Blatt4

Aufgabe1

a)

Vergleiche handschriftliche Abgabe.

b)

Die beiden Populationen P_0 (blau) und P_1 (rot) sind im folgenden 2D-Scatterplot dargestellt.

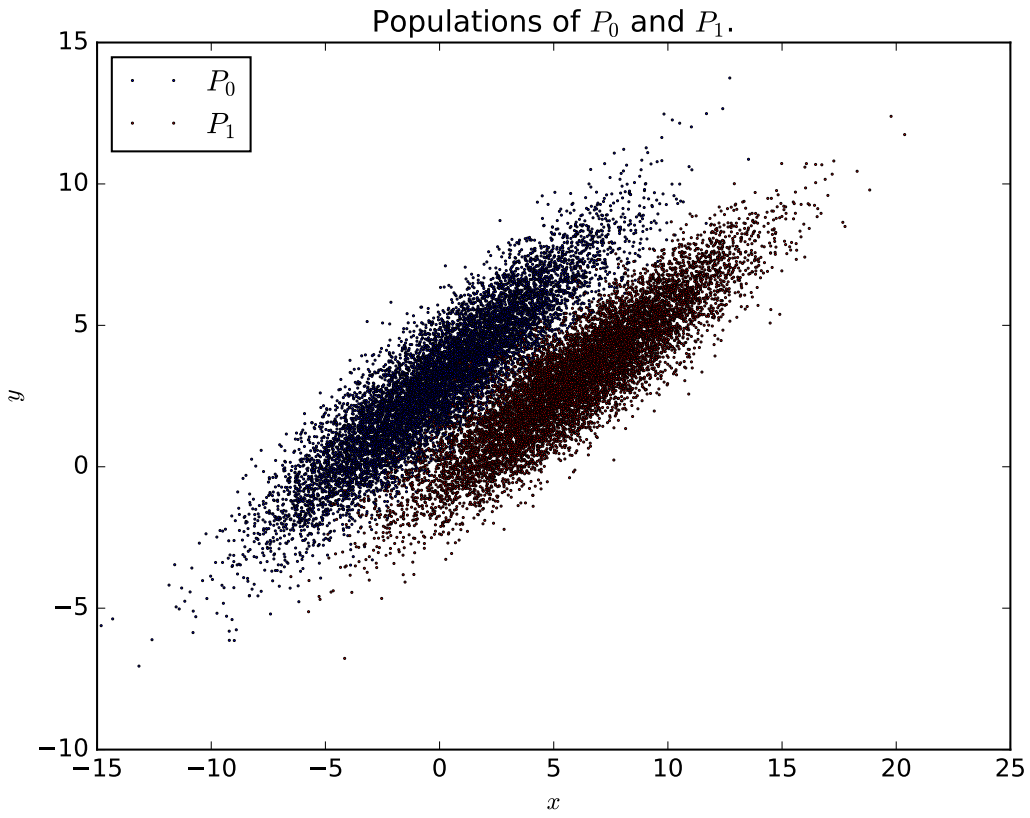

Abbildung 1: Populationen P_0 und P_1

c)

Die Kenngrößen der Population P_0 lauten:

P_0:

$$
\begin{aligned}
x_{\text{mean}} &= 0.0400 \\
y_{\text{mean}} &= 3.0145 \\
\text{Var}[x] &= 12.0826 \\
\text{Var}[y] &= 6.6144 \\
cov(x, y) &= 8.0453 \\
\rho &= 0.8999
\end{aligned}
$$

P_1:

$$
\begin{aligned}
x_{\text{mean}} &= 6.0247 \\
y_{\text{mean}} &= 3.1344 \\
\text{Var}[x] &= 11.9228 \\
\text{Var}[y] &= 5.2991 \\
cov(x, y) &= 7.1731 \\
\rho &= 0.9024
\end{aligned}
$$

P_{ges}:

$$
\begin{aligned}
x_{\text{mean}} &= 3.0323 \\
y_{\text{mean}} &= 3.0744 \\
\text{Var}[x] &= 20.9568 \\
\text{Var}[y] &= 5.9600 \\
cov(x, y) &= 7.7882 \\
\rho &= 0.6969
\end{aligned}
$$

d)

Siehe root-file.

Aufgabe2

a)

Der Plot mit den Projektionsgeraden.

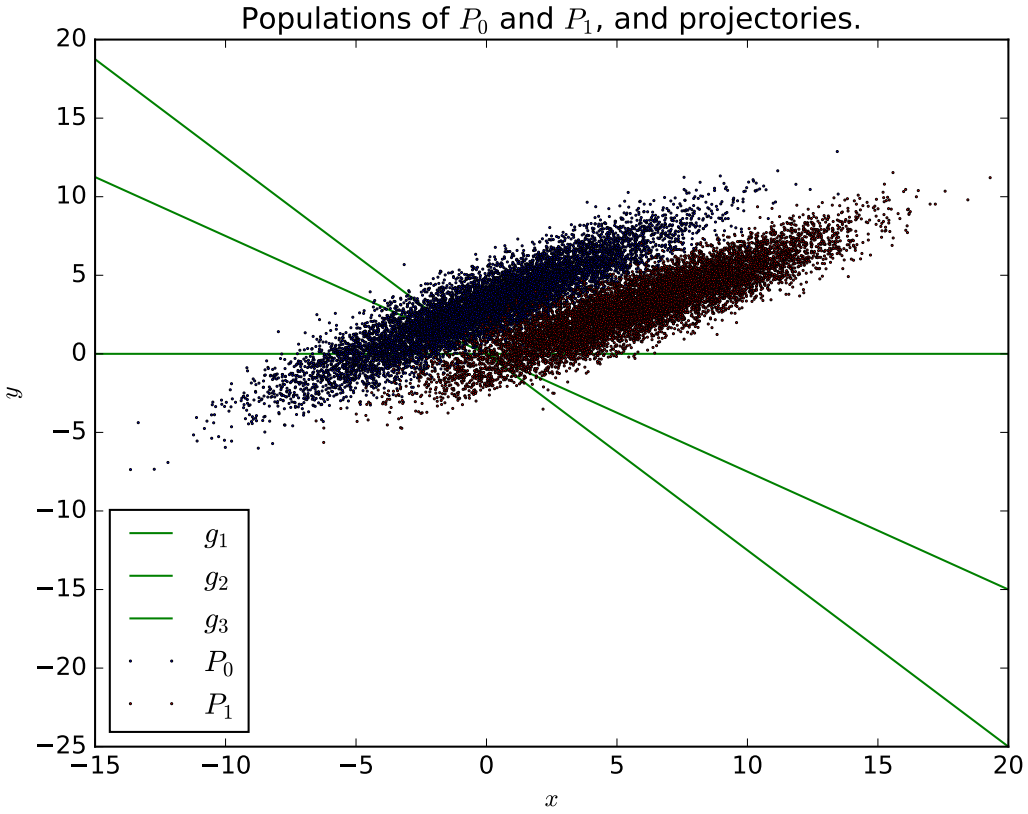

Abbildung 2: Scatterplot der Populationen mit Projektionsgeraden

b)

Alle Histogramme gehören zur b).

c)

Darstellungen der Effizienz und Reinheit.

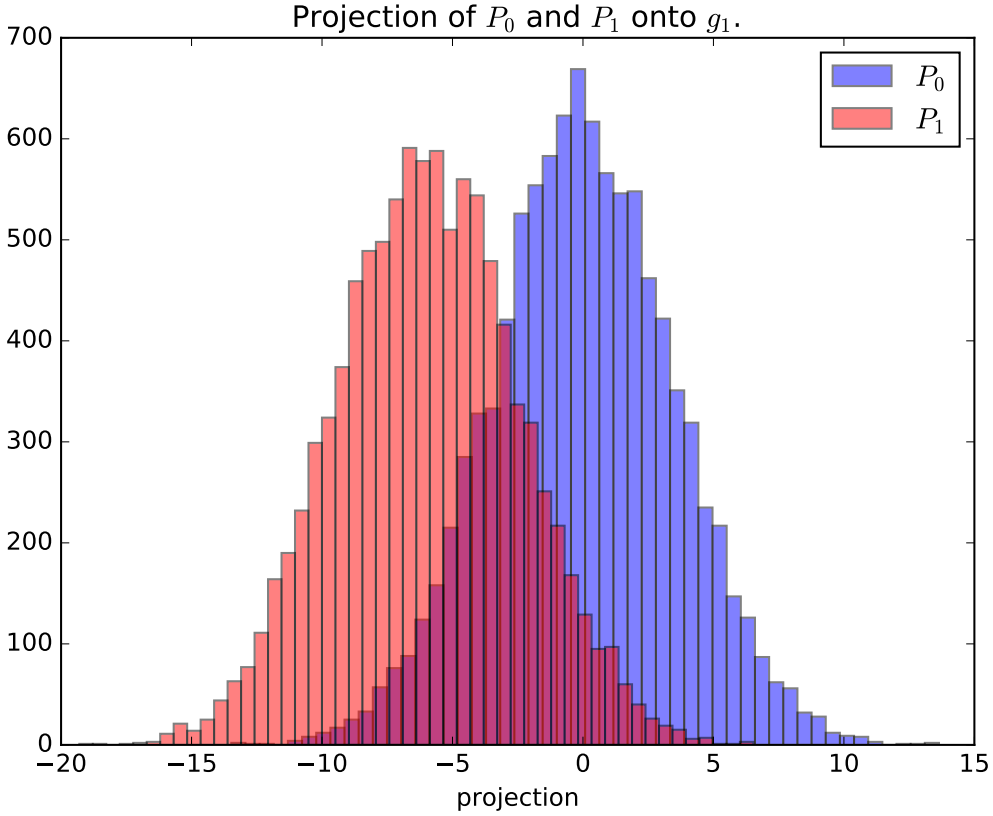

Abbildung 3: Populationen auf Projektionsgerade g_1

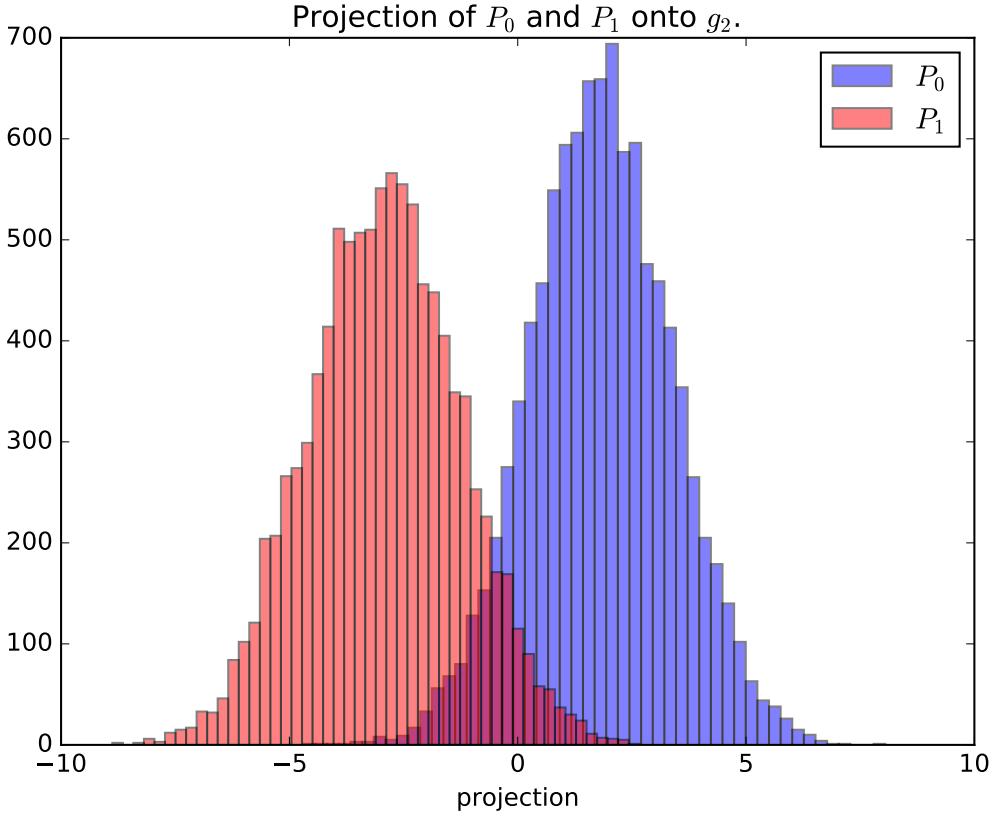

Abbildung 4: Populationen auf Projektionsgerade g_2

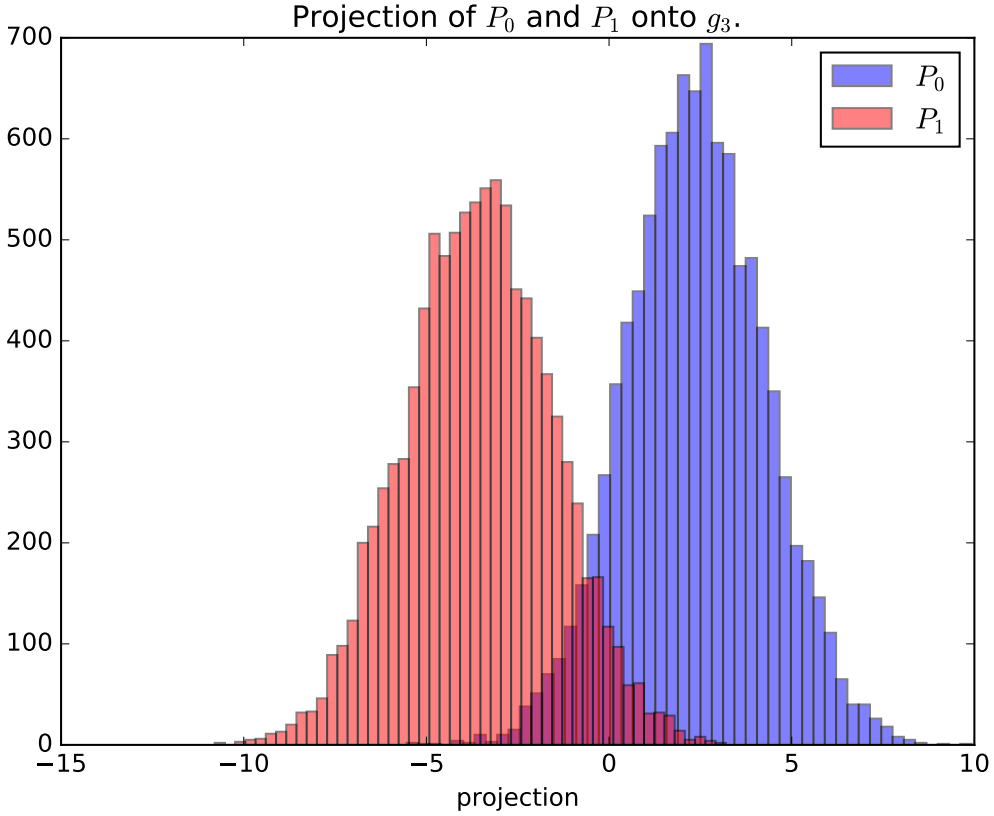

Abbildung 5: Populationen auf Projektionsgerade g_3

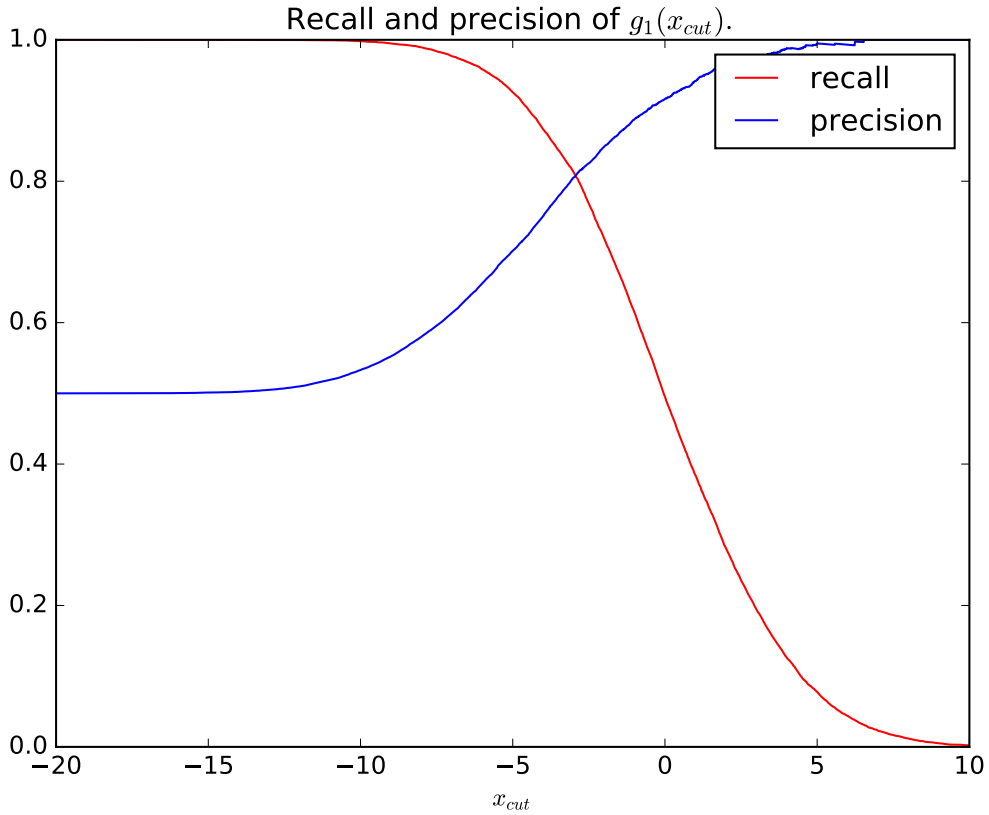

Abbildung 6: Reinheit und Effizienz im Bezug auf g_1

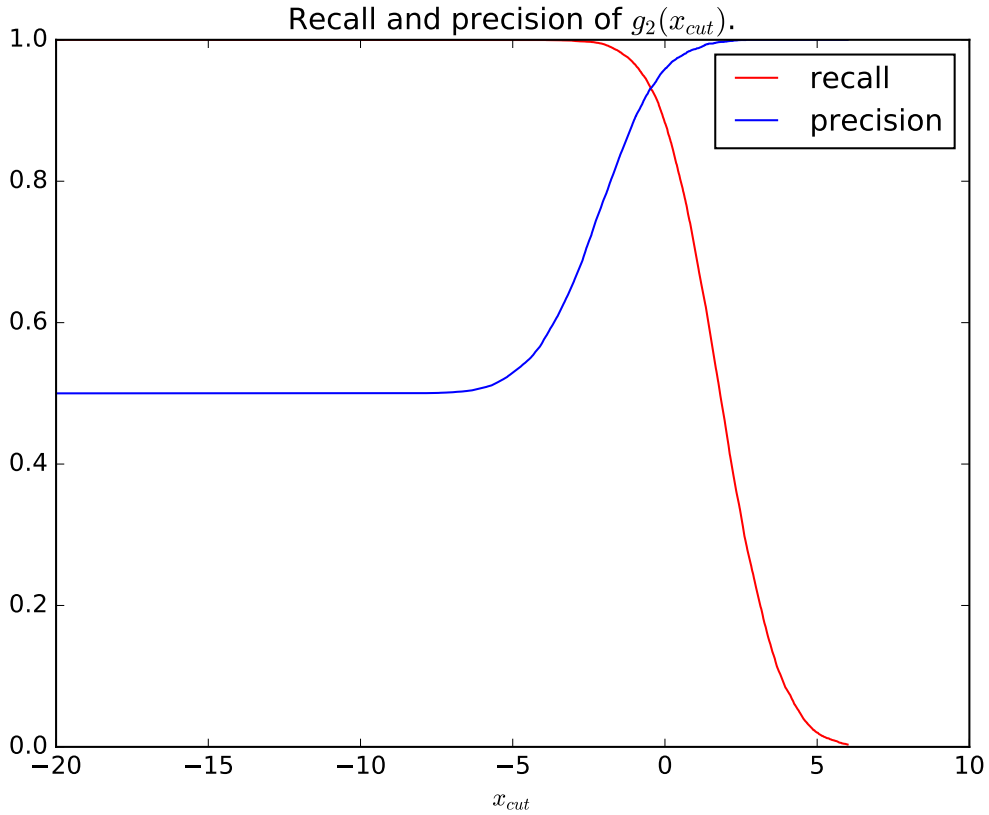

Abbildung 7: Reinheit und Effizienz im Bezug auf g_2

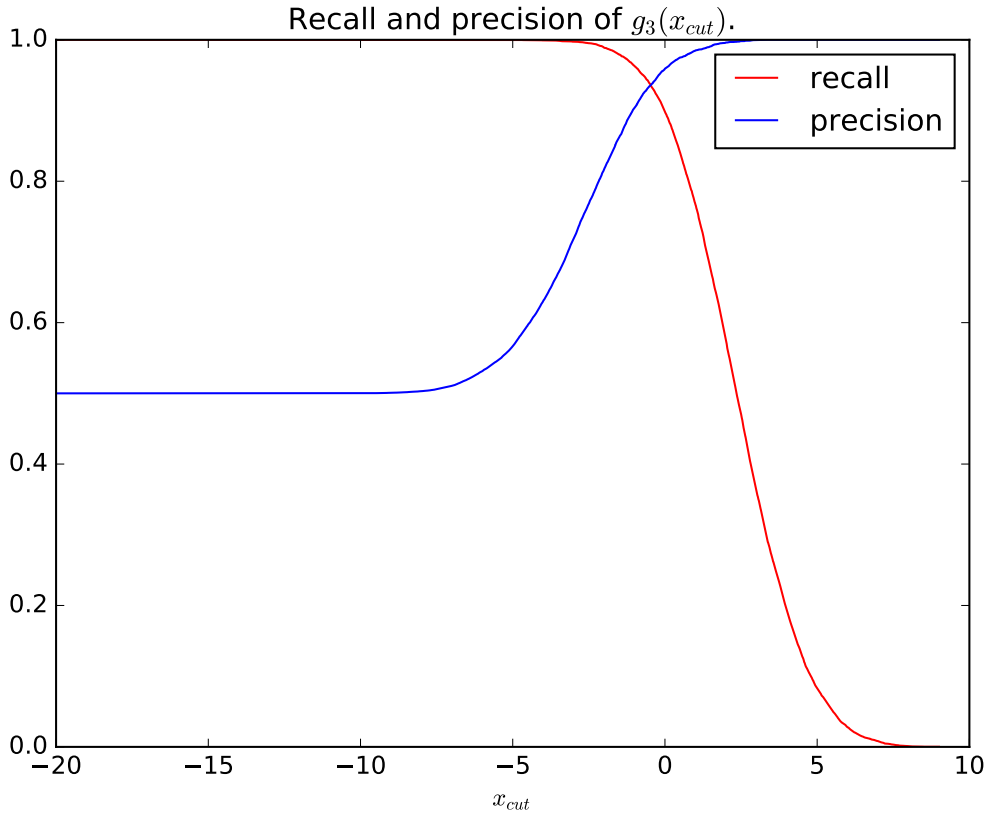

Abbildung 8: Reinheit und Effizienz im Bezug auf g_3

Aufgabe 4

a)

Trennung von Signal und Untergrund etwa beim Ice Cube, um nur die zu beobachtenden Ereignisse herauszufiltern. Es treten etwa atmosphärische Myonen auf, die eine ähnlich Spur wie die zu messenden durch Neutrinos erzeugten Myonen besitzen.

Entfernen inkonsistenter Messungen wenn verschiedene Parameter einer Messung nicht zueinander passen oder zu wenig Parameter bestimmt wurden, muss der Datensatz als ungültig gekennzeichnet werden. Beispielsweise, wenn jemand in einer Umfrage als Alter 15 Jahre und als Beruf Ingenieur angibt.

Entfernung von unnötigen Datensätzen teilweise befinden sich Felder in den Daten, die für manche Messungen unwichtig sind, z.B. kann beim Auslesen einer Datenübertragung von einer Wetterstation der mitgesendete Modellname verworfen werden, wenn du der Luftdruck interessant ist.

Glätten eines Signals wenn ein Wert aus einem stark rauschendem Sensor ausgelesen werden soll sind einzelne Werte wenig aussagekräftig, es interessiert eher der Mittelwert innerhalb eines bestimmten Intervalls. Das kann z.B. durch Faltung des Messsignals mit einer entsprechenden Gaußkurve geschehen.

Auswahl der qualitativ hochwertigsten Daten um die Hardware nicht zu überfordern, etwa bei hoher Ereignisfrequenz (CERN), geringer Übertragungskapazität (Ice Cube Uplink) oder um die Rechenlast zu verringern.

b)

Ja, etwa um Daten lesbarer oder Wahrscheinlichkeiten direkt ersichtlich zu machen.

c)

- Ersetzten durch Mittlung der umgebenden Datenpunkte (à la blur).
- Löschen der entsprechenden Datenzeile.
- Beim überschreiten von Messgrenzen ersetzen durch sehr kleine oder sehr große Werte (wenn das physikalisch vertretbar ist).
- Verwerfen der gesamten Messung.

d)

- Die Daten sollten kompatibel sein.
- Neue Datenstruktur muss Eigenheiten beider Datensätze berücksichtigen, etwa sollte das Abschneiden zusätzlicher Felder von Satz B beim Einfügen in Satz A vermieden werden, wenn diese Daten noch gebraucht werden.