

Contents

Preface	xi
A Word to the Practitioner.....	xi
The Organization of the Book.....	xi
Required Software	xii
Accessing the Supplementary Content	xii
Chapter 1 Introducing Partial Least Squares.....	1
Modeling in General.....	1
Partial Least Squares in Today's World.....	2
Transforming, and Centering and Scaling Data.....	3
An Example of a PLS Analysis	4
The Data and the Goal.....	4
The Analysis	5
Testing the Model	9
Chapter 2 A Review of Multiple Linear Regression.....	11
The Cars Example	11
Estimating the Coefficients	15
Underfitting and Overfitting: A Simulation.....	16
The Effect of Correlation among Predictors: A Simulation	18
Chapter 3 Principal Components Analysis: A Brief Visit	25
Principal Components Analysis.....	25
Centering and Scaling: An Example.....	25

The Importance of Exploratory Data Analysis in Multivariate Studies	31
Dimensionality Reduction via PCA.....	34
Chapter 4 A Deeper Understanding of PLS	37
Centering and Scaling in PLS	37
PLS as a Multivariate Technique	38
Why Use PLS?.....	39
How Does PLS Work?.....	45
PLS versus PCA.....	49
PLS Scores and Loadings	50
Some Technical Background	50
An Example Exploring Prediction	59
One-Factor NIPALS Model	60
Two-Factor NIPALS Model	63
Variable Selection	64
SIMPLS Fits	64
Choosing the Number of Factors	65
Cross Validation.....	65
Types of Cross Validation	66
A Simulation of K-Fold Cross Validation	69
Validation in the PLS Platform.....	69
The NIPALS and SIMPLS Algorithms	71
Useful Things to Remember About PLS	72
Chapter 5 Predicting Biological Activity.....	75
Background	75
The Data	76
Data Table Description.....	76
Initial Data Visualization.....	77
A First PLS Model.....	79
Our Plan	79
Performing the Analysis.....	79
The Partial Least Squares Report.....	81
The SIMPLS Fit Report.....	82
Other Options.....	83
A Pruned PLS Model	93

Model Fit	93
Diagnostics.....	95
Performance on Data from Second Study.....	96
Comparing Predicted Values for the Second Study to Actual Values	96
Comparing Residuals for Both Studies	99
Obtaining Additional Insight	101
Conclusion	104
Chapter 6 Predicting the Octane Rating of Gasoline	105
Background	105
The Data	106
Data Table Description.....	106
Creating a Test Set Indicator Column	107
Viewing the Data	108
Octane and the Test Set	108
Creating a Stacked Data Table	109
Constructing Plots of the Individual Spectra.....	111
Individual Spectra	112
Combined Spectra.....	113
A First PLS Model.....	116
Excluding the Test Set	116
Fitting the Model	117
The Initial Report	118
A Second PLS Model	120
Fitting the Model.....	120
High-Level Overview	120
Diagnostics.....	121
Score Scatterplot Matrices.....	125
Loading Plots	127
VIPs	129
Model Assessment Using Test Set.....	133
A Pruned Model.....	136
Chapter 7 Equation Chapter 1 Section 1 Water Quality in the Savannah River Basin	139
Background	140
The Data	141

Data Table Description.....	141
Initial Data Visualization.....	144
Missing Response Values.....	145
Impute Missing Data	146
Distributions	147
Transforming AGPT	148
Differences by Ecoregion	150
Conclusions from Visual Analysis and Implications	155
A First PLS Model for the Savannah River Basin	155
Our Plan	155
Performing the Analysis.....	156
The Partial Least Squares Report.....	159
The NIPALS Fit Report	159
Defining a Pruned Model	163
A Pruned PLS Model for the Savannah River Basin	166
Model Fit.....	166
Diagnostics.....	168
Saving the Prediction Formulas	169
Comparing Actual Values to Predicted Values for the Test Set	170
A First PLS Model for the Blue Ridge Ecoregion	173
Making the Subset.....	173
Reviewing the Data.....	174
Performing the Analysis.....	175
The NIPALS Fit Report	176
A Pruned PLS Model for the Blue Ridge Ecoregion	178
Model Fit.....	178
Comparing Actual Values to Predicted Values for the Test Set	179
Conclusion	181
Chapter 8 Baking Bread That People Like.....	183
Background	183
The Data	184
Data Table Description.....	184
Missing Data Check	186
The First Stage Model.....	187
Visual Exploration of Overall Liking and Consumer Xs	187

The Plan for the First Stage Model	189
Stage One PLS Model	190
Stage One Pruned PLS Model	195
Stage One MLR Model	197
Comparing the Stage One Models.....	200
Visual Exploration of Ys and Xs	202
Stage Two PLS Model	207
Stage Two MLR Model.....	212
The Combined Model for Overall Liking	215
Constructing the Prediction Formula	215
Viewing the Profiler	218
Conclusion	219
Appendix 1: Technical Details.....	221
Ground Rules.....	222
The Singular Value Decomposition of a Matrix.....	222
Definition.....	222
Relationship to Spectral Decomposition	223
Other Useful Facts.....	223
Principal Components Regression.....	223
The Idea behind PLS Algorithms	224
NIPALS	225
The NIPALS Algorithm.....	225
Computational Results.....	228
Properties of the NIPALS Algorithm	231
SIMPLS	237
Optimization Criterion	237
Implications for the Algorithm.....	237
The SIMPLS Algorithm	238
More on VIPs.....	244
The Standardize X Option	246
Determining the Number of Factors.....	246
Cross Validation: How JMP Does It.....	246
Appendix 2: Simulation Studies	249
Introduction	249
The Bias-Variance Tradeoff in PLS	250

Introduction	250
Two Simple Examples	250
Motivation.....	254
The Simulation Study	255
Results and Discussion.....	257
Conclusion.....	261
Using PLS for Variable Selection	263
Introduction	263
Structure of the Study	264
The Simulation	267
Computation of Result Measures	268
Results	270
Conclusion.....	280
References	281
Index	285