Közelítő és szimbolikus számítások I. Gyakorlat

Gyakorlatvezető: London András

11. Gyakorlat

Legkisebb négyzetek módszere (LNM)

A **regresszió analízis** standard eljárása:

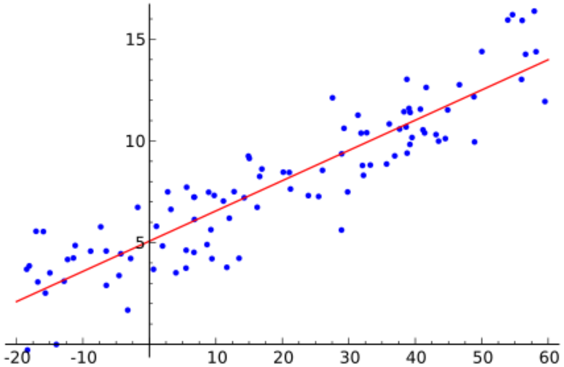
- Adott egy túlhatározott rendszer (pl. egy lineáris egyenletrendszer, ahol több az egyenlet, mint a változó) → keressünk közelítő megoldást

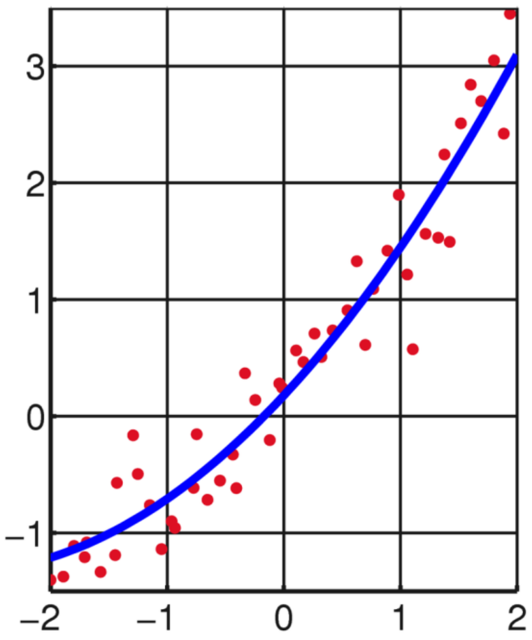
Legkisebb négyzetek: a megoldás olyan, hogy minimalizálja az egyes egyenletek esetén felmerülő (közelítésből származó) hibák négyzetösszegét

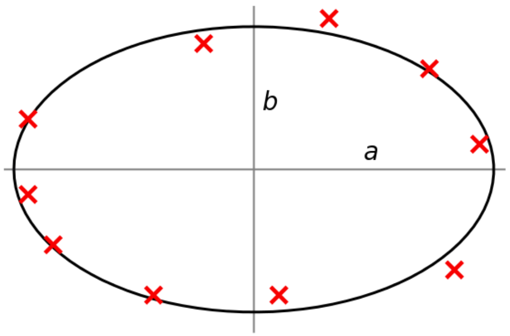
LNM - Motiváció

Függvényillesztés – **data fitting**

- Az adatpontokra legjobban illeszkedő függvény keressük „legkisebb négyzetek” értelemben







LNM - Függvényillesztés

- Adottak $(x_i, y_i), i = 1, 2, \ldots n$ adatpontok
- $y \sim f(x)$ típusú összefüggést keresünk a pontok között
- A modellfüggényünk $f(x, \beta)$, ahol β paraméterek halmaza
- A közelítés hibája az egyes pontokban:

$$r_i = y_i - f(x_i, \beta)$$

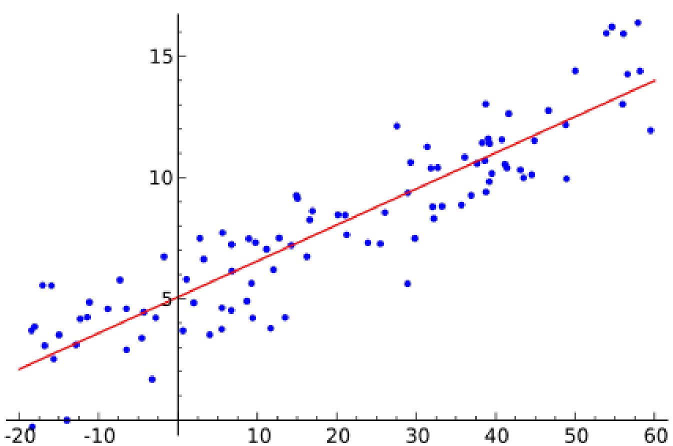
- Hogyan válasszuk meg β-t, hogy

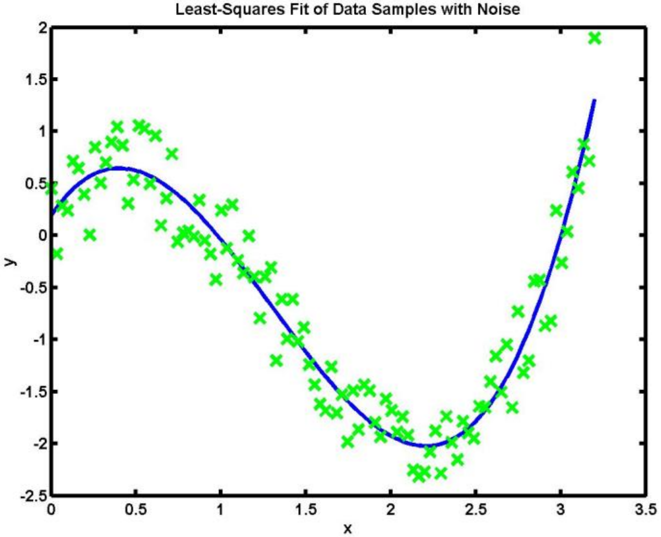
$$\sum_{i=1}^{n} r_i^2 \to \min$$

LNM - Függvényillesztés

Például

- $f(x, \beta) = \beta_0 + \beta_1 x_1$
- $f(x, \beta) = \sum_{j=1}^{m} \beta_i \phi(x_i)$





LNM Matlab (Least squares fit)

1. Lineáris egyenletrendszer közelítő megoldás

```
>>lsqlin()
```

2. Lineáris függvényillesztés

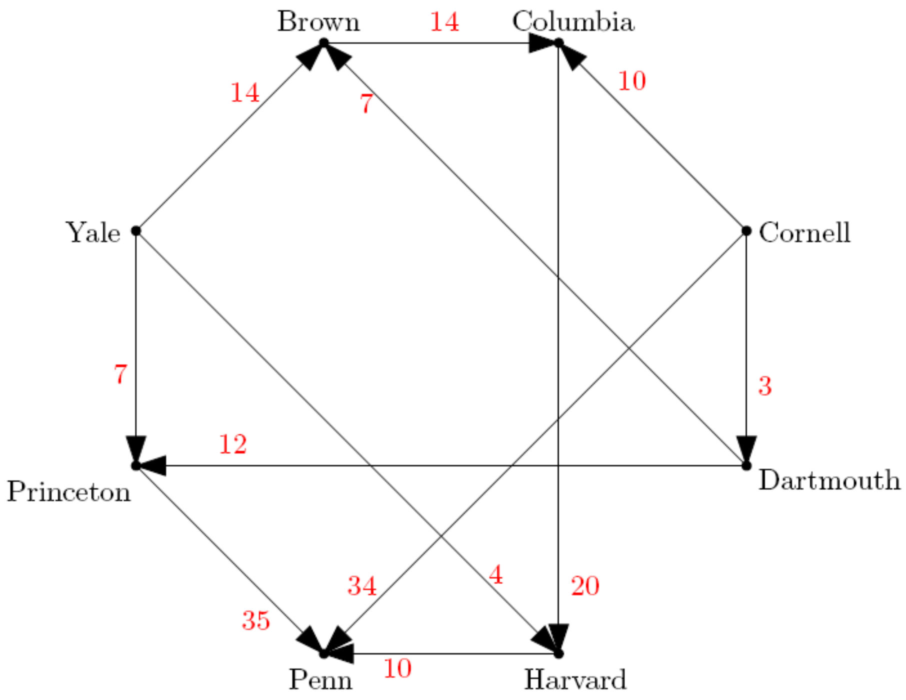
```
>>polyfit(x,y,2)
```

3. Nemlineáris függvényillesztés LNM-mel

```
>>polyfit(x,y,n)
>>lsqnonlin()
```

Egy érdekes alkalmazás

Massey rangsoroló módszere (Kenneth Massey, 1997, BSc szakdolgozat)



Ki a legjobb csapat?

Mi a csapatok közti rangsor?

Hogy alakulnak a relatív erőviszonyok?

Massey legkisebb négyzetek módszere

- A csapatok értékelését a pontbeli különbségek alapján szeretnénk tenni
- r_i = az *i*-edik csapat értékelése
- Ha *i* játszik *j* ellen, akkor a pontkülönbség függjön $r_i - r_j$ értéktől, pl.
  - $r_{Brown} - r_{Yale} = 14$
  - $r_{Columbia} - r_{Brown} = 14$
  - $r_{Columbia} - r_{Cornell} = 10$
- 12 egyenlet, 8 változó → van egyértelmű megoldás?

Massey legkisebb négyzetek módszere

Csapatok száma = n, összes lejátszott meccs = m

- m x n mátrix B
- Vektor $\vec{v} = (v_1, v_2, \dots, v_m)$
- Értékelő vektor $\vec{r} = (r_1, r_2, \dots, r_n)$

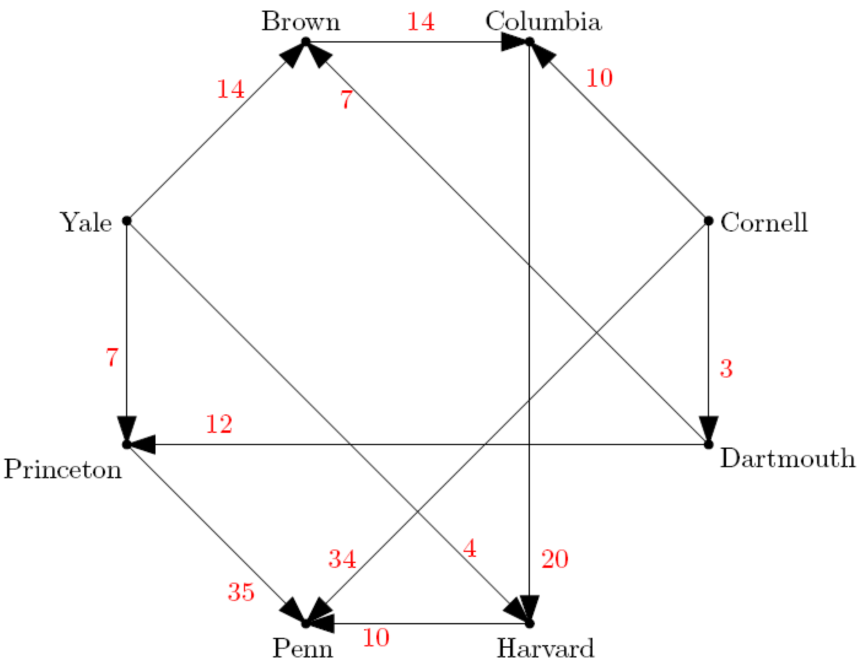
Ha a k-adik meccsen i megverte j-t, akkor

- $B_{ki} = 1, B_{kj} = -1,$ (és $B_{kl} = 0$ if $l \neq i, j$)
- $v_k =$ a pontkülönbség

A Massey-féle lineáris egyenletrendszer $B\vec{r} = \vec{v}$ - nincs egyértelmű megoldás

→ **Oldjuk meg $\|B\vec{r} - \vec{v}\|$ minimalizálási problémát**

Massey megoldás



Team	Massey Rating
Penn	25.25
Harvard	10.75
Columbia	0
Princeton	-3
Brown	-3.75
Yale	-7
Cornell	-11
Dartmouth	-11.25