high (same limitation as when NeuMiss is used for one outcome). In the results, it was revealed that MICE performs better than NeuMiss under all conditions.

Estimator	Mean	Bias
Naive Calculation	-0.06828	N/A
NeuMiss (no DR)	0.99043	-1.05872
MICE (no DR)	0.31918	-0.38747
NeuMiss (DR)	-0.90039	0.83210
MICE (DR)	0.06194	-0.13023

Table 9: Estimates of mean with various estimators with 5 features and 12000 samples.

Estimator	Mean	Bias
Naive Calculation	0.37006	N/A
NeuMiss (no DR)	3.34941	-2.97935
MICE (no DR)	0.68986	-0.31980
NeuMiss (DR)	-2.77694	3.14700
MICE (DR)	0.41624	-0.04618

Table 10: Estimates of mean with various estimators with 5 features and 1000 samples.

Estimator	Mean	Bias
Naive Calculation	-1.03173	N/A
NeuMiss (no DR)	-1.49641	0.46467
MICE (no DR)	-0.56546	-0.46627
NeuMiss (DR)	0.70930	-1.74104
MICE (DR)	-1.16382	0.13208

Table 11: Estimates of mean with various estimators with 5 features and 10 samples.

4 Variational Autoencoders

A Variational Autoencoder for Missing Data was implemented using code in `https://github.com/ProcessMonitoringStellenboschUniversity/IFAC-VAE-Imputation` from Variational Autoencoders for Missing Data Imputation with Application to a Simulated Milling Circuit article. Data was amputed with MCAR, MAR and MNAR missingness with various values of `missing-rate` and `prop-for-masking`, where appropriate. Note, `prop-for-masking` is the proportion of variables with no missing values that will be used for the logistic masking model.

The following training parameters were used:

- number of epochs = 500
- batch size = 250
- learning rate = 0.001

The reconstruction errors with VAEs when dealing with the various mechanisms of missingness were recorded:

Missingness Mechanism	Reconstruction Error (VAE)	Reconstruction Error (Mean)
MCAR	0.4139	0.8566
MAR	1.2622	0.9296
MNAR	0.6724	0.9299

Table 12: Reconstruction loss with missing-rate = 0.3 and prop-for-masking = 0.5

Missingness Mechanism	Reconstruction Error (VAE)	Reconstruction Error (Mean)
MCAR	0.5168	0.8468
MAR	0.7436	0.7783
MNAR	0.6473	0.8297

Table 13: Reconstruction loss with missing-rate = 0.5 and prop-for-masking = 0.7

Missingness Mechanism	Reconstruction Error (VAE)	Reconstruction Error (Mean)
MCAR	0.5724	0.8480
MAR	1.5578	0.7938
MNAR	0.7761	0.7761

Table 14: Reconstruction loss with missing-rate = 0.5 and prop-for-masking = 0.5

It was observed that when `missing-rate` was high, the reconstruction error with VAE was greater than the reconstruction error with mean imputation

Missingness Mechanism	Reconstruction Error (VAE)	Reconstruction Error (Mean)
MCAR	0.5238	0.8455
MAR	1.5861	0.8176
MNAR	0.7280	0.8794

Table 15: Reconstruction loss with missing-rate = 0.5 and prop-for-masking = 0.3

Missingness Mechanism	Reconstruction Error (VAE)	Reconstruction Error (Mean)
MCAR	0.5376	0.8478
MAR	0.9270	0.8584
MNAR	0.6831	0.8205

Table 16: Reconstruction loss with missing-rate = 0.5 and prop-for-masking = 0.2

Missingness Mechanism	Reconstruction Error (VAE)	Reconstruction Error (Mean)
MCAR	0.9169	0.8484
MAR	0.9506	0.7175
MNAR	0.8621	0.8136

Table 17: Reconstruction loss with missing-rate = 0.7 and prop-for-masking = 0.7

accross all three mechanisms of missingess, regardless of `prop-for-masking` values. When `missing-rate` is low(er), VAE performs better than mean imputation with MCAR and MNAR missingness but worse with MAR missingness.