Data and Decision-Making: Informative Missingness, Recommender Systems, and Personalised Medicine

Howard Bondell

School of Mathematics and Statistics
University of Melbourne
howard.bondell@unimelb.edu.au

5 November 2025

Outline

1. Informative Missingness in Recommender Systems
   - What is a Recommender System?
   - Common Approaches
   - Modelling Informative Missingness and Prediction
   - MovieLens Data Example
2. Optimal Treatment Decisions in Personalised Medicine
   - What is Personalised Medicine?
   - Constructing Optimal Decision Rules
   - Selection of Variables to Include in Decision
   - Application to AIDS Clinical Trial

Part I:

Informative Missingness in Recommender Systems

Motivation: Netflix Prize

- 2009: Netflix offered USD \$1M in competition to predict user ratings of movies.

Motivation: Netflix Prize

- 2009: Netflix offered USD \$1M in competition to predict user ratings of movies.
- Provided data on ratings with 480K users and 180K movies.
  - On average, each user rated only 200 movies.
- **Goal:** Predict ratings for remaining movies.
- Prize to team with smallest error - must also beat current method by 10%.

What is a Recommender System?

Recommender systems: Methods to take historical preference data and recommend items to users.

- Common Example: For a given user, wish to recommend a movie.
  - Ideally one that they would like!

What is a Recommender System?

Recommender systems: Methods to take historical preference data and recommend items to users.

- Common Example: For a given user, wish to recommend a movie.
  - Ideally one that they would like!
- Mathematical / Statistical / Machine Learning Problem:
  - Given a (large) matrix, rows are users, columns are movies
  - Entries are user ratings - mostly missing (typically $\approx$ 95%)
  - Predict missing values!

What is a Recommender System?

How to fill in missing ratings?

- If filled in, then can recommend movies to users.

Users	Movie1	Movie2	Movie3	Movie4	Movie5	Movie6	...
User1	?	?	4	?	1	?	...
User2	2	5	2	?	?	2	...
User3	?	?	5	3	2	4	...
User4	1	?	?	4	?	?	...
User5	2	3	?	?	?	?	...
...	...	...	...	...	...	...	...

Common Approaches

Methods typically fall into 3 main classes (but lots of crossover):

1. Low Rank Matrix Decompositions
   - Matrix completion (Pure Maths)
2. Collaborative Filtering
   - Predict via similarity (Computer Science)
3. Regression Modelling
   - Model response directly (Statistics)

Low Rank Matrix Decomposition

Matrix Completion Problem

- Seek low rank matrix - i.e. simple structure.
- Match observed values closely subject to rank K restriction.
- Other types of restrictions also proposed.
- Cold start problem - How to handle new user or movie?

Collaborative Filtering

To predict rating for a particular user-item pair

- Find users who have rating for item.
- Compare to current user.
- Predict via using those that are similar.
- Group partitions, nearest neighbours, kernel weighting, etc.

Regression Model - Intuition

- Consider a specific user rating of a specific movie.

  Rating = f (features) + User effect + Movie effect + Error

Regression Model - Intuition

- Consider a specific user rating of a specific movie.

  Rating = f (features) + User effect + Movie effect + Error

  Fixed Effects:

  - f (features): function of user and movie features - often linear.

    Example user features: Age, Sex, etc.
    Example movie features: Genre, Box Office Sales, etc.

Regression Model - Intuition

- Consider a specific user rating of a specific movie.

  Rating = f (features) + User effect + Movie effect + Error

  Fixed Effects:

  - f (features): function of user and movie features - often linear.

    Example user features: Age, Sex, etc.
    Example movie features: Genre, Box Office Sales, etc.

  Random Effects:

  - User effect: User tends to rate higher/lower than average user.
  - Movie effect: Movie tends to score higher/lower than average.
  - Error: Everything else not accounted for.

Problem: Informative Missingness

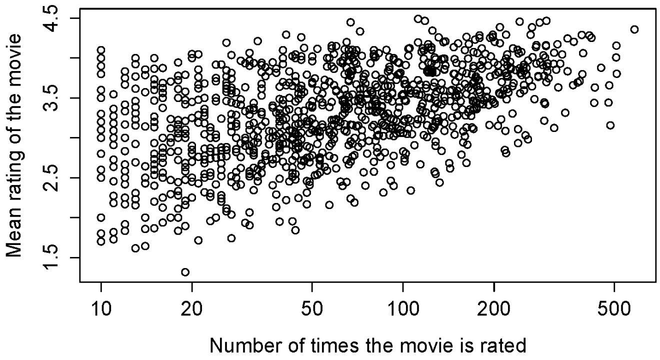
Problem: Data not Missing At Random!

- Users more likely to watch movies they would like.
- Observed ratings may be higher than typical rating.
- Missingness depends on underlying rating.
- Typical approaches do not account for this fact.
    - Biased parameter estimates.
    - More importantly, biased predictions!

Problem: Informative Missingness

Example: MovieLens Data

- 943 users, 1683 movies, approximately 6% rated.



Modelling Informative Missingness

Model using shared variable approach

- Unobserved "Value" determines if movie is watched/rated
- If Value > 0, then not missing, otherwise missing.

Modelling Informative Missingness

Model using shared variable approach

- Unobserved "Value" determines if movie is watched/rated
- If Value > 0, then not missing, otherwise missing.

  Value = f (Rating) + User effect + Movie effect + Error

Modelling Informative Missingness

Model using shared variable approach

- Unobserved "Value" determines if movie is watched/rated
- If Value > 0, then not missing, otherwise missing.

  $$\text{Value} = \text{f}\,(\text{Rating}) + \text{User effect} + \text{Movie effect} + \text{Error}$$

  - f (Rating): function of Rating - possibly linear.
  - Probability of rating being observed is monotone in true underlying rating

Modelling Informative Missingness

Model using shared variable approach

- Unobserved "Value" determines if movie is watched/rated
- If Value > 0, then not missing, otherwise missing.

  Value $=$ f (Rating) $+$ User effect $+$ Movie effect $+$ Error

  - f (Rating): function of Rating - possibly linear.
  - Probability of rating being observed is monotone in true underlying rating
    - User effect: accounts for users that watch (rate) lots/few movies.
    - Movie effect: accounts for often/rarely watched movies.

Joint Model

Joint Model for Rating and Missingness:

1. Model for rating:
    - Rating = f (features) + User effect + Movie effect + Error

Joint Model

Joint Model for Rating and Missingness:

1. Model for rating:
   - Rating $=$ f (features) $+$ User effect $+$ Movie effect $+$ Error
2. Model for missingness:
   - Value $=$ f (Rating) $+$ User effect $+$ Movie effect $+$ Error
   - Missing $=$ No, if Value > 0.

Joint Model

Joint Model for Rating and Missingness:

1. Model for rating:
   - Rating = f (features) + User effect + Movie effect + Error
2. Model for missingness:
   - Value = f (Rating) + User effect + Movie effect + Error
   - Missing = No, if Value $>$ 0.

- Feedback between the two layers.
- Include low rank latent factor model (user x movie) in model. Adds structure of low rank decomposition of error matrix
- Assume Normal Distributions for random effects and errors. Probit model for missingness.

Fitting Model

Given joint model setup:

- Expectation-Maximization (EM) Algorithm useful for computation.
  - Everything conditional normal or truncated normal!
- Alternative: Fully Bayesian Approach
  - Everything conditionally conjugate!

Prediction of Unobserved Ratings

Goal is to predict cell ratings.

- Use conditional expectation with random effects, as usual.
- But, given joint model, conditioning on rating observed gives different prediction than conditioning on it missing.
- Intuition: All else being equal, prediction for rating that is observed will be larger than for one that is missing.

MovieLens Data

Example: MovieLens Data (`http://www.movielens.org/`).

- 943 users and 1,682 movies for total of 1,586,126 cells.
- Approximately 6% observed $\approx$ 100,000.
- Covariates: For user - Age, Sex, Postcode (1st Digit), Occupation
  For movie - Genre

MovieLens Data

Example: MovieLens Data (`http://www.movielens.org/`).

- 943 users and 1,682 movies for total of 1,586,126 cells.
- Approximately 6% observed $\approx$ 100,000.
- Covariates: For user - Age, Sex, Postcode (1st Digit), Occupation
  For movie - Genre
- Split into 80% training and 20% test.
- Fit full joint model vs. single layer regression model only (no informative missingness).

MovieLens Data

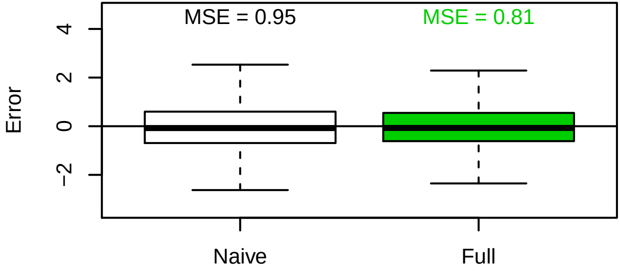
Example: MovieLens Data (`http://www.movielens.org/`).

- 943 users and 1,682 movies for total of 1,586,126 cells.
- Approximately 6% observed $\approx$ 100,000.
- Covariates: For user - Age, Sex, Postcode (1st Digit), Occupation
  For movie - Genre
- Split into 80% training and 20% test.
- Fit full joint model vs. single layer regression model only (no informative missingness).
- For joint model, since hold-out data were actually observed, predictions made conditional on being observed.

MovieLens Data

Example: MovieLens Data

- Boxplots of prediction errors made on the omitted data ($\approx 20,000$).
- MSE for full model, more than 10% improvement!
- 10% improvement on Netflix data was worth \$1 Million!



Informative Missingness in Recommender Systems

Benefits:

- Including missing data mechanism improves prediction accuracy.
- For larger problems, parallel computations using divide-and-conquer.
- Can be used with other recommender systems: Product rating, music rating, etc.

Informative Missingness in Recommender Systems

Benefits:

- Including missing data mechanism improves prediction accuracy.
- For larger problems, parallel computations using divide-and-conquer.
- Can be used with other recommender systems: Product rating, music rating, etc.

Additional Structure:

- Ideally, probability of watching movie monotone in true rating.
- But, conditionally on viewing movie, probability of submitting rating more likely in extremes, i.e. quadratic.
- Data not publicly available to fit model.

Part II:

Optimal Treatment Decisions in Personalised Medicine

Personalised Medicine

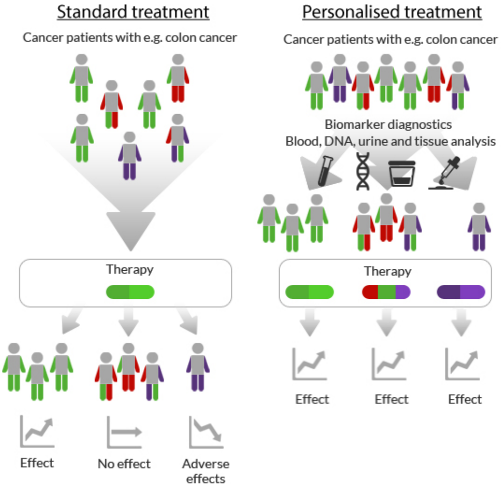


*Image adapted from the DNA Research Center

Personalised Medicine

- Same treatment is not always best for all patients.
- Which treatment should be assigned to which patient?

Personalised Medicine

- Same treatment is not always best for all patients.
- Which treatment should be assigned to which patient?
  - Depends on the patient's characteristics:
    - Age.
    - Sex.
    - Medical history.
    - Genetic information.
    - . . .

Personalised Medicine

- Same treatment is not always best for all patients.
- Which treatment should be assigned to which patient?
  - Depends on the patient's characteristics:
    - Age.
    - Sex.
    - Medical history.
    - Genetic information.
    - ...
  - Collecting and storing all information can complicate decision process and add cost.
- **Goal:** Identify characteristics important for treatment decisions.

Motivating Example

- AIDS Clinical Trials Group (ACTG) 175:

  Randomized, double-blind, placebo-controlled trial to compare efficacy among treatments in HIV-infected patients.

- ACTG175 dataset contains 16 baseline covariates including: age, sex, previous treatment history, etc.

Motivating Example

- AIDS Clinical Trials Group (ACTG) 175:

  Randomized, double-blind, placebo-controlled trial to compare efficacy among treatments in HIV-infected patients.

- ACTG175 dataset contains 16 baseline covariates including: age, sex, previous treatment history, etc.

Mean (standard error) of treatment effect at 96 weeks

Treatment	CD4 Cell Count Ratio to Baseline
Zidovudine (AZT)	0.798 (0.025)
AZT + Didanosine (ddI)	1.009 (0.028)
AZT + Zalcitabine (ddC)	1.000 (0.025)

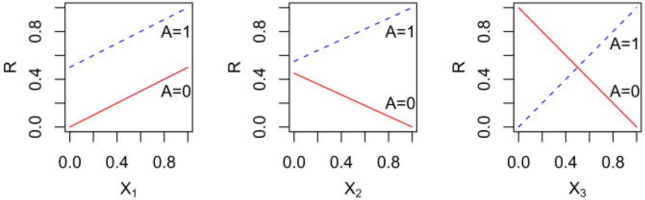
- Overall, combination of AZT + ddI is best (not significantly), but can we identify subgroups of patients who would benefit from the other combination, AZT + ddC?

Variable Selection for Treatment Decisions

- Typical variable selection considers variables that affect the response, i.e. those that help predict higher/lower CD4 count.
- Here, we seek variables to help determine treatment to employ.

Variable Selection for Treatment Decisions

- Typical variable selection considers variables that affect the response, i.e. those that help predict higher/lower CD4 count.
- Here, we seek variables to help determine treatment to employ.



(a) No interaction (b) Non-qualitative interaction (c) Qualitative Interaction

Notation and Formal Objective

- *Optimal treatment* for a patient is that which maximizes expected response.
- Decision rule will be a function of the set of variables $\boldsymbol{X}$.

Notation and Formal Objective

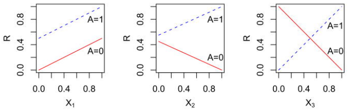
- *Optimal treatment* for a patient is that which maximizes expected response.
- Decision rule will be a function of the set of variables $\boldsymbol{X}$.
- Want decision rule resulting in optimal treatment for each patient.
    - Equivalently: Largest average response over all patients.

Notation and Formal Objective

- *Optimal treatment* for a patient is that which maximizes expected response.
- Decision rule will be a function of the set of variables $\boldsymbol{X}$.
- Want decision rule resulting in optimal treatment for each patient.
    - Equivalently: Largest average response over all patients.
- **Goal:** Find subset of variables, $\boldsymbol{X}_S$, that provides a rule with same average response as optimal one considering all $\boldsymbol{X}$.
    - These are only variables relevant to decision-making.

Simple Example

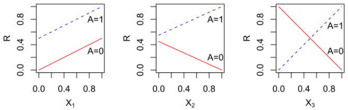
- Decision rule here is simple:
  - If $X_3 > 0.5$ then give Treatment 1. Otherwise Treatment 0.



(a) No interaction (b) Non-qualitative interaction (c) Qualitative Interaction

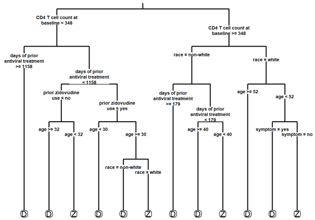
Simple Example

- Decision rule here is simple:
  - If $X_3 > 0.5$ then give Treatment 1. Otherwise Treatment 0.



(a) No interaction (b) Non-qualitative interaction (c) Qualitative Interaction

- In general, it would be more complex.



Common Approaches

1. Variable selection as usual:
   Regularization, Cross-Validation, AIC/BIC.
   - Targets on prediction error, not decision-making subset.

Common Approaches

1. Variable selection as usual:
   Regularization, Cross-Validation, AIC/BIC.
   - Targets on prediction error, not decision-making subset.
2. Include interactions between treatment and each variable.
   - Perform variable selection only on interaction terms.
   - Better, but still targets prediction error.
   - Variable may interact with treatment, but not be qualitative interaction.

Common Approaches

1. Variable selection as usual:
   Regularization, Cross-Validation, AIC/BIC.
   - Targets on prediction error, not decision-making subset.
2. Include interactions between treatment and each variable.
   - Perform variable selection only on interaction terms.
   - Better, but still targets prediction error.
   - Variable may interact with treatment, but not be qualitative interaction.
3. Univariate scores based on magnitudes and locations of crossing in interaction plots.
   - Rank variables based on individual scores.
   - Decide on where to make cutoff.

Proposed Approach: No Regret

- Consider "Regret" from using subset $\boldsymbol{X}_S$ relative to full $\boldsymbol{X}$.
    - Denote as R_S.
- **Idea:** For every patient, compare response obtained using rule based only on $\boldsymbol{X}_S$ to optimal response using full $\boldsymbol{X}$.

Proposed Approach: No Regret

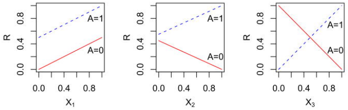
- Consider "Regret" from using subset $\boldsymbol{X}_S$ relative to full $\boldsymbol{X}$.
  - Denote as R_S.
- **Idea:** For every patient, compare response obtained using rule based only on $\boldsymbol{X}_S$ to optimal response using full $\boldsymbol{X}$.
- Formally: Regret is difference in average response obtained using decision rule based on full $\boldsymbol{X}$ compared to that using rule based solely on $\boldsymbol{X}_S$

Proposed Approach: No Regret

- Consider "Regret" from using subset $\boldsymbol{X}_S$ relative to full $\boldsymbol{X}$.
  - Denote as R_S.
- **Idea:** For every patient, compare response obtained using rule based only on $\boldsymbol{X}_S$ to optimal response using full $\boldsymbol{X}$.
- Formally: Regret is difference in average response obtained using decision rule based on full $\boldsymbol{X}$ compared to that using rule based solely on $\boldsymbol{X}_S$
- **Goal:** No Regret!
  - Formal hypothesis test on univariate parameter, $H_0 : R_S = 0$.

Simple Example of Regret

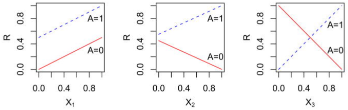
- Here, decision based on full model is
  - If $X_3 > 0.5$ then give Treatment 1. Otherwise Treatment 0.



(a) No interaction (b) Non-qualitative interaction (c) Qualitative Interaction

Simple Example of Regret

- Here, decision based on full model is
  - If $X_3 > 0.5$ then give Treatment 1. Otherwise Treatment 0.

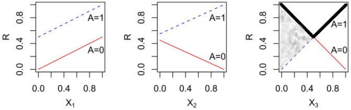


(a) No interaction (b) Non-qualitative interaction (c) Qualitative Interaction

- Consider subset with X_1 and X_2 only.
  - Then decision rule is: ALWAYS give treatment 1.

Simple Example of Regret

- Here, decision based on full model is
  - If $X_3 > 0.5$ then give Treatment 1. Otherwise Treatment 0.

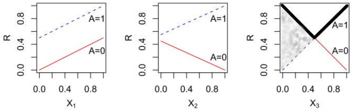


(a) No interaction (b) Non-qualitative interaction (c) Qualitative Interaction

- Consider subset with X_1 and X_2 only.
  - Then decision rule is: ALWAYS give treatment 1.
  - Apply this rule.
  - Regret is area between curves in region of incorrect decision!

Simple Example of Regret

- Here, decision based on full model is
    - If $X_3 > 0.5$ then give Treatment 1. Otherwise Treatment 0.



(a) No interaction (b) Non-qualitative interaction (c) Qualitative Interaction

- Consider subset with X_1 and X_2 only.
    - Then decision rule is: ALWAYS give treatment 1.
    - Apply this rule.
    - Regret is area between curves in region of incorrect decision!

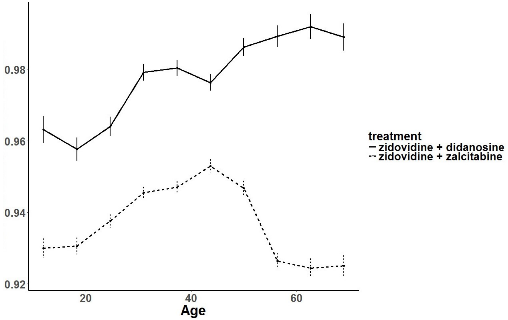
- If we throw out X_1 and X_2 and just keep X_3, NO REGRET.

Procedure

- To conduct hypothesis test, need model for the data.

Procedure

- To conduct hypothesis test, need model for the data.
- Use flexible regression model for response given each treatment condition.
    - Any is suitable, choose your favourite!
      Gaussian Process used here (and marginalised out over all other variables)



Backward Elimination Procedure

- Step 1: Estimate parameters in model.

Backward Elimination Procedure

- Step 1: Estimate parameters in model.
- Step 2: For each patient, now have predicted response under each treatment. So have estimated treatment rule.

Backward Elimination Procedure

- Step 1: Estimate parameters in model.
- Step 2: For each patient, now have predicted response under each treatment. So have estimated treatment rule.
- Step 3: Remove one predictor to get reduced subset, $\boldsymbol{X}_S$, and redo estimation under reduced model.

Backward Elimination Procedure

- Step 1: Estimate parameters in model.
- Step 2: For each patient, now have predicted response under each treatment. So have estimated treatment rule.
- Step 3: Remove one predictor to get reduced subset, $\boldsymbol{X}_S$, and redo estimation under reduced model.
- Step 4: Using the two decision rules, can directly compute plug-in estimate of Regret.

Backward Elimination Procedure

- Step 1: Estimate parameters in model.
- Step 2: For each patient, now have predicted response under each treatment. So have estimated treatment rule.
- Step 3: Remove one predictor to get reduced subset, $\boldsymbol{X}_S$, and redo estimation under reduced model.
- Step 4: Using the two decision rules, can directly compute plug-in estimate of Regret.
- Step 5: Repeat for all predictors, and choose the resulting subset with smallest Regret.

Backward Elimination Procedure

- Step 1: Estimate parameters in model.
- Step 2: For each patient, now have predicted response under each treatment. So have estimated treatment rule.
- Step 3: Remove one predictor to get reduced subset, $\boldsymbol{X}_S$, and redo estimation under reduced model.
- Step 4: Using the two decision rules, can directly compute plug-in estimate of Regret.
- Step 5: Repeat for all predictors, and choose the resulting subset with smallest Regret.
- Step 6: Repeat (3) - (5) until all subsets have *large* Regret.

Further Details on Backward Elimination

- Stop when all subsets have *large* Regret?

Further Details on Backward Elimination

- Stop when all subsets have *large* Regret?
- To test $H_0 : R_S = 0$, note that R_S is a function of the estimated parameters for full model and reduced model.

Further Details on Backward Elimination

- Stop when all subsets have *large* Regret?
- To test $H_0 : R_S = 0$, note that R_S is a function of the estimated parameters for full model and reduced model.
- Using joint distribution of estimated parameters, transform, to obtain distribution of univariate statistic, $\widehat{R_S}$.
- Use corrected p-value, due to multiple tests at each stage.

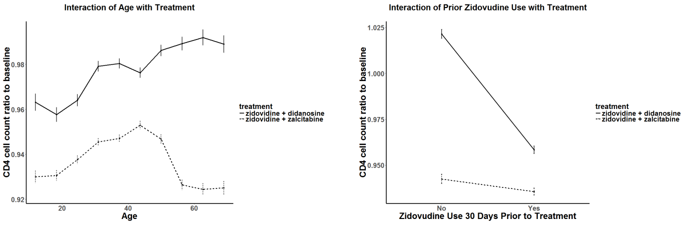
ACTG175 Analysis

- Two treatments for comparison:
  1. AZT + ddI (Best on average across all patients).
  2. AZT + ddC.
- Using 16 candidate predictors in Backward Elimination removes 10 and keeps 6.
  - Next one has corrected p-value around 0.20.

Variables selected as important for decision making, i.e. having qualitative interactions

Covariate	Cumulative Regret	P-value	Corrected P-value
number of days of previous antiretroviral therapy	0.01659	0.0029	0.0174
symptomatic indicator	0.02172	0.0004	0.0020
CD4 T cell count at baseline	0.03062	≤ 0.0001	≤ 0.0001
race	0.03787	≤ 0.0001	≤ 0.0001
indicator of prior AZT use	0.07317	≤ 0.0001	≤ 0.0001
age	0.07393	≤ 0.0001	≤ 0.0001

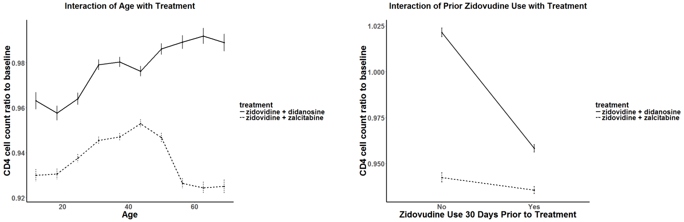
Marginal Covariates-Treatment Interaction



Marginal Covariates-Treatment Interaction Plot

Treatment response surfaces are predicted based on regression for each treatment, and marginal interaction plots are generated via Monte Carlo integration over the other covariates.

Marginal Covariates-Treatment Interaction

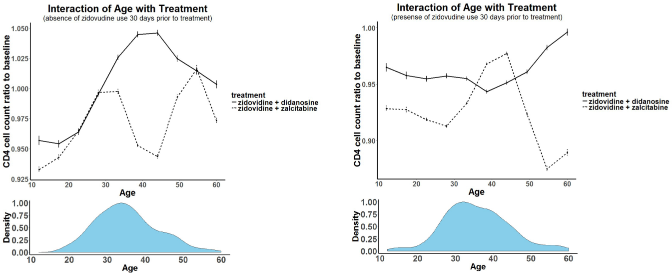


Marginal Covariates-Treatment Interaction Plot

Treatment response surfaces are predicted based on regression for each treatment, and marginal interaction plots are generated via Monte Carlo integration over the other covariates.

- Looks like no importance of *Age* or *AZT Usage* for decision-making?

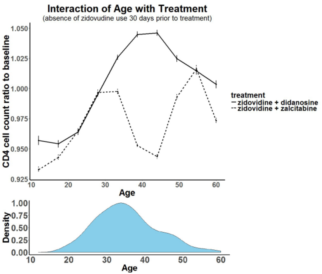
Conditional Covariates-Treatment Interaction



Conditional Covariates-Treatment Interaction Plot

Interaction of *Age* and *Treatment* stratified by *AZT Usage*, along with age distribution in each group.

Conditional Covariates-Treatment Interaction



Conditional Covariates-Treatment Interaction Plot

Interaction of *Age* and *Treatment* stratified by *AZT Usage*, along with age distribution in each group.

- Looks like a subgroup may benefit from alternate treatment!

Treatment Policy Value

- Cross-validation to evaluate the treatment policy.
- Value is estimated by computing mean response on holdout data for those observations consistent with the treatment policy found by each method.
  - Averaged over 100 splits of data (with standard deviation).

Value increase compared to AZT + ddC

Treatment Policy	Percent Patients Treated with AZT+ddI	Value Increase
All treated with AZT+ddI	100%	9.4 (3.1)
Policy from SA	51.1%	9.5 (5.3)
Policy from RV	50.4%	19.1 (4.5)

SA: sequential advantage (Fan et al., 2016). RV: Regret-based variable selection procedure.

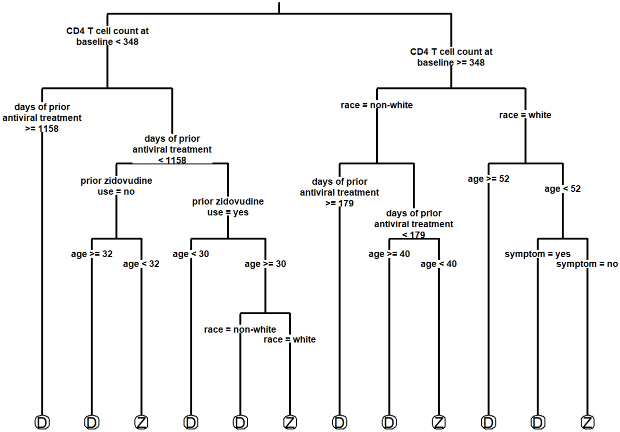
Implementation of Treatment Assignments

- How would medical professional use these results?
- Input patient values, have computer run procedure and spit out predicted best treatment.
  - May not appreciate "Black Box".

Implementation of Treatment Assignments

- How would medical professional use these results?
- Input patient values, have computer run procedure and spit out predicted best treatment.
  - May not appreciate “Black Box”.
- Instead, take results from this estimated policy and provide friendly interface.
  - Now have binary response data, i.e. predicted best treatment for each patient in data.
  - Construct decision tree from this new dataset.
  - Provide this visual representation of treatment rule.

Treatment Policy Tree



Decision tree approximation to the optimal treatment policy.
Z: AZT+ddC. D: AZT+ddI.

Further Issues

- Two (related) questions you may ask about the method:
  1. What happens with a large number of predictors? Backward elimination may not be well-suited.
  2. Why backward elimination, not forward selection?

Further Issues

- Two (related) questions you may ask about the method:
  1. What happens with a large number of predictors? Backward elimination may not be well-suited.
  2. Why backward elimination, not forward selection?
- First is straightforward.
  - Use screening method to reduce number of predictors.
  - Here, screen predictors in each treatment group separately.
  - Variables irrelevant in both can be removed.

Further Issues

- Two (related) questions you may ask about the method:
  1. What happens with a large number of predictors? Backward elimination may not be well-suited.
  2. Why backward elimination, not forward selection?
- First is straightforward.
  - Use screening method to reduce number of predictors.
  - Here, screen predictors in each treatment group separately.
  - Variables irrelevant in both can be removed.
- Second, a bit more complex.

Why Backward Elimination?

- For forward selection, we consider addition of each possible variable, one-at-a-time.
    - Need to fit model for each candidate subset obtained by adding one predictor.
    - Need to do this at every step.
    - Gaussian Process fitting can be expensive, although can be done in parallel.

Why Backward Elimination?

- For forward selection, we consider addition of each possible variable, one-at-a-time.
  - Need to fit model for each candidate subset obtained by adding one predictor.
  - Need to do this at every step.
  - Gaussian Process fitting can be expensive, although can be done in parallel.
- For backward elimination, same issue.
  - **So what's the difference?**

Why Backward Elimination?

- Actually use approximation to backward elimination.
  - Start with full Gaussian Process fit.
  - Using joint normal distribution of estimated parameters, can approximate results for any subset.

Why Backward Elimination?

- Actually use approximation to backward elimination.
  - Start with full Gaussian Process fit.
  - Using joint normal distribution of estimated parameters, can approximate results for any subset.
- Range parameter, ρ_{aj}, is effect of variable X_j on expected value in treatment group a.
  - Removing X_j from model is equivalent to $\rho_{0j} = \rho_{1j} = 0$.

Why Backward Elimination?

- Actually use approximation to backward elimination.
  - Start with full Gaussian Process fit.
  - Using joint normal distribution of estimated parameters, can approximate results for any subset.
- Range parameter, ρ_{aj}, is effect of variable X_j on expected value in treatment group a.
  - Removing X_j from model is equivalent to $\rho_{0j} = \rho_{1j} = 0$.
- Obtain distribution for any subset by conditioning on the appropriate set of $\rho_{aj} = 0$.
  - Once again, multivariate Normal.
- **No need to refit at all!**

Optimal Treatment Decisions in Personalized Medicine

Benefits:

- Selection of quantitative interactions represents variables relevant for decision-making.
- Approach directly targets difference from optimal decision.
- Can apply to multiple treatments, not just binary.

Optimal Treatment Decisions in Personalized Medicine

Benefits:

- Selection of quantitative interactions represents variables relevant for decision-making.
- Approach directly targets difference from optimal decision.
- Can apply to multiple treatments, not just binary.

Additional work:

- Adapt to dynamic treatments, i.e. multiple decision points.
  - Which variables?
  - Which time point?
  - What is the trigger to make a switch?