Lecture 6

Reduced-Rank Regression & Repeated Measures Analysis

Readings: Izenman 2008, Chapter 6; DSA 8020 Lectures 10 & 11 [Link]; DSA 8070 Week 4 MANOVA

DSA 8070 Multivariate Analysis

Whitney Huang
Clemson University

Notes

Agenda

1. Recap of Multivariate Regression
2. Reduced-Rank Regression
3. Repeated Measures Analysis

Notes

Review of Multiple Linear Regression

Setup (univariate response, multiple predictors):

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, \ldots, n,$$

$\varepsilon_i \overset{i.i.d.}{\sim} \mathrm{N}(0, \sigma^2)$

Classical workflow: model selection, least squares estimation, inference on β's, diagnostics

Limitation: When there are *multiple* response variables that may be correlated, separate univariate models can be inefficient and miss joint structure

Notes

What is Multivariate Regression?

Goal: Predict a vector response $\mathbf{Y}$ (dimension q) from predictors $\mathbf{X}$ (dimension p).

Model:

$$\mathbf{Y} = \mathbf{X}\mathbf{B}+\mathbf{E}, \quad \mathbf{Y} \in \mathbb{R}^{n\times q}, \mathbf{X} \in \mathbb{R}^{n\times p}, \mathbf{B} \in \mathbb{R}^{p\times q}, \mathbf{E} \in \mathbb{R}^{n\times q},$$

$$\mathbf{e}_i \overset{i.i.d}{\sim} \mathrm{N}_q(\mathbf{0}, \Sigma)$$

Note:

- Running OLS separately on each column of $\mathbf{Y}$ = doing multiple regressions for each response
- This *ignores* correlation among responses
- $\Rightarrow$ Estimates of coefficients stay the same, but inference changes because of correlated errors

Notes

Why Use Multivariate Regression?

Gains from joint modeling of responses:

- More efficient estimates when responses are correlated
- Ability to draw inferences about *joint* effects
- Dimension reduction: responses may lie in a lower-dimensional subspace $\Rightarrow$ Reduced-Rank Regression

Applications: repeated measurements, growth curves, etc

Notes

Izenman's Reduced-Rank Regression (RRR) – Intuition

Suppose the coefficient matrix has reduced rank: $\mathrm{rank}(\mathbf{B}) = r < \min(p, q)$, p predictors, q responses

- All responses are driven by a small number (r) of latent predictor combinations
- Equivalently, there exist $\mathbf{A} \in \mathbb{R}^{p\times r}$ and $\mathbf{C} \in \mathbb{R}^{r\times q}$ such that
$$\mathbf{B} = \mathbf{A}\mathbf{C}$$
- Interpretation: first map $\mathbf{X}$ to r latent predictors, then map to responses

Notes

Mathematical Formulation of RRR

Model: $\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}$, with $\mathrm{rank}(\mathbf{B}) \leq r$

Rank-constrained least squares:

$$\min_{\mathbf{B}:\ \mathrm{rank}(\mathbf{B}) \leq r} \|\mathbf{Y} - \mathbf{X}\mathbf{B}\|_F^2$$

Key idea:
Start with ordinary multivariate regression, then project the fitted values to a lower-rank approximation via Singular Value Decomposition (SVD)

Notes

Relationship to Canonical-Correlation Analysis (CCA) & PCA

PCA and CCA are coming up soon (Week 8 and 10) – here is a preview

- RRR connects to canonical variates: the latent predictor/response directions from Singular Value Decomposition (SVD) relate to canonical directions that maximize cross-covariance
- When predictors or responses have special structure, RRR also relates to PCA (dimension reduction in $\mathbf{Y}$-space or $\mathbf{X}$-space)
- Emphasis: capture *joint* structure between predictors and responses

Notes

Estimation via Singular Value Decomposition (SVD)

Steps:

1. Fit full multivariate least squares:

$$\widehat{\mathbf{B}}_{\text{OLS}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$$

2. Compute fitted values: $\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\mathbf{B}}_{\text{OLS}}$
3. SVD of fitted values: $\widehat{\mathbf{Y}} = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^\top$
4. Best rank-r approximation (retain top r singular values/vectors):

$$\widehat{\mathbf{Y}}^{(r)} = \mathbf{U}_r\,\mathbf{\Sigma}_r\,\mathbf{V}_r^\top$$

5. Corresponding RRR estimator:

$$\widehat{\mathbf{B}}_{\text{RRR}} = \widehat{\mathbf{B}}_{\text{OLS}}\,\mathbf{V}_r\,\mathbf{V}_r^\top,$$

   i.e., dimension reduction in the response space via $\mathbf{V}_r$

Notes

Choosing the Rank r

How many latent dimensions?

- Cross-validation: choose r with best predictive performance
- Information criteria: AIC, BIC (with appropriate df adjustments)
- Tests / LRTs under normality assumptions (when applicable)

Trade-offs:

- Underfit (r too small): bias
- Overfit (r too large): variance inflation

Notes

Simulation Example

Scenario: $p = 6$ predictors; $q = 4$ responses; true rank $r^\star = 2$

Compare: separate OLS vs. RRR with $r = 2$ (correct), $r = 0, 1$ (underfit), $r = 3, 4$ (overfit).

- When responses are correlated, RRR typically outperforms separate OLS in prediction.
- Evaluate predictive error or estimation error as a function of rank.

Check an example in the R session

Notes

Practical Considerations & Assumptions

- Linearity is assumed; normality needed for classical tests, or rely on large-sample results
- Error covariance and equal-variance (homoscedasticity) assumptions matter for inference
- Multicollinearity in $\mathbf{X}$ or high correlation in $\mathbf{Y}$ can make estimation tricky
- Computation can be challenging when p, q are large compared to n
- Regularized approaches (ridge + rank, nuclear norm) help in high-dimensional settings

Notes

Extensions: Regularized Reduced-Rank Regression

We learned some regularization techniques in **8020 Week 6**

- **Reduced-Rank Ridge Regression:** combines dimension reduction (low-rank) with shrinkage
- **Nuclear-norm penalty:** a convex tool that encourages low-rank structure
- **Why useful?** More stable, handles noise better, and generalizes well when p, q are large but n is only moderate

Notes

Summary & Take-Home Points

- Multivariate regression extends multiple regression to **multiple responses**
- Reduced-Rank Regression (RRR) takes advantage of correlations among responses for efficiency and dimension reduction
- RRR $\equiv$ constrained estimation: limit the rank of $\mathbf{B}$
- Choosing the right rank is key: too small $\rightarrow$ underfit, too large $\rightarrow$ overfit
- Modern extensions: add regularization (ridge, nuclear norm) for better performance, especially in high dimensions
- **Looking ahead:** repeated-measures data can be viewed as multivariate responses over time – methods like RRR help by modeling joint temporal structure and reducing dimensionality

Notes

Repeated Measures as Multivariate Responses

- Repeated measurements can be treated as a multivariate response vector (multivariate regression / MANOVA view)
- Typical covariance structures: compound symmetry, AR(1), etc
- **RRR perspective:** when responses over time are correlated and follow a lower-dimensional trajectory, RRR reduces response dimension while modeling joint temporal behavior

Notes

Dog Experiment [Source: PSU STAT 505]

A completely randomized block design was carried out to determine the effects of 4 surgical treatments on coronary potassium in a group of 36 dogs. There are 9, 8, 9, and 10 dogs in each treatment group, respectively. Each dog was measured at four different time points (1, 5, 9, and 13 minutes) following one of four experimental treatments:

- Control - no surgical treatment is applied
- Extrinsic cardiac denervation immediately prior to treatment
- Bilateral thoracic sympathectomy and stellectomy 3 weeks prior to treatment
- Extrinsic cardiac denervation 3 weeks prior to treatment

We are looking at the treatment effect on the coronary sinus potassium levels

Notes

Notation of Approaches

Let Y_{ijk} be the potassium level for treatment i in dog j at time k:

- there are $a = 4$ treatments (i.e., $i = 1, 2, 3, 4$)
- n_i dogs received treatment i (therefore, there are $n_1 + \cdots + n_a = 9 + 8 + 9 + 10 = 36$ dogs in total)
- $t = 4$, the number of observations over time (i.e., $k = 1, 2, 3, 4$)

Approaches

- Split-plot ANOVA
- MANOVA
- Mixed Models

Notes

Approach 1: Split-plot ANOVA

Model: $Y_{ijk} = \mu + \alpha_i + \delta_{j(i)} + \beta_k + (\alpha\beta)_{ik} + \varepsilon_{ijk}$, where

- α_i: effect of treatment i
- $\delta_{j(i)}$: random effect of dog j receiving treatment i
- β_k: effect of time k
- $(\alpha\beta)_{ik}$: treatment by time interaction
- ε_{ijk}: random error

Assumptions:

- $\varepsilon_{ijk} \overset{i.i.d}{\sim} \mathrm{N}(0, \sigma_\varepsilon^2)$
- $\delta_{j(i)} \overset{i.i.d.}{\sim} \mathrm{N}(0, \sigma_\delta^2)$
- β_k does not depend on the dog $\Rightarrow$ no time by dog interaction

Notes

Split-plot ANOVA Table

Source	df	MS	F
Trt	$a-1$	$\text{MS}_{trt} = \frac{\text{SS}_{trt}}{a-1}$	$F = \frac{\text{MS}_{trt}}{\text{MS}_{error_1}}$
Error 1	$N-a$	$\text{MS}_{error_1} = \frac{\text{SS}_{error_1}}{N-a}$	
Time	$t-1$	$\text{MS}_{time} = \frac{\text{SS}_{time}}{t-1}$	$F = \frac{\text{MS}_{time}}{\text{MS}_{error_2}}$
Trt × Time	$(a-1)(t-1)$	$\text{MS}_{trt \times time} = \frac{\text{SS}_{trt \times time}}{(a-1)(t-1)}$	$F = \frac{\text{MS}_{trt \times time}}{\text{MS}_{error_2}}$
Error 2	$(N-a)(t-1)$	$\text{MS}_{error_2} = \frac{\text{SS}_{error_2}}{(N-a)(t-1)}$	
Total	$Nt-1$		

Notes

Dog Experiment Split-plot Analysis


```
> library(lmerTest)
> fit <- lmer(Response ~ Treatment * Time + (1 | Dog_id), data = dat)
> anova(fit)
Type III Analysis of Variance Table with Satterthwaite's method
               Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
Treatment      3.3396 1.11319     3    32  6.0038  0.002297 **
Time           6.2043 2.06811     3    96 11.1540 2.404e-06 ***
Treatment:Time 3.4397 0.38219     9    96  2.0613  0.040573 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

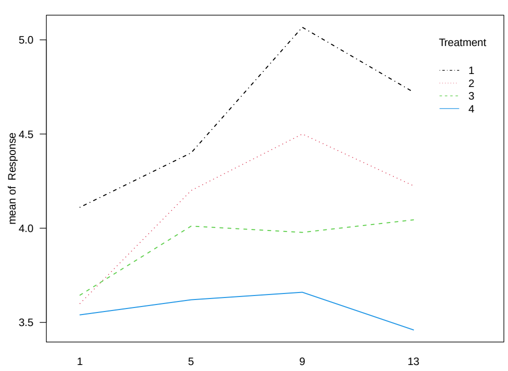
Hypothesis Tests:
We start with the interaction between treatment and time:

$$H_0 : (\alpha\beta)_{ik} = 0 \quad \forall i = 1, \cdots, a,\ k = 1, 2, \cdots, t$$

Result: We conclude the effect of treatment depends on time at $\alpha = 0.05$ level

Notes

Interaction Plot



Rejecting $H_0 : (\alpha\beta)_{ik} = 0$ means we reject the assumption of "parallelism"

Notes

Some Criticisms about the Split-ANOVA Approach

- The Split-plot ANOVA Approach assumes a constant correlation between any two observations from the same dog, that is, $\mathrm{Cor}(Y_{ijk}, Y_{ijk'}) = \frac{\sigma_\delta^2}{\sigma_\delta^2+\sigma_\varepsilon^2}$, this is the so-calle compound symmetry correlation structure
- This assumption is unlikely to be valid with repeated measurements over time as the correlation for two nearby time points is likely to be higher than the correlation for two far apart time points
- Next, we are going to take a multivariate approach (MANOVA) as an attempt to address this issue

Notes

Approach 2: MANOVA

Here we consider the observations over time from the same dog, dog j receiving treatment i as a single vector of interest

$$\boldsymbol{Y}_{ij} = (Y_{ij1}, Y_{ij2}, \cdots, Y_{ijt})^T,$$

and we will perform a one-way MANOVA

Assumptions:

- Dogs receiving treatment i have common mean vector $\boldsymbol{\mu}_i$
- All dogs have common covariance matrix $\boldsymbol{\Sigma}$
- Data from different dogs are independently sampled
- Data are multivariate normally distributed

Notes

Dog Experiment MANOVA Analysis


```
> dat <- read.table("dog1.txt")
> out <- manova(cbind(V3, V4, V5, V6) ~ as.factor(V1), data = dat)
> summary(out, test = "Wilks")
              Df   Wilks approx F num Df den Df  Pr(>F)
as.factor(V1)  3 0.48452    2.022     12 77.018 0.03316 *
Residuals     32
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Results: There are significant differences between at least one pair of treatments in at least one measurement of time

Criticism: MANOVA makes no assumptions regarding the temporal correlation structure, and hence, may be overparameterized leading to poor parameter estimates

Notes

Approach 3: Mixed Model Analysis

Main idea: Split-plot makes a too restrictive assumption while MANOVA makes no assumptions regarding the temporal correlation structure. The mixed model approach allows us to model the temporal correlation involving a limited number of parameters.

Model: $Y_{ijk} = \mu + \alpha_i + \delta_{j(i)} + \beta_k + (\alpha\beta)_{ik} + \varepsilon_{ijk}$.

Assumptions:

- $\varepsilon_{j(ik)} \overset{i.i.d}{\sim} \mathrm{N}(0, \sigma_\varepsilon^2)$
- $\delta_{j(i)} \overset{i.i.d.}{\sim} \mathrm{N}(0, \sigma_\delta^2)$
- The correlation between the errors for the same dog depends only on the difference in observation time points: $|k - k'|$, e.g., $\mathrm{Cor}(Y_{ijk}, Y_{ijk'}) = \rho^{|k-k'|}$ (Autoregressive with order 1)

Notes

Dog Experiment Mixed Model Analysis


```
> library(nlme)
> fit1 = gls(Response ~ Treatment * Time,
+            correlation = corCompSymm(form = ~ 1 | Dog_id), data = dat2)
> fit2 = gls(Response ~ Treatment * Time,
+            correlation = corAR1(form = ~ 1 | Dog_id), data = dat2)
> anova(fit1, fit2)
     Model df      AIC      BIC    logLik
fit1     1 18 275.8063 327.1429 -119.9032
fit2     2 18 277.5811 328.9177 -120.7906
```

Results:

- Based on both AIC/BIC, having an AR(1) does not necessarily improve the model fit (in this data)
- However, having the option of modeling repeated measurement error structure can be useful in general as it provides additional modeling choices

Notes

Linking Repeated Measures to RRR

- Repeated measures can be seen as a **multivariate response vector**, just like in multivariate regression.
- Traditional **Split-plot ANOVA** assumes **compound symmetry** (constant correlation) $\Rightarrow$ often too restrictive.
- **MANOVA** removes correlation assumptions but can be **overparameterized**, especially with many time points.
- **Mixed models** allow more flexible structures (e.g., AR(1)) with fewer parameters.
- **Connection to RRR:** Just as RRR reduces dimensionality by finding a low-rank trajectory for correlated responses, repeated-measures methods aim to balance **parsimony vs. flexibility** in modeling time-dependent correlations.

Notes

Repeated Measures Analysis Summary

We learned about three approaches to analyze repeated measurements:

- Split-plot ANOVA
- MANOVA
- Mixed Effects Model

In the next lecture, we will learn about Inference for Covariance Matrix

Notes

Notes

Notes