Blatt 7: Regression - lineare Ausgleichsrechnung MNEU

(Theorieteil)

Aufgabe 1 :

Die Punkte

$$P_1 = (-1, -1)\,,\; P_2 = (1, 8)\,,\; P_3 = (6, -1) \tag{1}$$

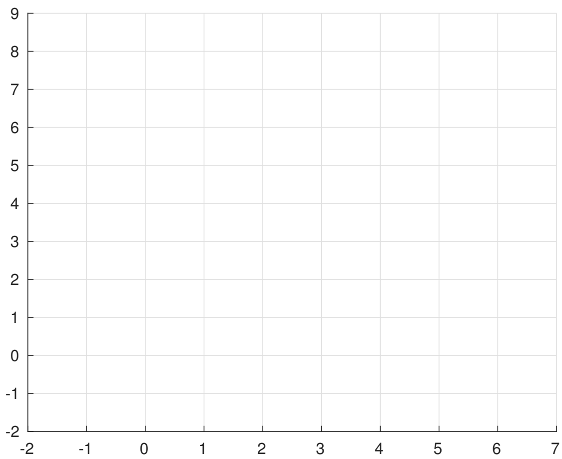
sollen durch eine Gerade $f(x) = m\,x + b$ bestmöglich repräsentiert werden.

(a)

(i) Tragen Sie die Punkte in das rechtsstehende Koordinatensystem ein.

(ii) Zeichnen Sie mit Lineal eine Gerade, von der Sie so nach Bauchgefühl denken, sie könnte die Punkte möglichst gut repräsentieren.

(iii) Zeichnen Sie mit Lineal die Residuenwerte $\mathrm{Res}_i = |P_y^i - f(P_x^i)|$ in die Graphik, berechnen Sie den Quadratfehler



$$Q(m, b) = \sum \mathrm{Res}_i^2 = \qquad \text{und} \qquad E(m, b) = \sqrt{Q(m, b)} =$$

(b) Berechnen Sie nun die Parameter m und b mit Hilfe der Formel aus der Vorlesung, so dass die Gerade $f(x) = m\,x + b$ zu einem minimalen Quadratfehler führt. Wie gross ist der Fehler jetzt?

$$E(m, b) = \sqrt{Q(m, b)} =$$

(c) Zeichnen Sie Ihr Ergebnis ebenfalls in die Graphik und vergleichen Sie "Bauchgefühlwerte" mit den Berechneten. Waren Sie gut?

Aufgabe 2 :

Die Messdaten P_1 bis P_4 in (1) sollen bestmöglich durch die Funktion

$$g(x) = c_1\,x^2 + c_2\,x$$

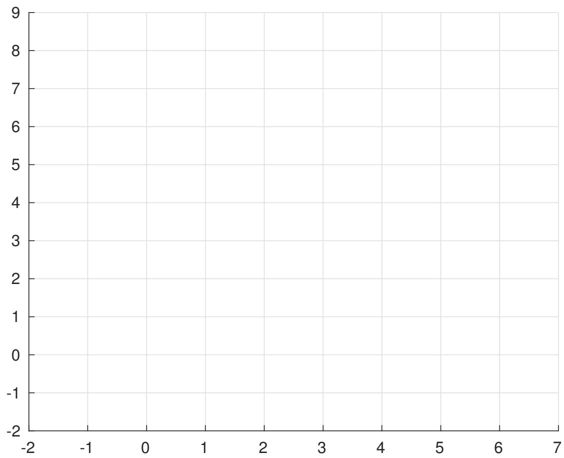
abgebildet werden.

(a)

(i) Tragen Sie die Punkte in das rechtsstehende Koordinatensystem ein.

(ii) Skizzieren Sie den Graphen g bestmöglich nach Bauchgefühl in das Koordinatensystem.

(iii) Tragen Sie abermals die Residuen ein und berechnen Sie den Quadratfehler



$$Q(m,b) = \sum \mathrm{Res}_i^2 = \qquad \text{und den Fehler} \qquad E(m,b) = \sqrt{Q(m,b)} =$$

(b) g ist ja ein quadratisches Polynom und wir haben drei Punkte. Drei Punkte legen ja ein quadratisches Polynom "fest". Was ist hier anders, warum machen wir dennoch eine Regression?

(c) Wie lautet die Funktion Q, die den Quadratfehler dieser Situation beschreibt?

$$Q(c_1, c_2) =$$

(d) Berechnen Sie den Gradienten ∇Q von Q.

$$\nabla Q = \begin{pmatrix} Q_{c_1} \\ Q_{c_2} \end{pmatrix} =$$

(e) Berechnen Sie stationäre Punkte von ∇Q.

(f) Wie groß ist der Quadratfehler

$$Q(c_1, c_2) = \sum \mathrm{Res}_i^2 = \qquad \text{und der Fehler} \qquad E(c_1, c_2) = \sqrt{Q(c_1, c_2)} =$$

(g) Geben Sie die Ergebnisfunktion

$$g(x) =$$

an und zeichnen Sie den Graphen in das Achsenkreuz. Waren Sie gut? ;-)

Aufgabe 3 :

Warum minimieren wir den Quadratfehler Q und nicht den Fehler $E = \sqrt{Q}$?

Aufgabe 4 :

Gegeben sind die Matrix $A = (a_{ij})_{1\leq i,j\leq n}$ und die Vektoren

$$x = (x_1, \ldots, x_n) \qquad\qquad y = (y_1, \ldots, y_n)^T$$

sowie deren arithemtische Mittel S_x und S_y.

(a) Schreiben Sie die folgenden Ausdrücke kurz und prägnant in Form von Normen und Skalarprodukten, sowie in Matlab-Syntax.

Ausdruck	Norm/Skalarprodukt	Matlab-Syntax
$\sum_{i=1}^{n} x_i y_i$	$=$	>>
$\sum_{i=1}^{n} x_i^2$	$=$	>>
$\sum_{i=1}^{n} \lvert x_i \rvert$	$=$	>>
$\frac{1}{n}\sum_{i=1}^{n} x_i$	$=$	>>

(b) Schreiben Sie ein Matlab-Skript mit for-Schleifen (ohne matlabinterne Routinen zu verwenden!) den folgenden Ausdruck:

$$y \cdot A^T\, x^T$$

(Praxisteil)

Aufgabe 5 :

Erstellen Sie ein Verzeichnis `Regression/`.

Aufgabe 6 :

Schreiben Sie ein Programm `Regression/MyLinReg.m`, das

(a) zunächst die vier Punkte in (1) in ein Feld $P \in \mathbb{R}^{4\times 2}$ speichert, dann

(b) eine lineare Ausgleichsrechnung zu den gegbenen Punkten in P durchführt, indem die entsprechende Formel ausgewertet wird.

(c) Das Ergebnis samt gegebener Punkte wird dann in einer Graphik geplottet; und zwar mit der schwarzen Linie für den Graphen und roten Kringeln für die Messdaten.

(d) Der resultierende Fehler werde schön formatiert mit fprintf ausgegeben und zwar so:

`Der Fehler zur Geraden f(x)=m*xb` betraegt err+

Dabei sind `m` und `b` Dezimalzahlen, die auf zwei Nachkommastellen angegeben sind und `err` ist eine Dezimalzahl in e-Form, mit einer Nachkommastelle.

Aufgabe 7 :

(a) Erzeugen Sie ein Verzeichnis `Regression/Daten/`

(b) Speichern Sie dort die Datei
http://axtr.xthelm.de/Lehre/MNEU/Praktikum/Regression/Daten/Filip.dat.

(c) Rufen Sie im Command-Window den Befehl >> `Q=load('Daten/Filip.dat');` auf. Sie sollten sich dabei im Verzeichnis `Regression` befinden. `Q` enthält nun zwei Spalten mit Zahlen. In der ersten Spalte, also `Q(:,1)` x-Werte und in der zweiten Spalte die y-Werte. Wieviele Zeilen enthält `Q`?

(d) Plotten Sie `Q`, indem Sie zweite Spalte (y-Achse) auf erste (x-Achse) auftragen.

(e) Plotten Sie die Punkte in `Q`, indem Sie erste Spalte auf zweite auftragen.

(f) Ersetzen im Programm aus Aufgabe 6 das Belegen des Feldes P durch das Einlesen dieser Datei. Welches ist die Funktion, die bestmöglich fittet?

$$f(x) = \_\_\_\_\_\, x + \_\_\_\_\_$$

Bei jeder weiteren Untersuchung müssen Sie nun nur den entsprechenden Datensatz einlesen und entscheiden welche Spalte der x- und welche der y-Achse zugeordnet werden soll.

Aufgabe 8 :

(a) Speichern Sie die Datei
http://axtr.xthelm.de/Lehre/MNEU/Praktikum/Regression/Daten/Mietspiegel.dat
im Verzeichnis `Daten`. Die Datei enthält Statistiken zu Mietpreis in Abhängigkeit von Wohnungsgröše und Ausstattung. Die Spalten waren wie folgt beschriftet:

```
miete mieteqm flaeche bjahr bad kueche bezv lage zh
```

(b) Führen Sie im Command-Window den Befehl

```
>> Q=load('Daten/Mietspiegel.dat');
```

aus und plotten Sie den Mietpreis in Abhängigkeit des Baujahres.

(c) Spielen Sie ein wenig die plots durch: Miete pro Baujahr, Fläche pro Baujahr, etc.

(d) Sie wollen nun selbst Ihre 120 qm-Wohnung vermieten und einen einigermašen vergleichbaren Preis dafür verlangen; ungeachtet des Baujahres und der sonstigen Ausstattung, rein Preis pro Fläche interressiert Sie. Leider gibt es kein vergleichbares Exemplar und auŝerdem wollen Sie Ihren Preis nicht von einem anderen Beispiel abhängig machen. Sie führen deshalb eine lineare Ausgleichsrechnung durch und ermitteln durch die berechnete Gerade Ihren Mietpreis.

(i) Mit welchen Spalten im Datensatz arbeiten Sie?

$$P_x = \underline{\hspace{2cm}} \quad \text{und} \quad P_y = \underline{\hspace{2cm}}$$

(ii) Wie lautet "Ihre" Gerade?

$$f(x) = \underline{\hspace{1.3cm}}\, x + \underline{\hspace{1.3cm}}$$

(iii) Wie teuer ist Ihre Wohnung?

$$€ = \underline{\hspace{1.3cm}}$$

(Theorieteil)

Lösung 1:

(b)

$$\begin{aligned} P_x &= (-1, 1, 6) & P_y &= (-1, 8, -1) \\ < P_x, P_y > &= 3 & \|P_x\|^2 &= 38 \\ S_x &= 2 & S_y &= 2 \\ N &= 3 \end{aligned}$$

Daraus folgt

$$f(x) = \frac{-9}{26}(x-2) + 2 = -\frac{9}{26}x + \frac{35}{13}$$

und

$$\begin{aligned} Q &= \left(-1 - \frac{9}{26} - \frac{35}{13}\right)^2 + \left(8 + \frac{9}{26} - \frac{35}{13}\right)^2 + \left(-1 + \frac{27}{13} - \frac{35}{13}\right)^2 \\ &= \frac{1}{26^2}(26 + 9 + 70)^2 + \frac{1}{26^2}(208 + 9 - 70)^2 + \frac{1}{13^2}(-13 + 27 - 35)^2 \\ &= \frac{105^2}{26^2} + \frac{147^2}{26^2} + \frac{21^2}{13^2} \approx 50.89 \end{aligned}$$

Lösung 2:

(c)

$$\begin{aligned} Q(c_1, c_2) &= \sum_{i=1}^{3} (P_y^i - g(P_x^i))^2 \\ &= (-1 - g(-1))^2 + (8 - g(1))^2 + (-1 - g(6))^2 \\ &= (-1 - c_1 + c_2)^2 + (8 - c_1 - c_2)^2 + (-1 - 36\,c_1 - 6\,c_2)^2 \end{aligned}$$

(d)

$$\begin{aligned} Q_{c_1}(c_1, c_2) &= -2\,(-1 - c_1 + c_2) - 2\,(8 - c_1 - c_2) - 72\,(-1 - 36\,c_1 - 6\,c_2) \\ &= 58 + 2596\,c_1 + 432\,c_2 \\ Q_{c_2}(c_1, c_2) &= 2\,(-1 - c_1 + c_2) - 2\,(8 - c_1 - c_2) - 12\,(-1 - 36\,c_1 - 6\,c_2) \\ &= -6 + 432\,c_1 + 76\,c_2 \end{aligned}$$

(e)

$$\nabla Q = 0 \quad \Leftrightarrow \quad \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} -6.56 \cdot 10^{-1} \\ 3.81 \end{pmatrix}$$

(f)

$$Q \approx 35.56 \qquad \text{und} \qquad E = \sqrt{Q} \approx 5.96$$

(g)

$$g(x) = -6.56 \cdot 10^{-1}\, x^2 + 3.81\, x$$

Lösung 3:

Die Wurzelfunktion ist streng monoton. Wenn also $E = \sqrt{Q}$ minimal ist, so ist es auch Q und umgekehrt. Mit beiden Ausdrücken erreichen wir unser Ziel. Ohne das Wurzelzeichen ist der Ausdruck etwas handlich, da bei der Ableitung sonst immer der Wurzelausdruck im Nenner mitgeschleppt werden muss.

Lösung 4:

(a)

Ausdruck	Norm/Skalarprodukt	Matlab-Syntax
$\sum_{i=1}^{n} x_i y_i$	$=< x, y >$	`>> x*y`
$\sum_{i=1}^{n} x_i^2$	$=< x, x >= \lVert x\rVert^2$	`>> x*x'`
$\sum_{i=1}^{n} \lvert x_i\rvert$	$= \lVert x\rVert_1$	`>>norm(x,1)`
$\frac{1}{n}\sum_{i=1}^{n} x_i$	$= S_x$	`>>sum(x)/N`

(b)

$$y \cdot A^T\, x^T$$

erhalten wir im Matlab-Skript durch

```
sum = 0;
for i=1:N
   sumj = 0;
   for j=1:N
      sumj = sumj + A(j,i)*x(j);
   end
   sum = sum + y(i)*sumj;
end
```

(Praxisteil)

Lösung 5:

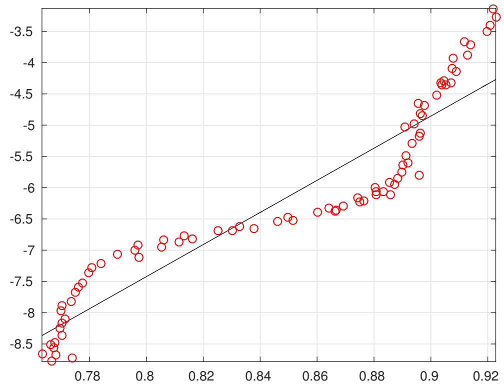
...

Lösung 6:

http://axtr.xthelm.de/Lehre/MNEU/Praktikum/Regression/MyLinReg.m

Lösung 7:

```
Der Fehler zur Gerade f(x)=25.67 * x + (-27.96) betraegt Err = 4.8e+00
```



Lösung 8:

(i) Spalten im Datensatz:

```
Px = P(:,3);
Py = P(:,1);
```

(ii) `Der Fehler zur Gerade f(x)=9.40 * x + (262.45) betraegt Err = 1.7e+04`

(iii) Wie teuer ist Ihre Wohnung?

$$€ = f(120) = 1'390.68$$