

Minsta-kvadratmetoden

1 Inledning

Ett ofta förekommande problem inom teknik och vetenskap är modellering av data, dvs. att koppla samman mätdata med en formel eller kurva som man vill verifiera eller bygga upp.

Som exempel tar vi det klassiska problemet att anpassa en rät linje $y = a + b \cdot t$ till givna mätdata (t_i, y_i) , $i = 1, \dots, n$. Hur skall vi välja a och b ? Vi kan ju inte få den räta linjen att gå igenom mer än högst två punkter. Problemet vi skall lösa är följande överbestämde ekvationssystem

$$\begin{cases} a + b \cdot t_1 = y_1 \\ a + b \cdot t_2 = y_2 \\ \vdots \\ a + b \cdot t_n = y_n \end{cases}$$

(det är a och b som är de obekanta!) På matrisform,

$$\underbrace{\begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_n \end{bmatrix}}_{\mathbf{A}} \begin{bmatrix} a \\ b \end{bmatrix} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{\mathbf{y}}$$

Grundidén i minsta-kvadratmetoden är att projicera vektorn \mathbf{y} ortogonalt på kolonnrummet för matrisen \mathbf{A} ($\mathcal{R}(\mathbf{A})$) och sedan lösa ekvationen $\mathbf{A}\mathbf{x} = \mathbf{p}$ där \mathbf{p} är projektionen. På så vis får vi en lösning $\hat{\mathbf{x}}$ där avståndet $\|\mathbf{A}\hat{\mathbf{x}} - \mathbf{y}\|$ är det minsta möjliga och väljer vi nu $a = \hat{x}_1$, $b = \hat{x}_2$ så har vi minimerat summan av kvadraterna på avvikelserna:

$$\sum_{i=1}^n (a + b \cdot t_i - y_i)^2$$

Lösningen $\hat{\mathbf{x}}$ till problemet $\mathbf{A}\mathbf{x} = \mathbf{p}$ ovan säges vara *minsta-kvadratlösningen* till det ursprungliga ekvationssystemet $\mathbf{A}\mathbf{x} = \mathbf{y}$. Vi finner den genom att lösa den så kallade *normalekvationen*

$$\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{A}^T \mathbf{y}$$

(Notera att vi inte behöver bestämma projektionen \mathbf{p} för att bestämma $\hat{\mathbf{x}}$.)

Det *kvadratiska medelfelet*, den genomsnittliga avvikelserna, ges av

$$\varepsilon = \|\mathbf{A}\hat{\mathbf{x}} - \mathbf{y}\| / \sqrt{n}$$

där n är antalet mätdata.

Låt oss nu bestämma den räta linje som i minsta-kvadratmening är bäst anpassad till följande data:

t	-1	0	1	2	3	4	6	7
y	-0.75	0.3	3	4	5.6	7	6.4	8.4

Då är

$$\mathbf{A}^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 2 & 3 & 4 & 6 & 7 \end{bmatrix}, \quad \mathbf{A}^T \mathbf{A} = \begin{bmatrix} 8 & 22 \\ 22 & 116 \end{bmatrix}, \quad \mathbf{A}^T \mathbf{y} = \begin{bmatrix} 33.95 \\ 153.75 \end{bmatrix}$$

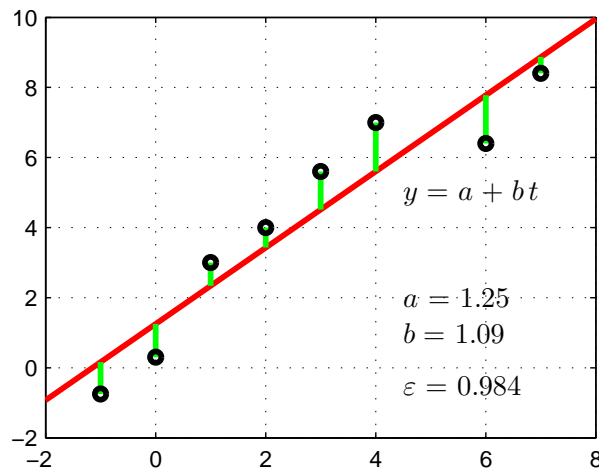
och normalekvationen är

$$\begin{bmatrix} 8 & 22 \\ 22 & 116 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 33.95 \\ 153.75 \end{bmatrix}$$

I MATLAB löser vi normalekvationen med backslash-kommandot \ enligt

```
>> td=[-1 0 1 2 3 4 6 7]';           % t-data
>> yd=[-0.75 0.3 3 4 5.6 7 6.4 8.4]'; % y-data
>> A=[ones(size(td)) td];           % Designmatrisen
>> x=A\yd; a=x(1); b=x(2);          % Minsta-kvadratlösningen
>> n=length(td);                     % Antalet mätdata
>> e=norm(A*x-yd)/sqrt(n);          % Kvadratiska medelfelet
```

Vi kan nu rita upp följande figur



Vi minimerar summan av kvadraterna på de lodräta sträckorna.

Uppgift 1. Antag att variablerna t och y uppfyller sambandet $y = a + b \cdot t$. För att bestämma koefficienterna a och b utför vi mätningar av t och y :

t	5	6	7	8	9	10
y	19.5888	23.4043	25.5754	29.1231	31.9575	35.8116

- Lös normalekvationen med minsta-kvadratmetoden i MATLAB. Bestäm också kvadratiska medelfelet.
- Rita upp datapunkterna (t_i, y_i) och den anpassade funktionen $y = a + b \cdot t$ i samma figur.

2 Allmän formulering

Härnäst skall vi bestämma den kurva

$$y = c_1 \cdot f_1(t) + c_2 \cdot f_2(t) + \dots + c_k \cdot f_k(t)$$

där f_1, f_2, \dots, f_k är kända funktioner, som i minsta-kvadratmening är bäst anpassad till givna mätdata

$$\begin{array}{c|cccc} \mathbf{t} & t_1 & t_2 & \dots & t_n \\ \hline \mathbf{y} & y_1 & y_2 & \dots & y_n \end{array}$$

Matrisformuleringen av det överbestämde ekvationssystem vi intresserar oss för ges av

$$\underbrace{\begin{bmatrix} f_1\left(\begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix}\right) & f_2\left(\begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix}\right) & \dots & f_k\left(\begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix}\right) \end{bmatrix}}_{\mathbf{A} \text{ (designmatrisen)}} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{\mathbf{y}}$$

Precis som tidigare finner vi minsta-kvadratlösningen genom att lösa normalekvationen.

Låt oss återgå till vårt första exempel ovan och istället anpassa ett andragradspolynom,

$$y = c_1 + c_2 \cdot t + c_3 \cdot t^2,$$

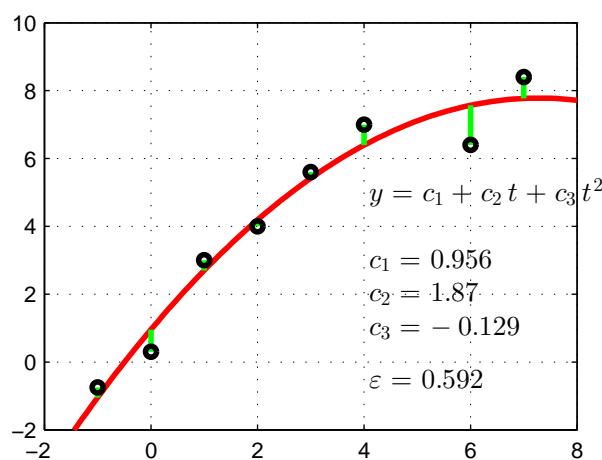
till givna mätpunkter. Här är

$$f_1(t) = 1, f_2(t) = t, f_3(t) = t^2$$

och designmatrisen \mathbf{A} och minsta-kvadratlösningen skapas i MATLAB på liknande sätt som tidigare

```
>> td=[-1 0 1 2 3 4 6 7]';           % t-data
>> yd=[-0.75 0.3 3 4 5.6 7 6.4 8.4]'; % y-data
>> A=[ones(size(td)) td td.^2];     % Designmatrisen
>> x=A\yd;                          % Minsta-kvadratlösningen
>> n=length(td);                   % Antalet mätdata
>> e=norm(A*x-yd)/sqrt(n);         % Kvadratiska medelfelet
```

Vi kan nu rita figuren



Vi minimerar summan av kvadraterna på de lodräta sträckorna.

Uppgift 2. Tabellen nedan visar medeltemperaturen under tidsperioden 1961-1990 för årets tolv månader i Göteborg.

Månad	1	2	3	4	5	6	7	8	9	10	11	12
Temp °C	-0.9	-0.9	2.0	6.0	11.6	15.5	16.6	16.2	12.8	9.1	4.4	1.0

Eftersom detta är ett periodiskt förlopp skall vi anpassa följande modell till mätdata:

$$y(t) = c_1 + c_2 \sin(\omega t) + c_3 \cos(\omega t)$$

Här är

$$f_1(t) = 1, f_2(t) = \sin(\omega t), f_3(t) = \cos(\omega t)$$

Välj ett lämpligt värde på ω (tänk på periodlängder) och bestäm c_1, c_2 och c_3 med minsta-kvadratmetoden. Bestäm också kvadratiska medelfelet. Rita grafen av funktionen $y(t)$ och rita ut mätdata i samma bild så att man kan se om modellen ansluter väl till mätdata.

Uppgift 3. Tabellen nedan visar (en del av) utfallet y av ett försök då man varierar två faktorer s och t .

s	41.9	43.4	43.9	44.5	47.3	47.5	47.9	50.2	...
t	29.1	29.3	29.5	29.7	29.9	30.3	30.5	30.7	...
y	251.3	251.3	248.3	267.5	273.0	276.5	270.3	274.9	...

(a). Data-filen `labdata.mat` på MATLAB-hemsidan innehåller hela uppsättningen data lagrad i tre variabler `sd`, `td` och `yd`. Hämta filen och ladda in i MATLAB med `load('labdata')`.

(b). Anpassa en linjär modell

$$y(s, t) = a + b \cdot s + c \cdot t$$

till data med minsta-kvadratmetoden. Rita ut data med en markör och ett plan som beskriver modellen.

1 Målsättning

Avsikten med denna laborationen att visa hur man kan anpassa linjära modeller till mätdata genom att sätta upp och lösa överbestämda linjära ekvationssystem i MATLAB med minsta-kvadratmetoden.

2 Kommentarer och förklaringar

Vanligtvis skall man inte bild normalekvationerna explicit i MATLAB för att sedan lösa dem, man skall istället lösa de överbestämde ekvationssystemen direkt med backslash-kommandot. När MATLAB löser de överbestämde systemen används metoder som är mycket effektiva och som är lämpade för datorberäkningar.

Modellen i sista uppgiften beror av två variabler, dvs. vi har en funktion i flera variabler. Sådana skall vi studera i flervariabelkursen.

3 Lärandemål

Efter denna laboration skall du kunna

- sätta upp det överbestämde ekvationssystemet som bestämmer parametrarna i modellen
- lösa överbestämde ekvationssystem med minsta-kvadratmetoden med hjälp av backslash-kommandot
- på lämpligt sätt rita upp mätdata och modell