

Uvod v strojno učenje

Jure Žabkar


jure.zabkar@fri.uni-lj.si



A.I. LAB
Ljubljana

Uvod v strojno učenje

Jure Žabkar, jure.zabkar@fri.uni-lj.si

April 14	April 21 17:00 	April 28	
Maj 5	Maj 12	Maj 19	Maj 26

Vsebina

Strojno učenje

Nadzorovano učenje

Regresija

Linearna regresija

Cenitev nepremičnin klasično

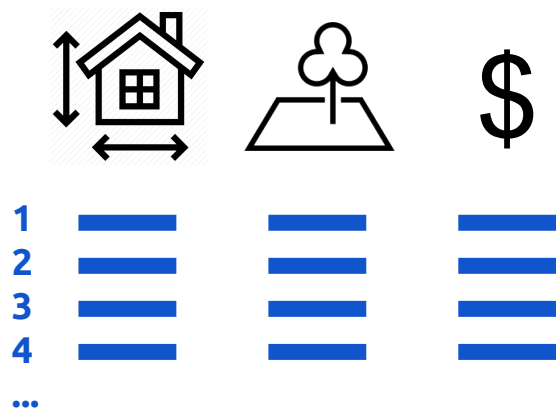
$$\text{\$} = 1.2 \times \text{house icon} + 3.14 \times \text{tree icon}$$

Cenitev nepremičnin s strojnim učenjem

$$\text{\$} = A \times \text{house_size} + B \times \text{land_area}$$

The equation represents a linear model for real estate pricing. The dollar sign (\$) is in blue. The coefficient A is in orange, followed by a blue multiplication sign (x). The house icon, which includes vertical and horizontal double-headed arrows indicating size, is enclosed in a light gray dashed box. This is followed by a blue plus sign (+), an orange coefficient B, another blue multiplication sign (x), and a tree icon on a trapezoidal base representing land area.

Cenitev nepremičnin s **strojnim učenjem**



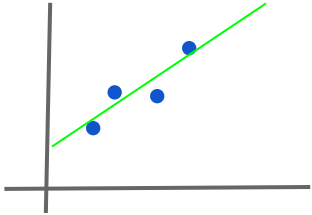
Strojno učenje

Nadzorovano

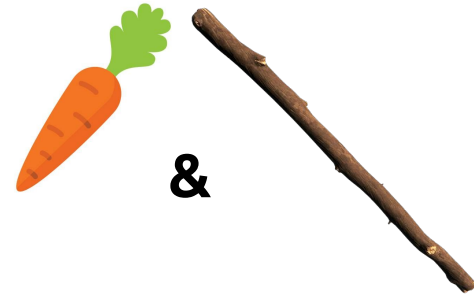
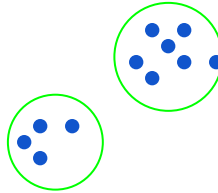
Nenadzorovano

**Spodbujevalno
učenje**

Regresija, Klasifikacija



Gručenje, povezovalna pravila



Terminologija

- **Razred** (angl. Class, Label, Outcome)
odvisna spremenljivka, ki jo napovedujemo,
- **Atribut** ali značilka (angl. Feature)
vhodna / neodvisna spremenljivka
- **Primer** (angl. Example)
vektor vrednosti atributov in razreda
osnovni element množice podatkov

Regresija

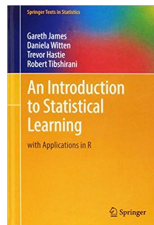
Množica podatkov **D**:

$\{\mathbf{e} = [x_1, \dots, x_n, y] \mid x_i \text{ vrednosti atributov}, y \text{ vrednost razreda}\}$

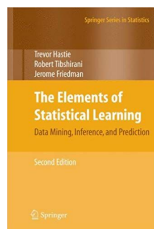
Vrednosti **atributov**: lahko **diskretne** ali **zvezne** vrednosti

Vrednost **razreda**: **zvezna**

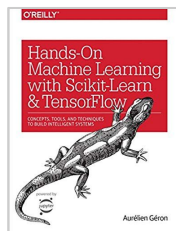
Literatura



3.1 - 3.3



2.3.1, 3.2

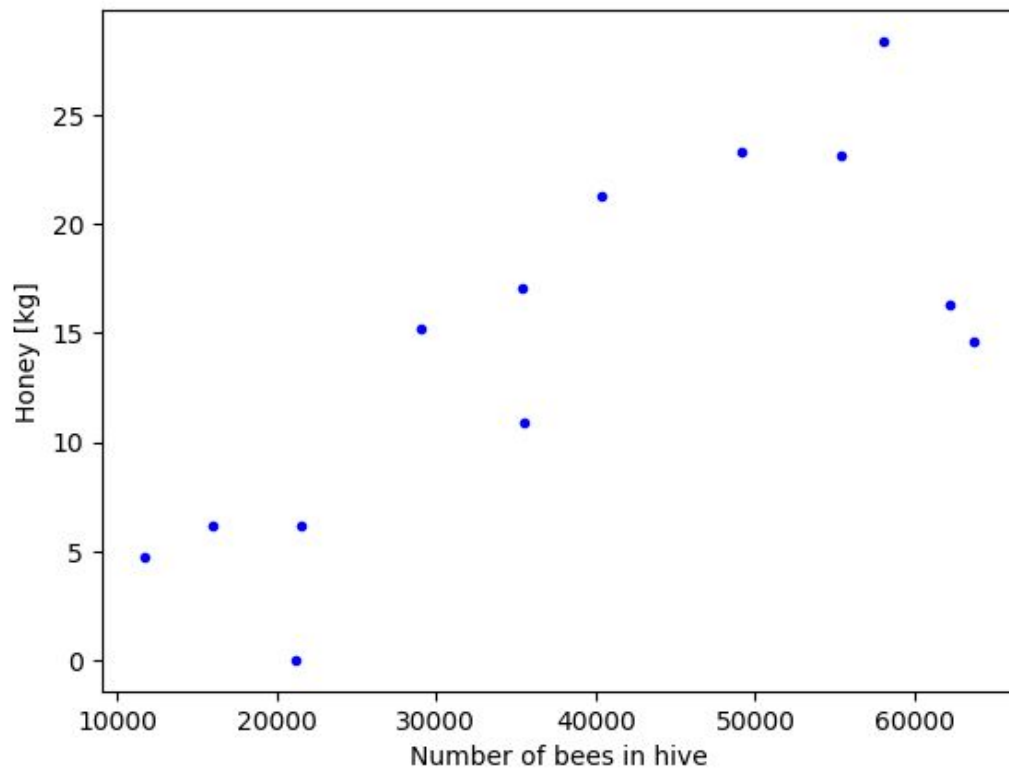


str. 102-130

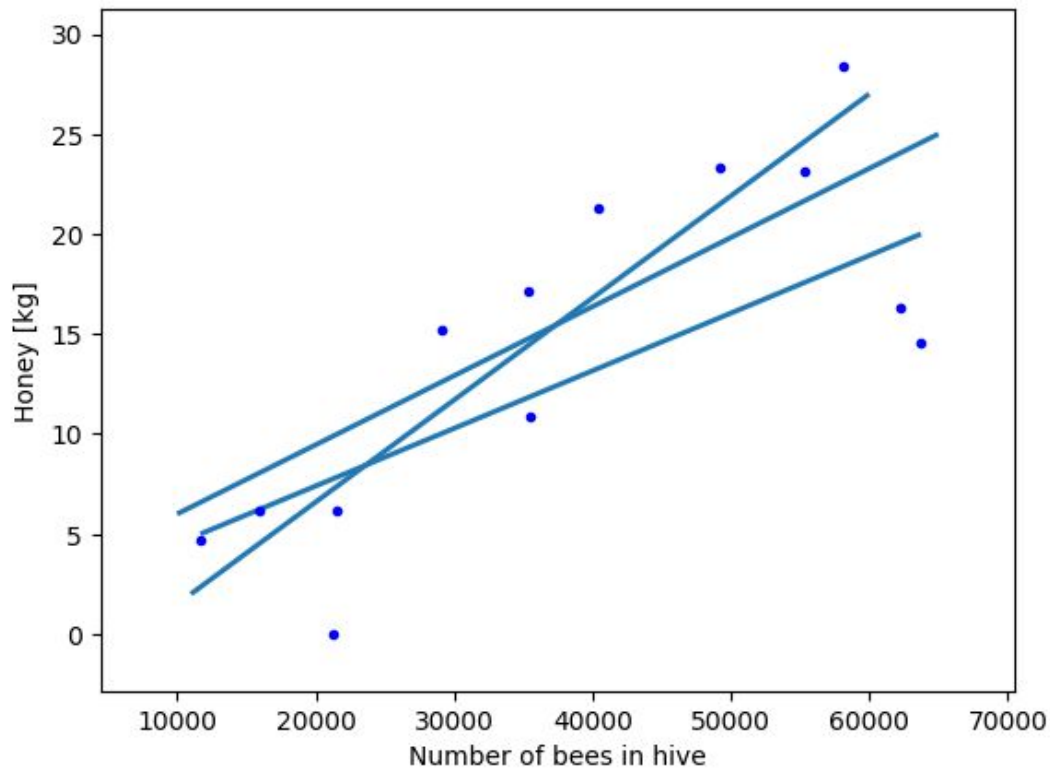
Linearna regresija



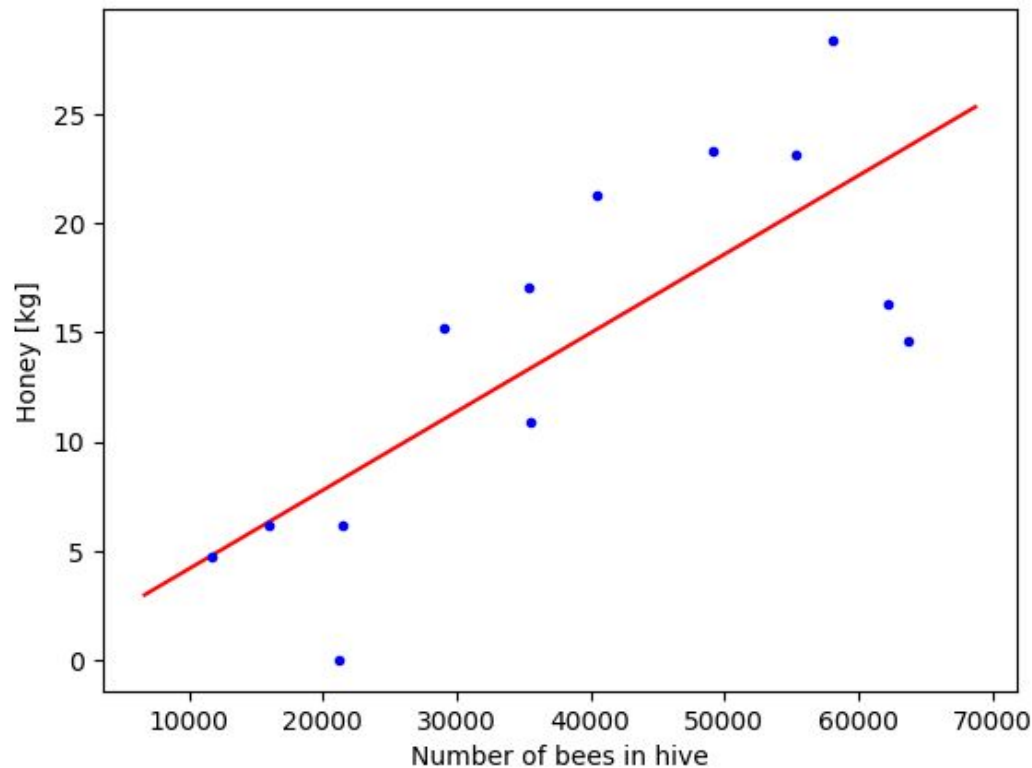
Pridelek medu na panj



Pridelek medu na panj **možni modeli**



Pridelek medu na panj **najboljši model**



Med = f (#čebel, Temperatura, Vlaga)



$$Y = X\beta + \epsilon$$

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

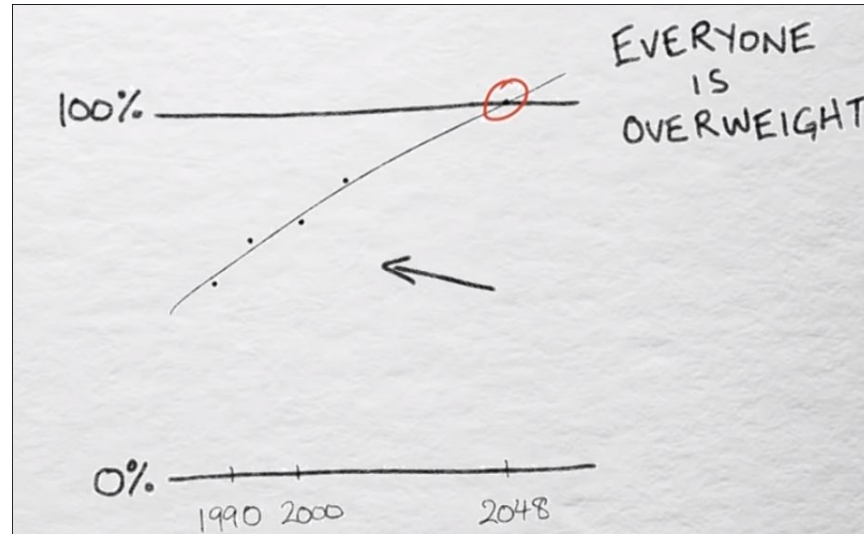
Skaliranje atributov



- Število čebel med 10000 in 60000
- Dnevna temperatura zraka spomladi/poleti med 10°C in 35°C
- Relativna zračna vlaga med 0 in 1 (0-100%)

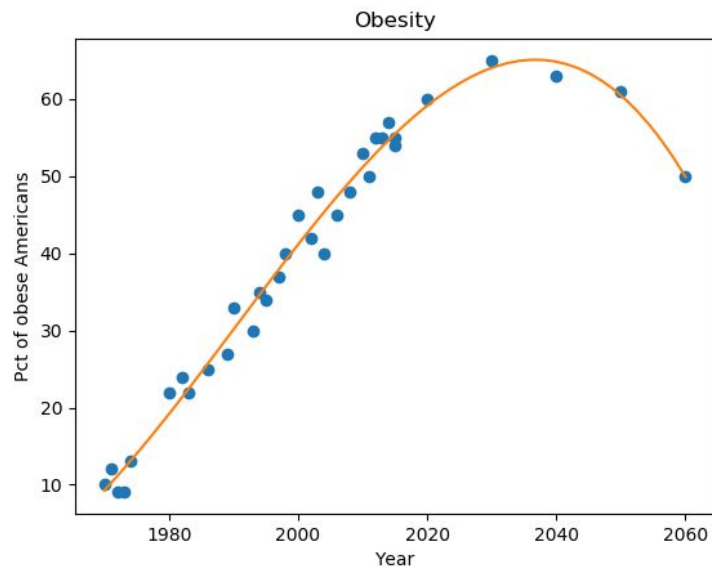
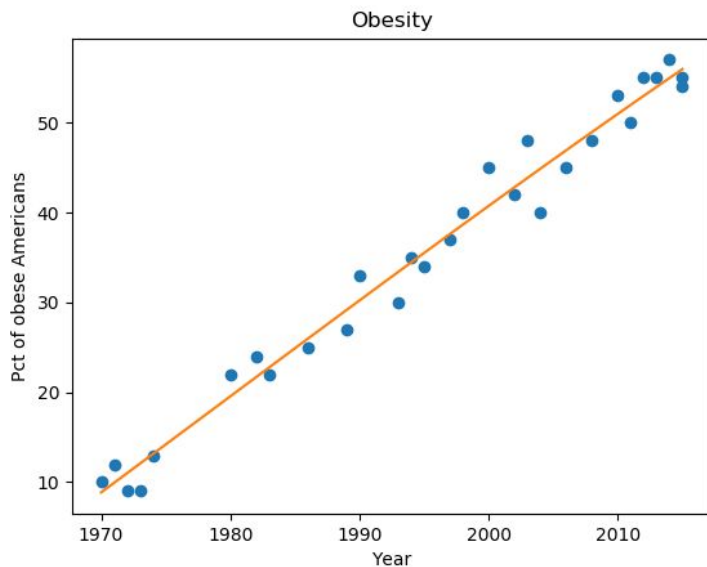
Obesity apocalypse

abcNEWS: "By 2048, all American adults would become overweight or obese."

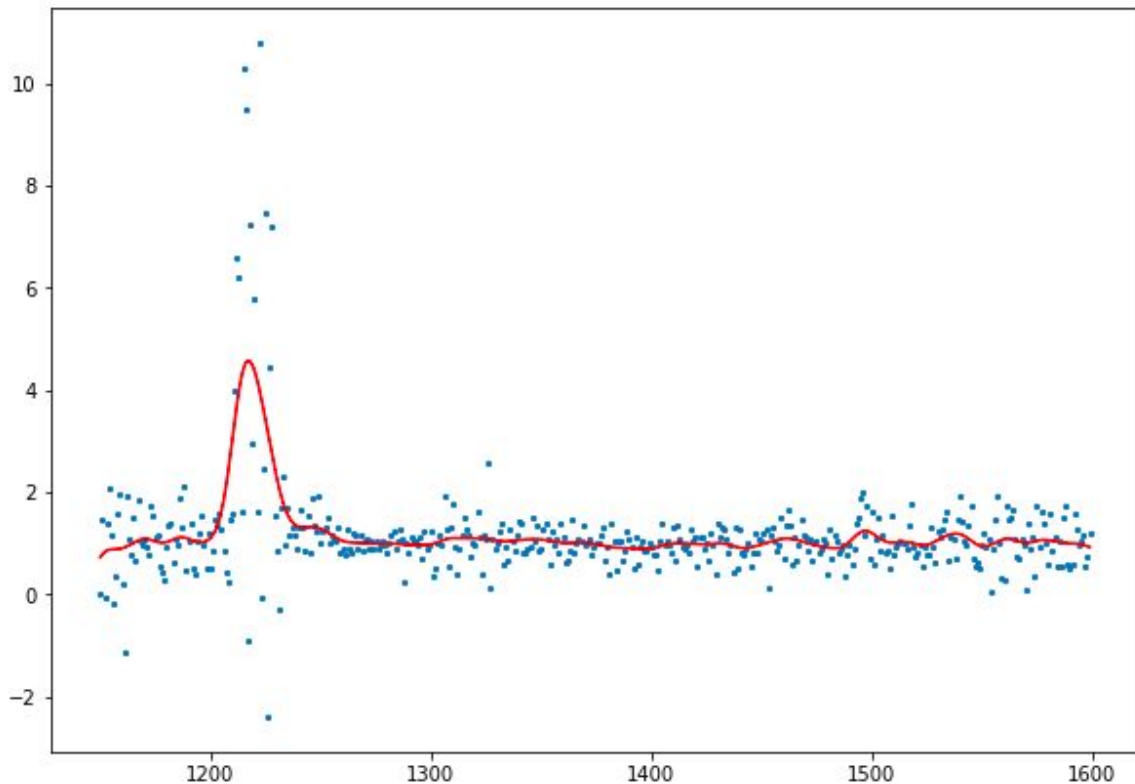
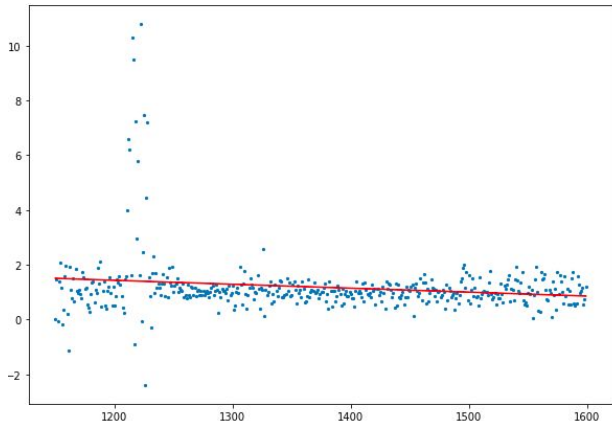


Polinomska regresija

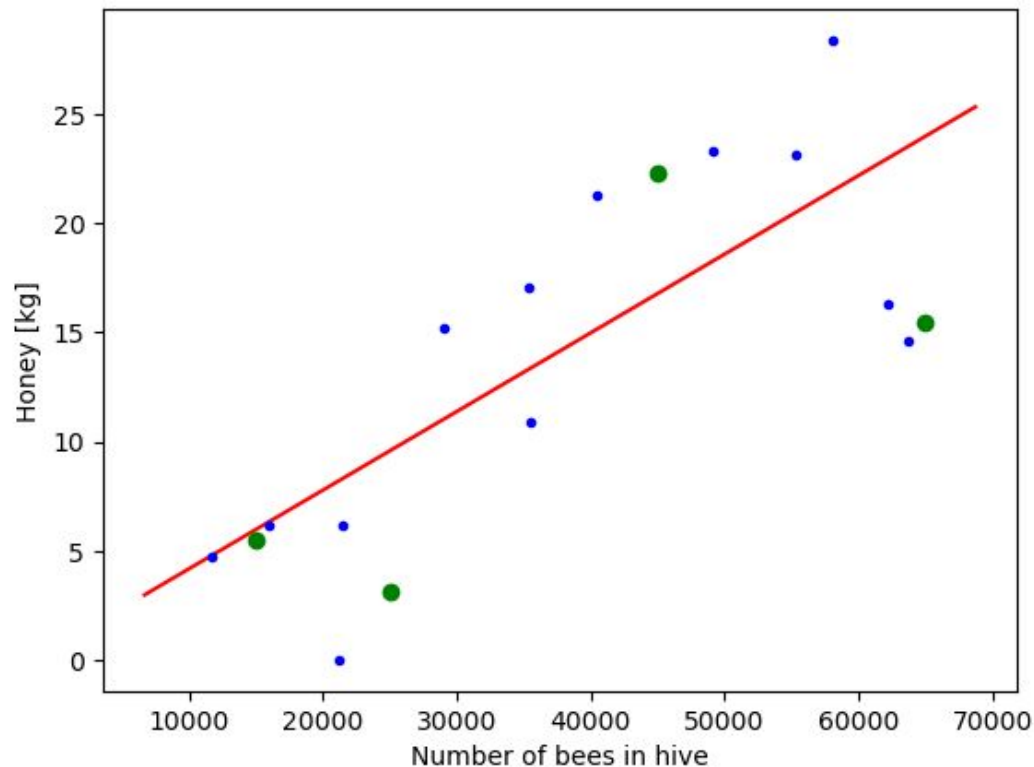
Izmišljeni podatki!



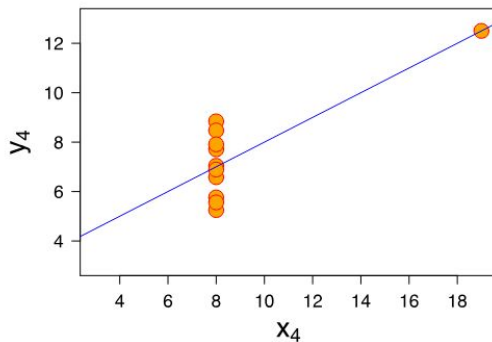
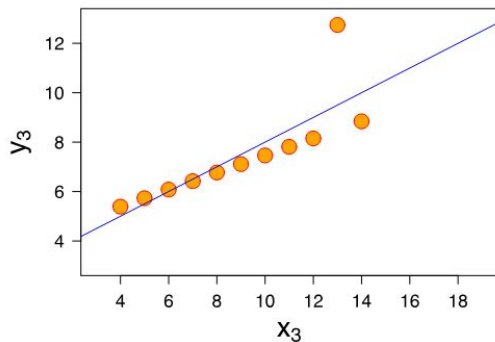
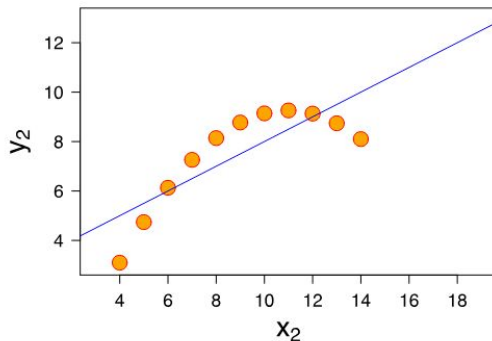
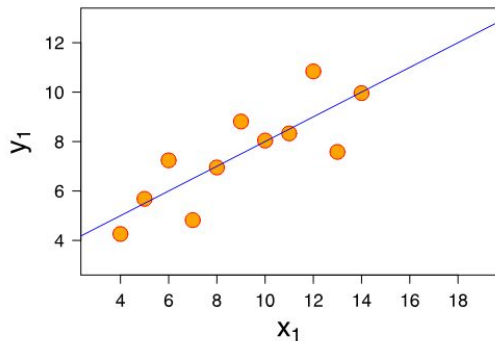
Lokalno utežena regresija (Locally weighted regression)



Med na panj **kNN regresija**



Nariši podatke



4 baze podatkov

skoraj **identična** statistika

zelo **različni** grafi

vir:
https://en.wikipedia.org/wiki/Anscombe%27s_quartet