

Inhaltsverzeichnis

Vorwort *XI*

1 Einführung in die multivariate Datenanalyse 1

- 1.1 Was ist multivariate Datenanalyse? 1
- 1.2 Datensätze in der multivariaten Datenanalyse 4
- 1.3 Ziele der multivariaten Datenanalyse 5
 - 1.3.1 Einordnen, Klassifizierung der Daten 5
 - 1.3.2 Multivariate Regressionsverfahren 6
 - 1.3.3 Möglichkeiten der multivariaten Verfahren 7
- 1.4 Prüfen auf Normalverteilung 8
 - 1.4.1 Wahrscheinlichkeitsplots 10
 - 1.4.2 Box-Plots 12
- 1.5 Finden von Zusammenhängen 16
 - 1.5.1 Korrelationsanalyse 16
 - 1.5.2 Bivariate Datendarstellung – Streudiagramme 18
- Literatur 20*

2 Hauptkomponentenanalyse 21

- 2.1 Geschichte der Hauptkomponentenanalyse 21
- 2.2 Bestimmen der Hauptkomponenten 22
 - 2.2.1 Prinzip der Hauptkomponentenanalyse 22
 - 2.2.2 Was macht die Hauptkomponentenanalyse? 24
 - 2.2.3 Grafische Erklärung der Hauptkomponenten 25
 - 2.2.4 Bedeutung der Faktorenwerte und Faktorenladungen (Scores und Loadings) 29
 - 2.2.5 Erklärte Varianz pro Hauptkomponente 35
- 2.3 Mathematisches Modell der Hauptkomponentenanalyse 36
 - 2.3.1 Mittenzentrierung 37
 - 2.3.2 PCA-Gleichung 38
 - 2.3.3 Eigenwert- und Eigenvektorenberechnung 38

2.3.4	Berechnung der Hauptkomponenten mit dem NIPALS-Algorithmus	40
2.3.5	Rechnen mit Scores und Loadings	42
2.4	PCA für drei Dimensionen	46
2.4.1	Bedeutung von Bi-Plots	48
2.4.2	Grafische Darstellung der Variablenkorrelationen zu den Hauptkomponenten (Korrelation-Loadings-Plots)	52
2.5	PCA für viele Dimensionen: Gaschromatographische Daten	56
2.6	Standardisierung der Messdaten	65
2.7	PCA für viele Dimensionen: Spektren	72
2.7.1	Auswertung des VIS-Bereichs (500–800 nm)	74
2.7.2	Auswertung des NIR-Bereichs (1100–2100 nm)	81
2.8	Wegweiser zur PCA bei der explorativen Datenanalyse	86
	<i>Literatur</i>	88
3	Multivariate Regressionsmethoden	89
3.1	Klassische und inverse Kalibration	90
3.2	Univariate lineare Regression	92
3.3	Maßzahlen zur Überprüfung des Kalibriermodells (Fehlergrößen bei der Kalibrierung)	93
3.3.1	Standardfehler der Kalibration	93
3.3.2	Mittlerer Fehler – RMSE	94
3.3.3	Standardabweichung der Residuen – SE	95
3.3.4	Korrelation und Bestimmtheitsmaß	96
3.4	Signifikanz und Interpretation der Regressionskoeffizienten	97
3.5	Grafische Überprüfung des Kalibriermodells	97
3.6	Multiple lineare Regression (MLR)	99
3.7	Beispiel für MLR – Auswertung eines Versuchsplans	100
3.8	Hauptkomponentenregression (Principal Component Regression – PCR)	103
3.8.1	Beispiel zur PCR – Kalibrierung mit NIR-Spektren	105
3.8.2	Bestimmen des optimalen PCR-Modells	106
3.8.3	Validierung mit unabhängigem Testset	110
3.9	Partial Least Square Regression (PLS-Regression)	111
3.9.1	Geschichte der PLS	112
3.10	PLS-Regression für eine Y-Variable (PLS1)	113
3.10.1	Berechnung der PLS1-Komponenten	114
3.10.2	Interpretation der P-Loadings und W-Loadings bei der PLS-Regression	117
3.10.3	Beispiel zur PLS1 – Kalibrierung von NIR-Spektren	117
3.10.4	Finden des optimalen PLS-Modells	118
3.10.5	Validierung des PLS-Modells mit unabhängigem Testset	121
3.10.6	Variablenselektion – Finden der optimalen X-Variablen	122
3.11	PLS-Regression für mehrere Y-Variablen (PLS2)	127

- 3.11.1 Berechnung der PLS2-Komponenten 127
- 3.11.2 Wahl des Modells: PLS1 oder PLS2? 129
- 3.11.3 Beispiel PLS2: Bestimmung von Gaskonzentrationen in der Verfahrenstechnik 130
- 3.11.4 Beispiel 2 zur PLS2: Berechnung der Konzentrationen von Einzelkomponenten aus Mischungsspektren 141
Literatur 151

- 4 Kalibrieren, Validieren, Vorhersagen 153**
 - 4.1 Zusammenfassung der Kalibrierschritte – Kalibrierfehler 154
 - 4.2 Möglichkeiten der Validierung 155
 - 4.2.1 Kreuzvalidierung (Cross Validation) 156
 - 4.2.2 Fehlerabschätzung aufgrund des Einflusses der Datenpunkte (Leverage Korrektur) 157
 - 4.2.3 Externe Validierung mit separatem Testset 159
 - 4.3 Bestimmen des Kalibrier- und Validierdatensets 162
 - 4.3.1 Kalibrierdatenset repräsentativ für Y-Datenraum 164
 - 4.3.2 Kalibrierdatenset repräsentativ für X-Datenraum 164
 - 4.3.3 Vergleich der Kalibriermodelle 165
 - 4.4 Ausreißer 168
 - 4.4.1 Finden von Ausreißern in den X-Kalibrierdaten 169
 - 4.4.2 Grafische Darstellung der Einflüsse auf die Kalibrierung 172
 - 4.4.2.1 Einfluss-Grafik: Influence Plot mit Leverage und Restvarianz 172
 - 4.4.2.2 Residuenplots 174
 - 4.5 Vorhersagebereich der vorhergesagten Y-Daten 175
 - 4.5.1 Grafische Darstellung des Vorhersageintervalls 177
Literatur 181

- 5 Datenvorverarbeitung bei Spektren 183**
 - 5.1 Spektroskopische Transformationen 183
 - 5.2 Spektrennormierung 185
 - 5.2.1 Normierung auf den Mittelwert 186
 - 5.2.2 Vektornormierung auf die Länge eins (Betrag-1-Norm) 186
 - 5.3 Glättung 187
 - 5.3.1 Glättung mit gleitendem Mittelwert 187
 - 5.3.2 Polynomglättung (Savitzky-Golay-Glättung) 187
 - 5.4 Basislinienkorrektur 190
 - 5.5 Ableitungen 193
 - 5.5.1 Ableitung nach der Differenzenquotienten-Methode (Punkt-Punkt-Ableitung) 193
 - 5.5.2 Ableitung über Polynomfit (Savitzky-Golay-Ableitung) 195
 - 5.6 Korrektur von Streueffekten 198
 - 5.6.1 MSC (Multiplicative Signal Correction) 198
 - 5.6.2 EMSC (Extended Multiplicative Signal Correction) 199

- 5.6.3 Standardisierung der Spektren (Standard Normal Variate (SNV) Transformation) 202
- 5.7 Vergleich der Vorbehandlungsmethoden 203
Literatur 210

- 6 Eine Anwendung in der Produktionsüberwachung – von den Vorversuchen zum Einsatz des Modells 211**

 - 6.1 Vorversuche 211
 - 6.2 Erstes Kalibriermodell 217
 - 6.3 Einsatz des Kalibriermodells – Validierphase 220
 - 6.4 Offset in den Vorhersagewerten der zweiten Testphase 224
 - 6.5 Zusammenfassung der Schritte bei der Erstellung eines Online-Vorhersagemodells 227

- 7 Tutorial zum Umgang mit dem Programm „The Unscrambler“ der Demo-CD 229**

 - 7.1 Durchführung einer Hauptkomponentenanalyse (PCA) 229
 - 7.1.1 Beschreibung der Daten 229
 - 7.1.2 Aufgabenstellung 230
 - 7.1.3 Datendatei einlesen 230
 - 7.1.4 Definieren von Variablen- und Objektbereichen 231
 - 7.1.5 Speichern der Datentabelle 232
 - 7.1.6 Plot der Rohdaten 233
 - 7.1.7 Verwendung von qualitativen Variablen (kategoriale Variable) 235
 - 7.1.8 Berechnen eines PCA-Modells 238
 - 7.1.9 Interpretation der PCA-Ergebnisse 241
 - 7.1.9.1 Erklärte Varianz (Explained Variance) 241
 - 7.1.9.2 Scoreplot 242
 - 7.1.9.3 Loadingsplot 247
 - 7.1.9.4 Einfluss-Plot (Influence Plot) 250
 - 7.2 Datenvorverarbeitung 253
 - 7.2.1 Berechnung der zweiten Ableitung 253
 - 7.2.2 Glättung der Spektren 256
 - 7.2.3 Berechnen der Streukorrektur mit EMSC 257
 - 7.3 Durchführung einer PLS-Regression mit einer Y-Variablen 261
 - 7.3.1 Aufgabenstellung 261
 - 7.3.2 Interpretation der PLS-Ergebnisse 266
 - 7.3.2.1 PLS-Scoreplot 266
 - 7.3.2.2 Darstellung der Validierungsrestvarianzen (Residual Validation Variance) 269
 - 7.3.2.3 Darstellung der Regressionskoeffizienten 270
 - 7.3.2.4 Darstellung der vorhergesagten und der gemessenen Theophyllinkonzentrationen (Predicted versus Measured Plot) 271
 - 7.3.2.5 Residuenplot 273

- 7.4 Verwenden des Regressionsmodells –
Vorhersage des Theophyllingehalts für Testdaten 276
- 7.5 Export der Unscrambler-Modelle zur Verwendung in beliebigen
Anwendungen 278
- 7.5.1 Kalibriermodell für Feuchte erstellen 279
- 7.5.2 Export des PLS-Regressionsmodells für die Feuchte 283
- 7.5.2.1 Umwandeln der Grafikanzeige in numerische Daten 283
- 7.5.2.2 Export des Regressionsmodells als Text-Datei (ASCII Model) 285
- 7.5.2.3 Berechnung der Feuchte in Excel 286
- 7.6 Checkliste für spektroskopische Kalibrierungen mit dem
Unscrambler 287
Literatur 290

Anhänge A–D 291

- Anhang A 292
- Anhang B 302
- Anhang C 304
- Anhang D 310

Stichwortverzeichnis 313