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 0.2
- $P(\Delta = 1|Y, \mathbf{V}) = P(\Delta = 1|\mathbf{V}) \equiv \pi(\mathbf{V}) > 0$
- $\mu = E(Y) = E\{E(Y|\mathbf{V})\}$

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

Estimator	$\hat{E(Y)}$	Bias
Naive Mean Calculation	-0.033967	N/A
Horvitz-Thompson	-0.024974	-0.009
Outcome Regression	-0.040475	0.007
Doubly Robust	-0.030882	-0.003

Table 5: Estimates of $E(Y)$ with three estimators