DATA SMOOTHING

Maak één van de onderstaande opgaven. Zorg ervoor dat alle opgaven onder de studenten verdeeld worden. Voor elke opgave geeft iemand zich op om die in het practicum op te lossen (email aan `Franky.Backeljauw@uantwerpen.be`). De benodigde software vind je in de GSL library.

1) Gegeven zijn de volgende 25 datapunten $y_i = f(t_i)$ voor $t_i = i, 1 \leq i \leq 25$:

$$y = (\,5.0291, 6.5099, 5.3666, 4.1272, 4.2948, 6.1261, 12.5140, 10.0502, 9.1614, 7.5677, 7.2920, 10.0357, 11.0708, 13.4045, 12.8415, 11.9666, 11.0765, 11.7774, 14.5701, 17.0440, 17.0398, 15.9069, 15.4850, 15.5112, 17.6572\,)^t$$

- Fit de gegevens met een rechte $y(t) = a_0 + a_1 t$ volgens het principe van de kleinste kwadraten (ℓ_2). Plot de gegevens en het model. Plot ook de residuvector. Je stelt een kandidaat outlier vast.
- Fit de gegevens opnieuw met een rechte (ℓ_2), nu zonder rekening te houden met het datapunt dat de outlier oplevert.
- Bereken het model $y(t) = b_0 + b_1 t + b_2 \sin(t)$ ook zonder rekening te houden met de outlier (ℓ_2). Plot opnieuw de gegevens en het model. Plaats de outlier mee op de tekening.

2) Gegeven zijn onderstaande data:

j	x_j	y_j	j	x_j	y_j
0	0.0	−0.80	10	3.6	0.74
1	0.6	−0.34	11	4.7	−0.82
2	1.5	0.59	12	5.2	−1.27
3	1.7	0.59	13	5.7	−0.92
4	1.9	0.23	14	5.8	−0.92
5	2.1	0.10	15	6.0	−1.04
6	2.3	0.28	16	6.4	−0.79
7	2.6	1.03	17	6.9	−0.06
8	2.8	1.50	18	7.6	1.00
9	3.0	1.44	19	8.0	0.00

Neem de lineair onafhankelijke basisfuncties

$$\phi_{3i}(x) = \binom{3}{i} x^i (1-x)^{3-i} \qquad i = 0, 1, 2, 3$$

en bereken de optimale lineaire combinatie

$$\phi(x) = \lambda_0\phi_{3,0}(x) + \lambda_1\phi_{3,1}(x) + \lambda_2\phi_{3,2}(x) + \lambda_3\phi_{3,3}(x)$$

die de fout

$$\sqrt{\sum_{j=0}^{19}(\phi(x_j) - y_j)^2}$$

minimaal maakt.

Geef ook het conditiegetal van de rechthoekige matrix in het kleinste kwadratenprobleem. Plot de berekende $\phi(x)$ samen met de datapunten.

3) Vervang nu de basisfuncties $\phi_{3,i}(x)$ in bovenstaande opgave, enerzijds door de veeltermen

$$T_i(x) = 2xT_{i-1}(x) - T_{i-2}(x) \qquad i = 0, \dots, 3$$
$$T_0(x) = 1 \qquad T_1(x) = x$$

en anderzijds door de monomiale veeltermbasis $x^i, i = 0, \dots, 3$. Bereken de optimale modellen

$$\sum_{i=0}^{3} \tau_i T_i(x) \qquad \sum_{j=0}^{19}\left(y_j - \sum_{i=0}^{3}\tau_i T_i(x)\right)^2 \text{ minimaal}$$

en

$$\sum_{i=0}^{3} \mu_i x^i \qquad \sum_{j=0}^{19}\left(y_j - \sum_{i=0}^{3}\mu x^i\right)^2 \text{ minimaal}$$

Vergelijk de conditiegetallen en bereken de residuvectoren van beide modellen.

4) Gegeven zijn de volgende 7 datapunten die resulteren uit metingen verricht in een oliereservoir:

x	0.635	1.435	2.235	3.035	3.835	4.635	5.435
y	7.50	4.35	2.97	2.20	1.70	1.28	1.00

- Maak een grafiek van de datapunten. We stellen een exponentieel dalend verloop vast. Plot nu ook $\log(y_i), i = 1, \dots, 7$. Deze grafiek suggereert duidelijk een eenvoudig model $g(x)$ voor $\log(y(x))$. Bereken dit model in de zin van de kleinste kwadraten (ℓ_2). Je kan nu $\exp(g(x))$ gebruiken als model voor $y(x)$. Plot $\exp(g(x))$ samen met de oorspronkelijke data.

5) Aan het National Institute of Standards (NIST) te Washington werden een aantal standaardizatietesten uitgevoerd waarvan eentje de dataset opleverde die je kan downloaden van `www.itl.nist.gov/div898/strd/lls/data/Filip.shtml`. Het gaat om 82 observaties van een grootheid $y(x)$ die zich redelijk laat benaderen (ℓ_2) door een veeltermmodel van graad 10. Bereken de coefficienten van dat model.