

# Adatmodellezés, függvényillesztés

Kormányos Andor

Komplex Rendszerek Fizikája Tanszék

2023. március 2.

## Kísérlet:

- bizonyos mennyiségeket mérünk
- sokszor más mennyiségek függvényében, pl. az idő vagy hely
- egy elméleti modell összefüggéseket feltételez a mérhető mennyiségek között

## Problémák:

- az adatokat torzítás és mérési hiba és zaj terheli

## Célok:

- ellenőrizni a fizikai törvény helyességét a mérési eredmények alapján
- megadni a modellben megjelenő fizikai konstansok értékét
- mennyire lehetünk abban bizonyosak, hogy a fizikai törvény valóban leírja a méréseket? Hogyan dönthetünk alternatív modellek között?

Fóliák:

- Mérési hiba, zaj
- Statisztikus hiba, hibaterjedés

## Kérdés:

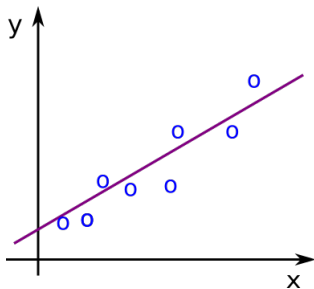
- adott egy  $\{(\mathbf{x}^{(i)}, y^{(i)})\}$  adathalmaz,  $i = 1 \dots N$  mérési pontok
- erre szeretnék illeszteni egy  $h(\mathbf{x}; \mathbf{a})$  függvényt
- hogyan határozhatjuk meg a  $\mathbf{a} = (a_0, a_1 \dots a_k)$  paramétervektort?
- hogyan számszerűsíthető, hogy mennyire illeszkedik jól a modell a mérésekhez?

**Példa:** egyenesillesztés  $\Rightarrow h(x; \mathbf{a}) = a_0 + a_1x$

# Függvényillesztés

Intuíció:

- az a  $h(\mathbf{x}; \mathbf{a})$  írja le legjobban az adatokat, amelyekre a mért  $y^{(i)}$  adat és a hipotézis  $h(\mathbf{x}^{(i)}; \mathbf{a})$  értéke **átlagosan** a legkevésbe tér el
- $h(\mathbf{x}; \mathbf{a})$  nem kell, hogy minden  $y^{(i)}$  ponton átmenjen, csak az, hogy átlagosan kicsi legyen a különbség



# Költségfüggvény

## Stratégia:

- definiálunk egy  $\text{cost}(y^{(i)}, h(\mathbf{x}^{(i)}; \mathbf{a}))$  függvényt
- ez adja meg, hogy mennyire tér el az  $y^{(i)}$  adattól a  $h(\mathbf{x}^{(i)}; \mathbf{a})$  hipotézis
- definiáljuk a  $J(\mathbf{a}; \mathbf{x}^{(i)}, y^{(i)})$  költségfüggvényt:

$$J(\mathbf{a}; \mathbf{x}^{(i)}, y^{(i)}) = \frac{1}{N} \sum_{i=1}^N \text{cost}(y^{(i)}, h(\mathbf{x}^{(i)}; \mathbf{a}))$$

- adott  $h(\mathbf{x}; \mathbf{a})$  hipotézis (függvényalak) mellett az eltérés átlagosan akkor lesz a legkisebb, ha megkeressük a  $J(\mathbf{a}; \mathbf{x}^{(i)}, y^{(i)})$  minimumát a  $a_0, a_1 \dots a_k$  függvényében

$$\min_{\mathbf{a}} [J(\mathbf{a}; \mathbf{x}^{(i)}, y^{(i)})] = ?$$

- $\Rightarrow$  függvény minimalizációs probléma

## Kérdések

- milyen  $\text{cost}(y^{(i)}, h(\mathbf{x}^{(i)}; \mathbf{a}))$ -t használjunk?
  - a legkisebb négyzetek módszer gyakran alkalmazott
  - másfajta költségfüggvény is létezik, lásd pl **logisztikus regresszió**
- milyen algoritmussal tudjuk megtalálni egy függvény minimumát vagy maximumát? (ezzel most nem foglalkozunk)