

Simbolička regresija

Seminar

Dino Šantl

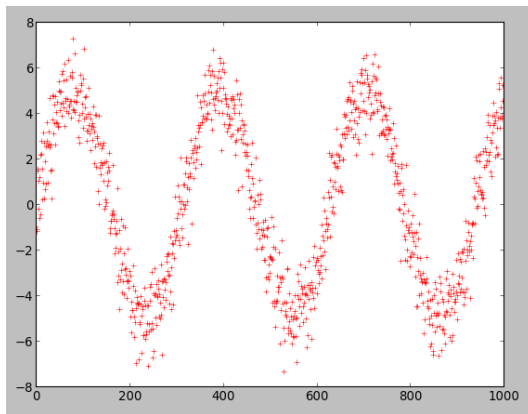
Mentor: Prof. dr. sc. Domagoj Jakobović

Fakultet elektrotehnike i računarstva

Zagreb, lipanj 2013.

Uvod

- ▶ Postupak pronalaženja matematičkog izraza iz empirijskih podataka



Uvod

- ▶ Osim traženja parametra modela traži se i sam model
- ▶ Model nije pretpostavljen kao npr. kod linearne regresije

Linearna regresija:

$$h(x) = \mathbf{w}_1 \cdot x^2 + \mathbf{w}_2 \cdot x + \mathbf{w}_0 \quad (1)$$

Simbolička regresija:

$$h(x) = 2 \cdot \ln(x) + 0.5 \cdot \sin(x) \quad (2)$$

ili

$$h(x) = e^x \cdot \cos^2(x) \quad (3)$$

Evolucijski algoritam

- ▶ Prostor stanja prevelik za iscrpno pretraživanje (engl. *brute force*) - heuristika → evolucijski algoritam
- ▶ Evolucijski algoritam
 - ▶ **jedinka**
 - ▶ populacija
 - ▶ selekcija
 - ▶ križanje
 - ▶ mutacija
 - ▶ dobrota jedinke - **račun pogreške**

Evolucijski algoritam

- ▶ Genetsko programiranje

Evolucijski algoritam

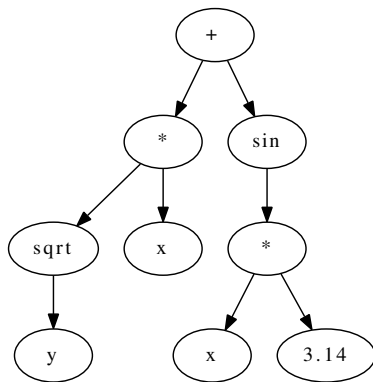
- ▶ Genetsko programiranje
- ▶ Gramatička evolucija

Evolucijski algoritam

- ▶ Genetsko programiranje
- ▶ Gramatička evolucija
- ▶ Analitičko programiranje

Genetsko programiranje

- ▶ Jedinka prikazana stablom: $\sqrt{y} \cdot x + \sin(3.14 \cdot x)$



Gramatička evolucija

- ▶ Jedinka prikazana u binarnom obliku - 0011 1110 0101 1000
- ▶ Formira se niz iz kojeg se gradi izraz - [3, 14, 5, 8]
- ▶ Izraz se gradi na temelju zadane gramatike (npr. BNF oblik)

Gramatička evolucija

- ▶ Jedinka prikazana u binarnom obliku - 0011 1110 0101 1000
- ▶ Formira se niz iz kojeg se gradi izraz - [3, 14, 5, 8]
- ▶ Izraz se gradi na temelju zadane gramatike (npr. BNF oblik)

<izraz>	::=	<izraz> <operator> <izaraz>	(0)
		<var>	(1)
		<func>(<izraz>)	(2)
<operator>	::=	+	(0)
		-	(1)
<var>	::=	X	(0)
		3.14	(1)
<func>	::=	Sin	(0)
		Log	(1)

Analitičko programiranje

- ▶ Jedinka zapisana pomoću niza - [3, 14, 42, 5]
- ▶ Definira se skup u kojem se nalaze operatori, funkcije, varijable i konstante
- ▶ Na temelju broja u nizu određuje se element u matematičkom izrazu
- ▶ { +, -, *, / , sin, cos, sqrt, X, Y, 3.14 }

Problemi u simboličkoj regresiji

- ▶ Predugi izrazi - regularizacija
- ▶ Nedefiniran izlaz funkcije - **intervalna aritmetika**
- ▶ Konstante - **linearno skaliranje**

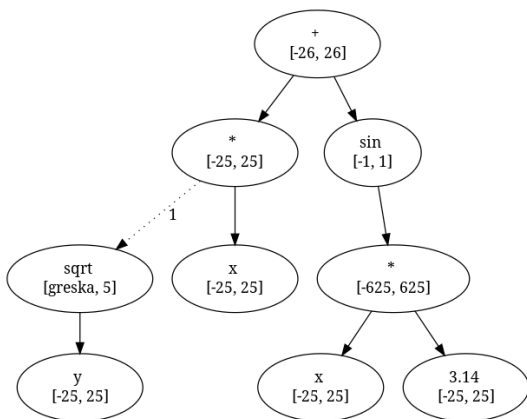
Intervalna aritmetika

- ▶ Raniji pokušaji: sigurni operatori: $f(x) = \frac{1}{x-5}$, $x = 5$?

Intervalna aritmetika

- ▶ Raniji pokušaji: sigurni operatori: $f(x) = \frac{1}{x-5}$, $x = 5$?
- ▶ Prikaz jedinke pomoću stabla
- ▶ Računa se izlaz generirane funkcije ali u intervalnoj aritmetici
- ▶ Ako je interval beskonačan ili se pojavi kompleksni broj potrebno je obrisati podstablo
- ▶ Korak prije računanja pogreške

Intervalna aritmetika



Linearno skaliranje

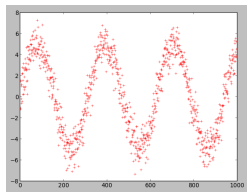
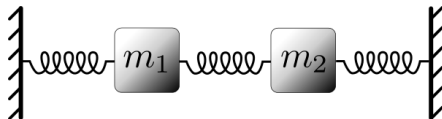
- ▶ Funkcija ima "dobar" oblik ali nije dobro skalirana
- ▶ Primjer: $h(x) = 12.42 \cdot x^2 + 10^8$

Linearno skaliranje

- ▶ Funkcija ima "dobar" oblik ali nije dobro skalirana
- ▶ Primjer: $h(x) = 12.42 \cdot x^2 + 10^8$
- ▶ Izraz se zapiše u obliku $h'(x) = \mathbf{a} \cdot h(x) + \mathbf{b}$
- ▶ Traže se konstante \mathbf{a} i \mathbf{b} koje minimiziraju srednju kvadratnu pogrešku (dobrota jedinke)

$$E = \frac{1}{N} \sum_{i=1}^N (y_i - (a \cdot h(\mathbf{x}_i) + b))^2 \quad (4)$$

Izlučivanje fizikalnih zakona iz empirijskih podataka



Bez znanja fizike:

$$a - 0.05v - 7.8x = 0$$

(5)

Izlučivanje fizikalnih zakona iz empirijskih podataka

- ▶ Empirijski podaci

Izlučivanje fizikalnih zakona iz empirijskih podataka

- ▶ Empirijski podaci
- ▶ Numeričko računanje parcijalnih derivacija $\frac{\Delta y}{\Delta x}$

Izlučivanje fizikalnih zakona iz empirijskih podataka

- ▶ Empirijski podaci
- ▶ Numeričko računanje parcijalnih derivacija $\frac{\Delta y}{\Delta x}$
- ▶ Simboličkom regresijom traži se funkcija $f(x, y)$ koja najbolje opisuje parcijalne derivacije: $\frac{\delta y}{\delta x} = \frac{\frac{\partial f}{\partial x}}{\frac{\partial f}{\partial y}}$

Izlučivanje fizikalnih zakona iz empirijskih podataka

- ▶ Empirijski podaci
- ▶ Numeričko računanje parcijalnih derivacija $\frac{\Delta y}{\Delta x}$
- ▶ Simboličkom regresijom traži se funkcija $f(x, y)$ koja najbolje opisuje parcijalne derivacije: $\frac{\delta y}{\delta x} = \frac{\frac{\partial f}{\partial x}}{\frac{\partial f}{\partial y}}$
- ▶ Račun pogreške:

$$E = -\frac{1}{N} \sum_{i=1}^N \log \left[1 + \text{abs} \left(\frac{\delta y_i}{\delta x_i} - \frac{\Delta y_i}{\Delta x_i} \right) \right] \quad (6)$$

Izlučivanje fizikalnih zakona iz empirijskih podataka

- ▶ Zanemarivanje nekih funkcija \rightarrow aproksimacije

Izlučivanje fizikalnih zakona iz empirijskih podataka

- ▶ Zanemarivanje nekih funkcija \rightarrow aproksimacije
- ▶ Računanje traje od par sati do nekoliko dana

Izlučivanje fizikalnih zakona iz empirijskih podataka

- ▶ Zanemarivanje nekih funkcija → aproksimacije
- ▶ Računanje traje od par sati do nekoliko dana
- ▶ Pokazuje se da je točno rješenje u trenutku kada se dogodi veliki skok u pogrešci

Hvala na pažnji!

Pitanja?

