

BaselineCorrection

Christoph Steinbeck, et al.

16/05/2018

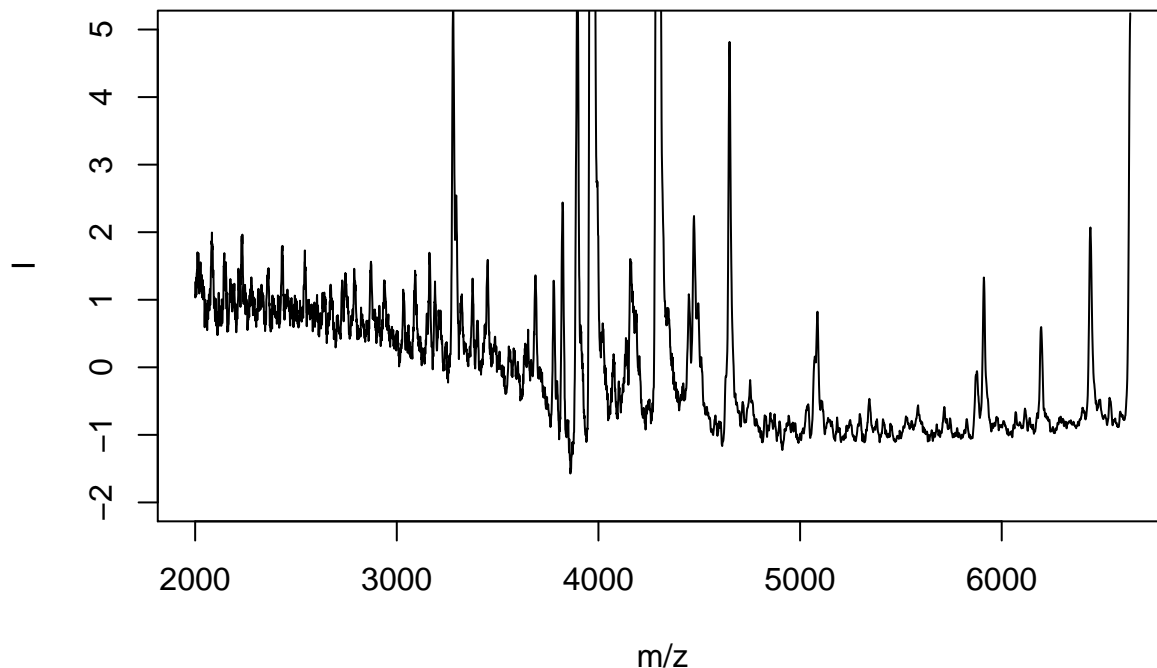
Präambel

Dies ist ein begleitendes Übungsblatt zur Vorlesung Analytische Chemie I (MC 2.1.1). Sie finden hier überwiegend den R-Code für die Beispiele, die in der Vorlesung besprochen wurden. Zu jedem Code-Beispiel, das Sie in der Umgebung R-Studio ausführen können, finden Sie in der Regel eine oder mehrere Aufgaben. Bitte konzentrieren Sie sich nicht nur auf die Aufgaben, sondern studieren Sie die Code-Beispiele und versuchen Sie sie im Detail zu verstehen. Merken Sie sich wiederverwendbare Muster in den Lösungen und nutzen Sie die Zeit mit diesen Beispielen zum Experimentieren und Spielen.

Baseline Correction

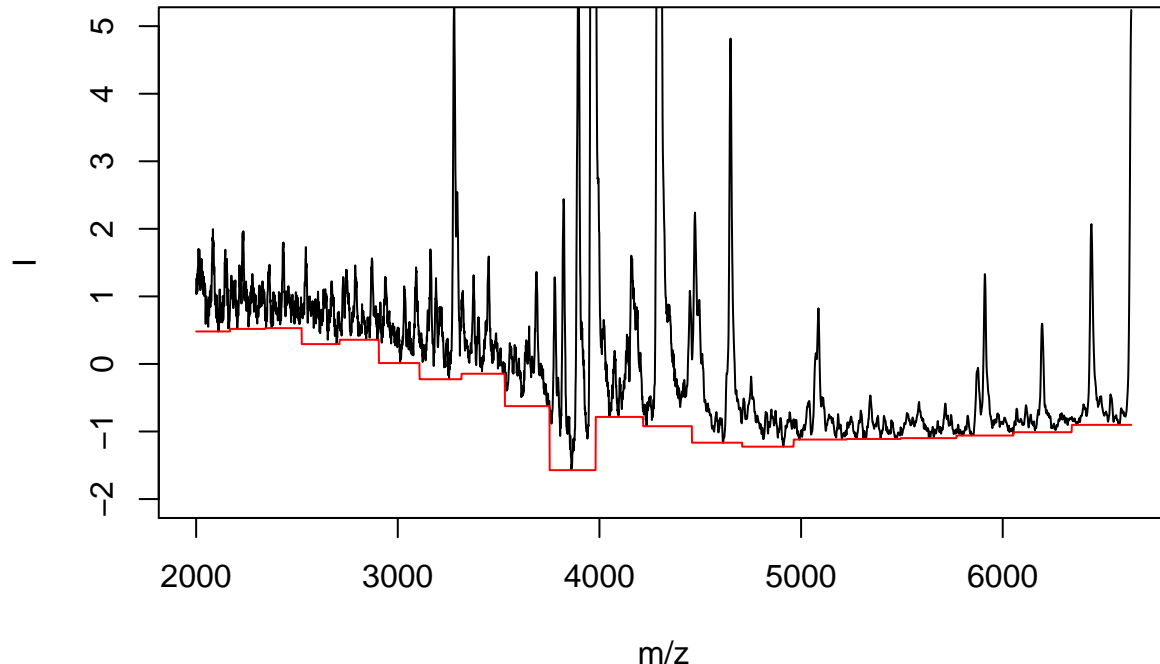
Um eine Basislinienkorrektur zu üben, erzeugen wir zunächst eine künstlich gestörte Basislinie durch Addition eines Cosinus-Ausschnittes zu unseren bereits bekannten Prostata-Daten. Beachten Sie bitte, das working directory im kommenden Code-Abschnitt an Ihre lokalen Gegebenheiten anzupassen.

```
setwd("/Users/steinbeck/Downloads/")
load("Prostate2000Raw.rda")
y <- Prostate2000Raw$intensity[1:4000, 1] + cos(1:4000/1000)
mz <- Prostate2000Raw$mz[1:4000]
plot(mz, y, type = "l", col = "black", ylim = c(-2, 5), xlab = "m/z", ylab = "I")
```



Basislinienkorrektur durch Binning

```
lsection <- 200
xmat <- matrix(y, nrow=lsection)
ymin <- apply(xmat, 2, min)
plot(mz, y, type = "l", col = "black", ylim = c(-2, 5), xlab = "m/z", ylab = "I")
lines(mz, rep(ymin, each = lsection), col="red")
```



AUFGABE: Untersuchen Sie die Funktion der Matrix `xmat` und was the folgende `apply()`-Funktion mit dieser Matrix macht.

AUFGABE: Ziehen Sie (durch Modifikation des obigen Codes) die berechnete Baseline vom Spektrum ab und plotten Sie das Ergebnis. Geniessen Sie die Vorteil der funktionalen Programmierung.

Basislinienkorrektur durch Lokale Polynome

LOESS und LOWESS (local weighted scatterplot smoothing) sind zwei stark verwandte nicht-parametrische Regressionsmethoden, die mehrere Regressionsmodelle in einem k-nearest-neighbor-basierten Metamodell kombinieren. “LOESS” ist eine spätere Verallgemeinerung von LOWESS

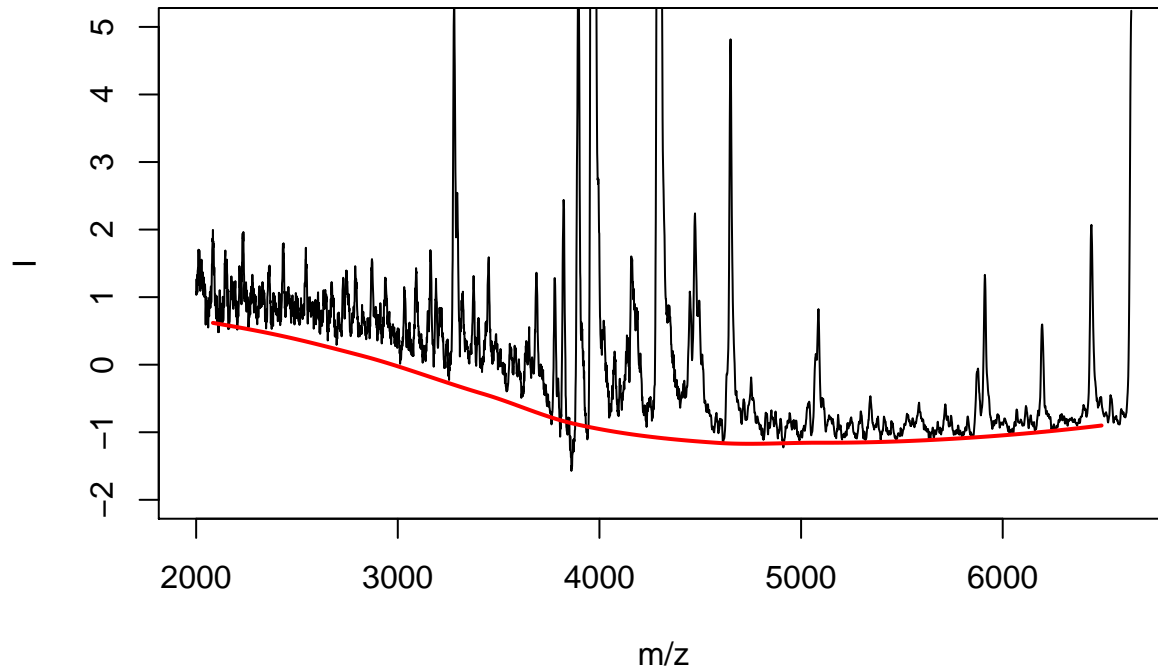
LOESS und LOWESS bauen auf “klassischen” Methoden wie der linearen und nichtlinearen kleinsten Quadrate auf. Sie sprechen Situationen an, in denen die klassischen Verfahren nicht gut funktionieren oder ohne übermäßige Arbeit nicht effektiv angewendet werden können. LOESS verbindet viel von der Einfachheit der linearen kleinsten Quadrate mit der Flexibilität der nichtlinearen Regression. Dies geschieht, indem einfache Modelle an lokalisierte Teilmengen der Daten angepasst werden, um eine Funktion aufzubauen, die den deterministischen Teil der Variation der Daten Punkt für Punkt beschreibt. Eine der Hauptvorteile dieser Methode besteht darin, dass der Datenanalytiker keine globale Funktion in irgendeiner Form angeben muss, um ein Modell an die Daten anzupassen, sondern nur Segmente der Daten.

```
setwd("/Users/steinbeck/Downloads/")
load("Prostate2000Raw.rda")
y <- Prostate2000Raw$intensity[1:4000, 1] + cos(1:4000/1000)
mz <- Prostate2000Raw$mz[1:4000]
```

```

lsection <- 200
xmat <- matrix(y, nrow=lsection)
ymin <- apply(xmat, 2, min)
bsln.loess <- loess(ymin ~ mz[seq(101, 4000, by = 200)])
plot(mz, y, type = "l", col = "black", ylim = c(-2, 5), xlab = "m/z", ylab = "I")
lines(mz, predict(bsln.loess, mz), lwd = 2, col="Red")

```



AUFGABE: Ziehen Sie wieder die berechnete Baseline vom Spektrum ab und begutachten Sie das Ergebnis.

Der Kompromiß für diese Eigenschaften ist eine erhöhte Berechnung. Weil es so rechenintensiv ist, wäre LOESS in der Zeit, in der die Regression der kleinsten Quadrate entwickelt wurde, praktisch unmöglich gewesen. Die meisten anderen modernen Methoden zur Prozessmodellierung ähneln in dieser Hinsicht LOESS. Diese Methoden wurden bewusst so konzipiert, dass sie unsere derzeitigen Computerfähigkeiten optimal nutzen, um Ziele zu erreichen, die mit herkömmlichen Ansätzen nicht leicht zu erreichen sind. Mehr unter https://en.wikipedia.org/wiki/Local_regression.

Basislinienkorrektur durch Asymmetric Least Squares

Asymmetric Least Squares Fitting ist ein alternativer und recht allgemeiner Ansatz für die Basislinienschätzung. Es erzeugt eine langsam variierende Schätzung der Basislinie. Im Gegensatz zur normalen Glättung der kleinsten Quadrate werden positive Abweichungen gegenüber der Basissschätzung jedoch (deutlich) geringer gewichtet als negative.

Die asymmetrische Glättung der kleinsten Quadrate ist für die Grundlinienschätzung attraktiv: 1) Sie ist schnell, auch bei großen Signalen; 2) die Flexibilität der Grundlinie lässt sich mit einem einzigen Parameter justieren 3) die Position der Grundlinie kann mit einem weiteren Parameter eingestellt werden. Gegeben der beiden Parameter sind die Berechnungen vollständig reproduzierbar.

```

library(ptw)
plot(mz, y, col = "gray", type = "l", ylim = c(-2, 5), xlab = "m/z", ylab = "I")
lines(mz, asysm(y), lwd = 2, col="red")

```

