A Matrix Factorization Approach to Multiple Imputation

Knowledge Lab Team Presentation

Nandana Sengupta
(co-authors: Madeleine Udell, James Evans, Nati Srebro)

February 10, 2016

Introduction

Introduction

- Missing data arise in almost all empirical analysis.

Introduction

- Missing data arise in almost all empirical analysis.
- Distracts from main goal of study.

Introduction

- Missing data arise in almost all empirical analysis.
- Distracts from main goal of study.
- Social science

Introduction

- Missing data arise in almost all empirical analysis.
- Distracts from main goal of study.
- Social science
    - opinion surveys

Introduction

- Missing data arise in almost all empirical analysis.
- Distracts from main goal of study.
- Social science
  - opinion surveys
  - longitudnal surveys

Introduction

- Missing data arise in almost all empirical analysis.
- Distracts from main goal of study.
- Social science
  - opinion surveys
  - longitudnal surveys
- Ad-hoc methods

Introduction

- Missing data arise in almost all empirical analysis.
- Distracts from main goal of study.
- Social science
  - opinion surveys
  - longitudnal surveys
- Ad-hoc methods
  - Complete case analysis (fully observed rows).

Introduction

- Missing data arise in almost all empirical analysis.
- Distracts from main goal of study.
- Social science
  - opinion surveys
  - longitudnal surveys
- Ad-hoc methods
  - Complete case analysis (fully observed rows).
  - Available case analysis (fully observed columns).

Introduction

- Missing data arise in almost all empirical analysis.
- Distracts from main goal of study.
- Social science
  - opinion surveys
  - longitudnal surveys
- Ad-hoc methods
  - Complete case analysis (fully observed rows).
  - Available case analysis (fully observed columns).
  - Mean Imputation.

Introduction

- Missing data arise in almost all empirical analysis.
- Distracts from main goal of study.
- Social science
    - opinion surveys
    - longitudnal surveys
- Ad-hoc methods
    - Complete case analysis (fully observed rows).
    - Available case analysis (fully observed columns).
    - Mean Imputation.
- Concerns about validity of inferences.

Introduction

- Missing data arise in almost all empirical analysis.
- Distracts from main goal of study.
- Social science
  - opinion surveys
  - longitudnal surveys
- Ad-hoc methods
  - Complete case analysis (fully observed rows).
  - Available case analysis (fully observed columns).
  - Mean Imputation.
- Concerns about validity of inferences.
- Types of Missing Data

Introduction

- Missing data arise in almost all empirical analysis.
- Distracts from main goal of study.
- Social science
  - opinion surveys
  - longitudnal surveys
- Ad-hoc methods
  - Complete case analysis (fully observed rows).
  - Available case analysis (fully observed columns).
  - Mean Imputation.
- Concerns about validity of inferences.
- Types of Missing Data
  - Missing Completely at Random.

Introduction

- Missing data arise in almost all empirical analysis.
- Distracts from main goal of study.
- Social science
  - opinion surveys
  - longitudnal surveys
- Ad-hoc methods
  - Complete case analysis (fully observed rows).
  - Available case analysis (fully observed columns).
  - Mean Imputation.
- Concerns about validity of inferences.
- Types of Missing Data
  - Missing Completely at Random.
  - Missingness depends on unobservables.

Introduction

- Missing data arise in almost all empirical analysis.
- Distracts from main goal of study.
- Social science
  - opinion surveys
  - longitudnal surveys
- Ad-hoc methods
  - Complete case analysis (fully observed rows).
  - Available case analysis (fully observed columns).
  - Mean Imputation.
- Concerns about validity of inferences.
- Types of Missing Data
  - Missing Completely at Random.
  - Missingness depends on unobservables.
  - Missing at Random.

Multiple Imputation

Multiple Imputation

- Rubin (1976), Schafer (1998), Van Buuren et al (1999), King et al (2000, 2015)

Multiple Imputation

- Rubin (1976), Schafer (1998), Van Buuren et al (1999), King et al (2000, 2015)
- Idea: Analysis should reflect uncertainty inherent in imputation.

Multiple Imputation

- Rubin (1976), Schafer (1998), Van Buuren et al (1999), King et al (2000, 2015)
- Idea: Analysis should reflect uncertainty inherent in imputation.
- Complete data D (dimension $n \times p$), observed data D^{obs}, Missingness Matrix M

Multiple Imputation

- Rubin (1976), Schafer (1998), Van Buuren et al (1999), King et al (2000, 2015)
- Idea: Analysis should reflect uncertainty inherent in imputation.
- Complete data D (dimension $n \times p$), observed data D^{obs}, Missingness Matrix M
- Assumption 1: Missing at Random: $P(M|D) = P(M|D^{obs})$

Multiple Imputation

- Rubin (1976), Schafer (1998), Van Buuren et al (1999), King et al (2000, 2015)
- Idea: Analysis should reflect uncertainty inherent in imputation.
- Complete data D (dimension $n \times p$), observed data D^{obs}, Missingness Matrix M
- Assumption 1: Missing at Random: $P(M|D) = P(M|D^{obs})$
- Assumption 2: Distributional $D \sim N_p(\mu, \Sigma)$.

Multiple Imputation

- Rubin (1976), Schafer (1998), Van Buuren et al (1999), King et al (2000, 2015)
- Idea: Analysis should reflect uncertainty inherent in imputation.
- Complete data D (dimension $n \times p$), observed data D^{obs}, Missingness Matrix M
- Assumption 1: Missing at Random: $P(M|D) = P(M|D^{obs})$
- Assumption 2: Distributional $D \sim N_p(\mu, \Sigma)$.
- 3 stage scheme

Multiple Imputation

- Rubin (1976), Schafer (1998), Van Buuren et al (1999), King et al (2000, 2015)
- Idea: Analysis should reflect uncertainty inherent in imputation.
- Complete data D (dimension $n \times p$), observed data D^{obs}, Missingness Matrix M
- Assumption 1: Missing at Random: $P(M|D) = P(M|D^{obs})$
- Assumption 2: Distributional $D \sim N_p(\mu, \Sigma)$.
- 3 stage scheme
  - Imputation : Expectation Maximization, Chained equations.

Multiple Imputation

- Rubin (1976), Schafer (1998), Van Buuren et al (1999), King et al (2000, 2015)
- Idea: Analysis should reflect uncertainty inherent in imputation.
- Complete data D (dimension $n \times p$), observed data D^{obs}, Missingness Matrix M
- Assumption 1: Missing at Random: $P(M|D) = P(M|D^{obs})$
- Assumption 2: Distributional $D \sim N_p(\mu, \Sigma)$.
- 3 stage scheme
  - Imputation : Expectation Maximization, Chained equations.
  - Analysis

Multiple Imputation

- Rubin (1976), Schafer (1998), Van Buuren et al (1999), King et al (2000, 2015)
- Idea: Analysis should reflect uncertainty inherent in imputation.
- Complete data D (dimension $n \times p$), observed data D^{obs}, Missingness Matrix M
- Assumption 1: Missing at Random: $P(M|D) = P(M|D^{obs})$
- Assumption 2: Distributional $D \sim N_p(\mu, \Sigma)$.
- 3 stage scheme
  - Imputation : Expectation Maximization, Chained equations.
  - Analysis
  - Combining Results

Multiple Imputation

- Rubin (1976), Schafer (1998), Van Buuren et al (1999), King et al (2000, 2015)
- Idea: Analysis should reflect uncertainty inherent in imputation.
- Complete data D (dimension $n \times p$), observed data D^{obs}, Missingness Matrix M
- Assumption 1: Missing at Random: $P(M|D) = P(M|D^{obs})$
- Assumption 2: Distributional $D \sim N_p(\mu, \Sigma)$.
- 3 stage scheme
  - Imputation : Expectation Maximization, Chained equations.
  - Analysis
  - Combining Results
- 'R' Packages: Amelia, MICE, MI.

Multiple Imputation

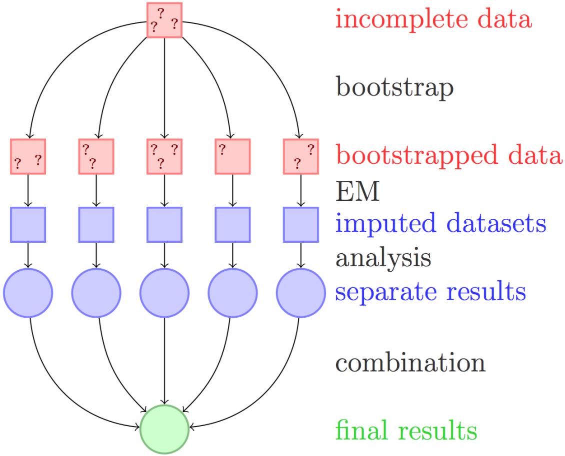


Image credit: https://cran.r-project.org/web/packages/Amelia/vignettes/amelia.pdf

Matrix Factorization: Generalized Low Rank Models

Matrix Factorization: Generalized Low Rank Models

- Low Rank and Low Norm approaches.

Matrix Factorization: Generalized Low Rank Models

- Low Rank and Low Norm approaches.
- Srebro (2004), Udell et al (2014).

Matrix Factorization: Generalized Low Rank Models

- Low Rank and Low Norm approaches.
- Srebro (2004), Udell et al (2014).
- Approximate matrix D (dimension $n \times p$) by $X'Y$.

Matrix Factorization: Generalized Low Rank Models

- Low Rank and Low Norm approaches.
- Srebro (2004), Udell et al (2014).
- Approximate matrix D (dimension $n \times p$) by $X'Y$.
- minimize $\sum_{i,j} L_{i,j}(x_i y_j, d_{ij}) + \gamma \sum_{i=1}^{n} r_i(x_i) + \gamma \sum_{j=1}^{p} r_j(y_j)$.

Matrix Factorization: Generalized Low Rank Models

- Low Rank and Low Norm approaches.
- Srebro (2004), Udell et al (2014).
- Approximate matrix D (dimension $n \times p$) by $X'Y$.
- minimize $\sum_{i,j} L_{i,j}(x_i y_j, d_{ij}) + \gamma \sum_{i=1}^{n} r_i(x_i) + \gamma \sum_{j=1}^{p} r_j(y_j)$.
    - L: Loss function (over columns) – quadratic, ordinal hinge, logistic, classification error etc.

Matrix Factorization: Generalized Low Rank Models

- Low Rank and Low Norm approaches.
- Srebro (2004), Udell et al (2014).
- Approximate matrix D (dimension $n \times p$) by $X'Y$.
- minimize $\sum_{i,j} L_{i,j}(x_i y_j, d_{ij}) + \gamma \sum_{i=1}^{n} r_i(x_i) + \gamma \sum_{j=1}^{p} r_j(y_j)$.
  - L: Loss function (over columns) – quadratic, ordinal hinge, logistic, classification error etc.
  - $r(.)$: regularization functions – trace norm, max norm etc.

Matrix Factorization: Generalized Low Rank Models

- Low Rank and Low Norm approaches.
- Srebro (2004), Udell et al (2014).
- Approximate matrix D (dimension $n \times p$) by $X'Y$.
- minimize $\sum_{i,j} L_{i,j}(x_i y_j, d_{ij}) + \gamma \sum_{i=1}^{n} r_i(x_i) + \gamma \sum_{j=1}^{p} r_j(y_j)$.
  - L: Loss function (over columns) – quadratic, ordinal hinge, logistic, classification error etc.
  - $r(.)$: regularization functions – trace norm, max norm etc.
  - X, Y : SVD good initialization.

Matrix Factorization: Generalized Low Rank Models

- Low Rank and Low Norm approaches.
- Srebro (2004), Udell et al (2014).
- Approximate matrix D (dimension $n \times p$) by $X'Y$.
- minimize $\sum_{i,j} L_{i,j}(x_i y_j, d_{ij}) + \gamma \sum_{i=1}^{n} r_i(x_i) + \gamma \sum_{j=1}^{p} r_j(y_j)$.
  - L: Loss function (over columns) – quadratic, ordinal hinge, logistic, classification error etc.
  - $r(.)$: regularization functions – trace norm, max norm etc.
  - X, Y : SVD good initialization.
  - k, γ: chosen via crossvalidation.

Matrix Factorization: Generalized Low Rank Models

- Low Rank and Low Norm approaches.
- Srebro (2004), Udell et al (2014).
- Approximate matrix D (dimension $n \times p$) by $X'Y$.
- minimize $\sum_{i,j} L_{i,j}(x_i y_j, d_{ij}) + \gamma \sum_{i=1}^{n} r_i(x_i) + \gamma \sum_{j=1}^{p} r_j(y_j)$.
  - L: Loss function (over columns) – quadratic, ordinal hinge, logistic, classification error etc.
  - $r(.)$: regularization functions – trace norm, max norm etc.
  - X, Y : SVD good initialization.
  - k, γ: chosen via crossvalidation.
- Low Norm Models: $r(x)$.

Matrix Factorization: Generalized Low Rank Models

- Low Rank and Low Norm approaches.
- Srebro (2004), Udell et al (2014).
- Approximate matrix D (dimension $n \times p$) by $X'Y$.
- minimize $\sum_{i,j} L_{i,j}(x_i y_j, d_{ij}) + \gamma \sum_{i=1}^{n} r_i(x_i) + \gamma \sum_{j=1}^{p} r_j(y_j)$.
  - L: Loss function (over columns) – quadratic, ordinal hinge, logistic, classification error etc.
  - $r(.)$: regularization functions – trace norm, max norm etc.
  - X, Y : SVD good initialization.
  - k, γ: chosen via crossvalidation.
- Low Norm Models: $r(x)$.
- Low Rank Models: $Rank(X'Y) \leq k$.

Matrix Factorization: Generalized Low Rank Models

- Low Rank and Low Norm approaches.
- Srebro (2004), Udell et al (2014).
- Approximate matrix D (dimension $n \times p$) by $X'Y$.
- minimize $\sum_{i,j} L_{i,j}(x_i y_j, d_{ij}) + \gamma \sum_{i=1}^{n} r_i(x_i) + \gamma \sum_{j=1}^{p} r_j(y_j)$.
  - L: Loss function (over columns) – quadratic, ordinal hinge, logistic, classification error etc.
  - $r(.)$: regularization functions – trace norm, max norm etc.
  - X, Y : SVD good initialization.
  - k, γ: chosen via crossvalidation.
- Low Norm Models: $r(x)$.
- Low Rank Models: $Rank(X'Y) \leq k$.
- Low Rank, Low Norm Models: Both

Matrix Factorization: Generalized Low Rank Models

- Low Rank and Low Norm approaches.
- Srebro (2004), Udell et al (2014).
- Approximate matrix D (dimension $n \times p$) by $X'Y$.
- minimize $\sum_{i,j} L_{i,j}(x_i y_j, d_{ij}) + \gamma \sum_{i=1}^{n} r_i(x_i) + \gamma \sum_{j=1}^{p} r_j(y_j)$.
  - L: Loss function (over columns) – quadratic, ordinal hinge, logistic, classification error etc.
  - $r(.)$: regularization functions – trace norm, max norm etc.
  - X, Y : SVD good initialization.
  - k, γ: chosen via crossvalidation.
- Low Norm Models: $r(x)$.
- Low Rank Models: $Rank(X'Y) \leq k$.
- Low Rank, Low Norm Models: Both
- Julia Implementation: LowRankModels

Interpretations: Generalized Low rank Models

Interpretations: Generalized Low rank Models

- Low dimensional embedding

Interpretations: Generalized Low rank Models

- Low dimensional embedding
- Latent Variables

Interpretations: Generalized Low rank Models

- Low dimensional embedding
- Latent Variables
- Compression

Interpretations: Generalized Low rank Models

- Low dimensional embedding
- Latent Variables
- Compression
- Denoising

Interpretations: Generalized Low rank Models

- Low dimensional embedding
- Latent Variables
- Compression
- Denoising
- Probabilistic Interpretation

Interpretations: Generalized Low rank Models

- Low dimensional embedding
- Latent Variables
- Compression
- Denoising
- **Probabilistic Interpretation** $\Leftarrow$ Equivalent to Multiple Imputation assumption when full rank.

Empirical Applications

- General Social Survey Data (GSS)
- National Longitudnal Survey of Youth (NLSY)
- Evaluation Strategy
- Caveats

Empirical Applications

- General Social Survey Data (GSS)
  - Sociological survey: adults in randomly selected US households.
- National Longitudnal Survey of Youth (NLSY)
- Evaluation Strategy
- Caveats

Empirical Applications

- General Social Survey Data (GSS)
  - Sociological survey: adults in randomly selected US households.
  - Data on attitudes and demographic characteristics.
- National Longitudnal Survey of Youth (NLSY)
- Evaluation Strategy
- Caveats

Empirical Applications

- General Social Survey Data (GSS)
  - Sociological survey: adults in randomly selected US households.
  - Data on attitudes and demographic characteristics.
- National Longitudnal Survey of Youth (NLSY)
  - Longitudnal dataset: Tracking cohort of young men and women over time.
- Evaluation Strategy
- Caveats

Empirical Applications

- General Social Survey Data (GSS)
    - Sociological survey: adults in randomly selected US households.
    - Data on attitudes and demographic characteristics.
- National Longitudnal Survey of Youth (NLSY)
    - Longitudnal dataset: Tracking cohort of young men and women over time.
    - Data on range of economic, psychological, demographic characteristics.
- Evaluation Strategy
- Caveats

Empirical Applications

- General Social Survey Data (GSS)
  - Sociological survey: adults in randomly selected US households.
  - Data on attitudes and demographic characteristics.
- National Longitudnal Survey of Youth (NLSY)
  - Longitudnal dataset: Tracking cohort of young men and women over time.
  - Data on range of economic, psychological, demographic characteristics.
- Evaluation Strategy
  - Subsets of the data used
- Caveats

Empirical Applications

- General Social Survey Data (GSS)
  - Sociological survey: adults in randomly selected US households.
  - Data on attitudes and demographic characteristics.
- National Longitudnal Survey of Youth (NLSY)
  - Longitudnal dataset: Tracking cohort of young men and women over time.
  - Data on range of economic, psychological, demographic characteristics.
- Evaluation Strategy
  - Subsets of the data used
  - 10% observed data held-out at random.
- Caveats

Empirical Applications

- General Social Survey Data (GSS)
  - Sociological survey: adults in randomly selected US households.
  - Data on attitudes and demographic characteristics.
- National Longitudnal Survey of Youth (NLSY)
  - Longitudnal dataset: Tracking cohort of young men and women over time.
  - Data on range of economic, psychological, demographic characteristics.
- Evaluation Strategy
  - Subsets of the data used
  - 10% observed data held-out at random.
  - Imputation models: Low Rank (Scaled), Low Rank (Unscaled), Trace Norm (Full Rank), Trace Norm (Low Rank), MICE
- Caveats

Empirical Applications

- General Social Survey Data (GSS)
  - Sociological survey: adults in randomly selected US households.
  - Data on attitudes and demographic characteristics.
- National Longitudnal Survey of Youth (NLSY)
  - Longitudnal dataset: Tracking cohort of young men and women over time.
  - Data on range of economic, psychological, demographic characteristics.
- Evaluation Strategy
  - Subsets of the data used
  - 10% observed data held-out at random.
  - Imputation models: Low Rank (Scaled), Low Rank (Unscaled), Trace Norm (Full Rank), Trace Norm (Low Rank), MICE
  - Loss calculated over hold out sample
- Caveats

Key Results: GSS

Key Results: GSS

- Overall Trace (Full Rank) had lowest loss, all Low Rank and Low Norm models outperformed MICE

Key Results: GSS

- Overall Trace (Full Rank) had lowest loss, all Low Rank and Low Norm models outperformed MICE
- Column-wise: $\approx 80\%$ columns had lower loss compared to MICE

Key Results: GSS

- Overall Trace (Full Rank) had lowest loss, all Low Rank and Low Norm models outperformed MICE
- Column-wise: $\approx 80\%$ columns had lower loss compared to MICE
  - Summary Table

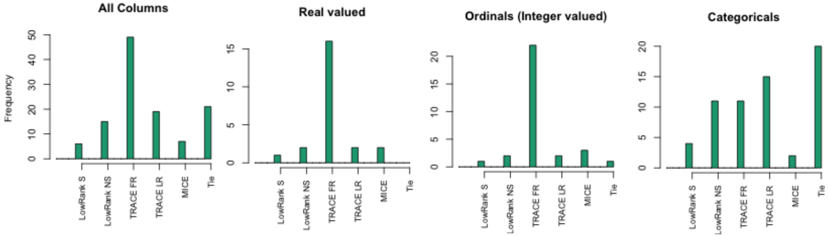
	LowRank (S)	LowRank (NS)	Trace (FR)	Trace (LR)	MICE
Loss/(10^3)	18.50	15.80	14.40	15.80	20.60
%age reduction over MICE	10.10 %	23.40 %	30.10 %	23.00 %	–
%age cols w/ lower loss	73.50 %	84.60 %	87.20 %	84.60 %	–

Key Results: GSS

- Overall Trace (Full Rank) had lowest loss, all Low Rank and Low Norm models outperformed MICE
- Column-wise: $\approx 80\%$ columns had lower loss compared to MICE
  - Summary Table

	LowRank (S)	LowRank (NS)	Trace (FR)	Trace (LR)	MICE
Loss/(10^3)	18.50	15.80	14.40	15.80	20.60
%age reduction over MICE	10.10 %	23.40 %	30.10 %	23.00 %	–
%age cols w/ lower loss	73.50 %	84.60 %	87.20 %	84.60 %	–

  - Method with lowest loss across columns



Key Results: NLSY

Key Results: NLSY

- Overall Trace (Full Rank) had lowest loss, all Low Rank and Low Norm models outperformed MICE

Key Results: NLSY

- Overall Trace (Full Rank) had lowest loss, all Low Rank and Low Norm models outperformed MICE
- Column-wise: $\approx 90\%$ columns had lower loss compared to MICE

Key Results: NLSY

- Overall Trace (Full Rank) had lowest loss, all Low Rank and Low Norm models outperformed MICE
- Column-wise: $\approx 90\%$ columns had lower loss compared to MICE
  - Summary Table

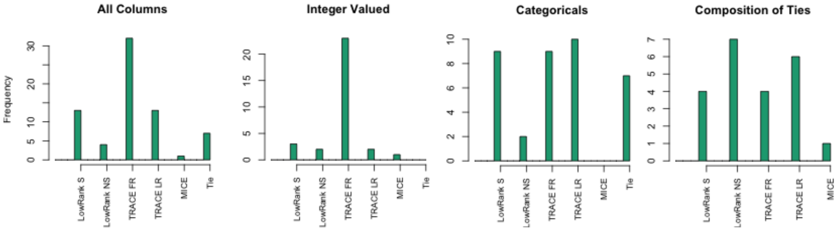
	LowRank (S)	LowRank (NS)	Trace (FR)	Trace (LR)	MICE
Loss/(10^3)	31.40	28.20	25.90	28.20	37.00
%age reduction over MICE	15.20 %	23.70 %	30.00 %	23.70 %	–
%age cols w/ lower loss	75.70%	92.90 %	94.30 %	94.30 %	–

Key Results: NLSY

- Overall Trace (Full Rank) had lowest loss, all Low Rank and Low Norm models outperformed MICE
- Column-wise: $\approx 90\%$ columns had lower loss compared to MICE
  - Summary Table

	LowRank (S)	LowRank (NS)	Trace (FR)	Trace (LR)	MICE
Loss/(10^3)	31.40	28.20	25.90	28.20	37.00
%age reduction over MICE	15.20 %	23.70 %	30.00 %	23.70 %	–
%age cols w/ lower loss	75.70%	92.90 %	94.30 %	94.30 %	–

  - Method with lowest loss across columns



Next Steps

- Probabilistic losses

Next Steps

- Probabilistic losses
- Max Norm regularizer

Next Steps

- Probabilistic losses
- Max Norm regularizer
- Replicating missingness patterns

Next Steps

- Probabilistic losses
- Max Norm regularizer
- Replicating missingness patterns
- Wrapper for Multiple Imputation

Next Steps

- Probabilistic losses
- Max Norm regularizer
- Replicating missingness patterns
- Wrapper for Multiple Imputation
- Extending GLRM to longitudnal data using Tensor Decomposition

Next Steps

- Probabilistic losses
- Max Norm regularizer
- Replicating missingness patterns
- Wrapper for Multiple Imputation
- **Extending GLRM to longitudnal data using Tensor Decomposition** $\Leftarrow$ Future Work.

Thank you!
(Comments and Suggestions Welcome)