



Kompetenčni
center za
umetno
inteligenco

Slovenian
National
AI Competence
Centre



Uvod v strojno učenje: Kako iz podatkov nastanejo napovedi

Miran Meža, Marko Meža, Andraž Sivec
Delavnica se izvaja v prostoru LUCAMI (klet nove stavbe)



REPUBLIKA SLOVENIJA
MINISTRSTVO ZA NOTRANJE ZADEVE
IN JAVNO UPRAVO



REPUBLIKA SLOVENIJA
MINISTRSTVO ZA IZOBRAŽEVANJE, ZNANOST
IN MLADINO



REPUBLIKA SLOVENIJA
MINISTRSTVO ZA GOSPODARSTVO
DELO IN ŠPORT



REPUBLIKA SLOVENIJA
MINISTRSTVO ZA OBRAMBO





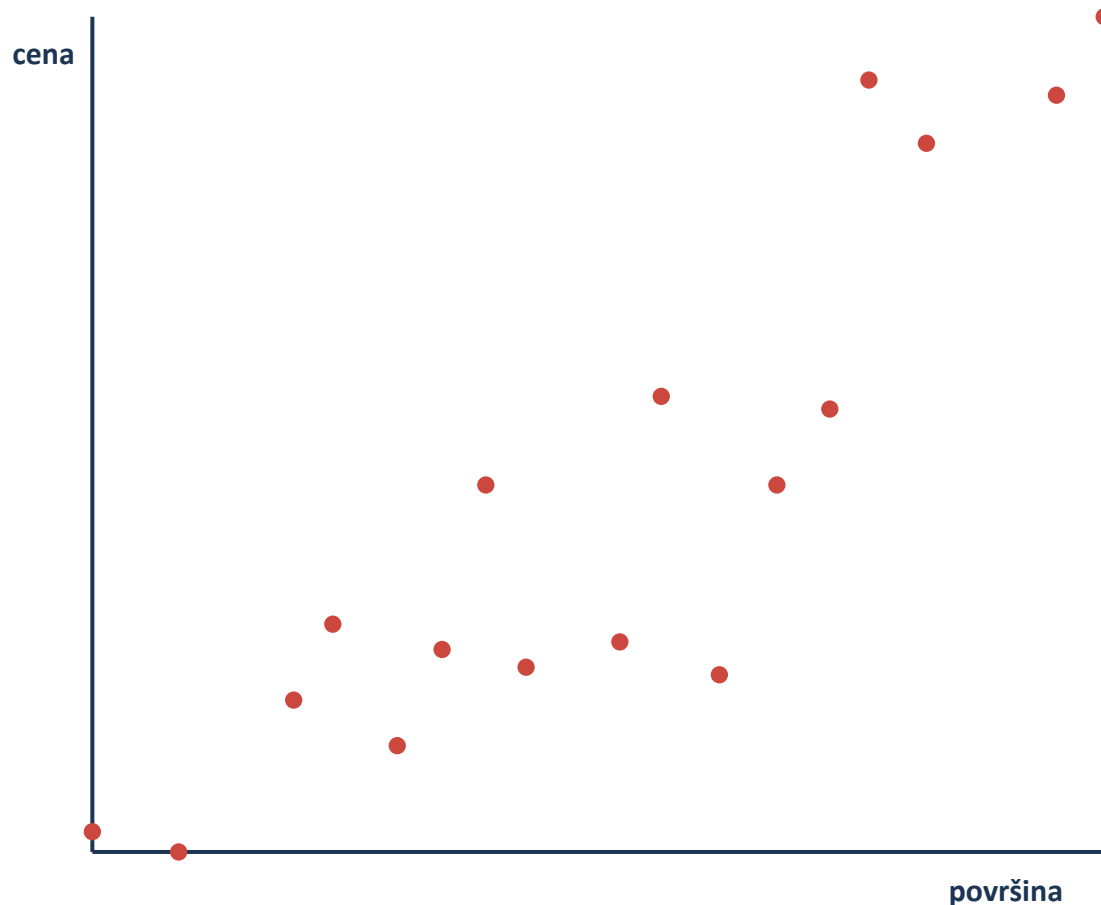
Linearna regresija





Regresijski problem

Napovedujemo številsko vrednost



Primer: cena hiše

Za vsak učni vzorec poznamo pravo vrednost cene.

Cilj

Na podlagi površine napovedati ceno nove hiše.

Regresija = napoved zvezne oziroma številске vrednosti.





Učna množica

Primeri, iz katerih se model uči

površina (x)	cena (y)
2104	460
1416	232
1534	315
852	178
...	...

x = vhodna značilka

y = ciljna spremenljivka

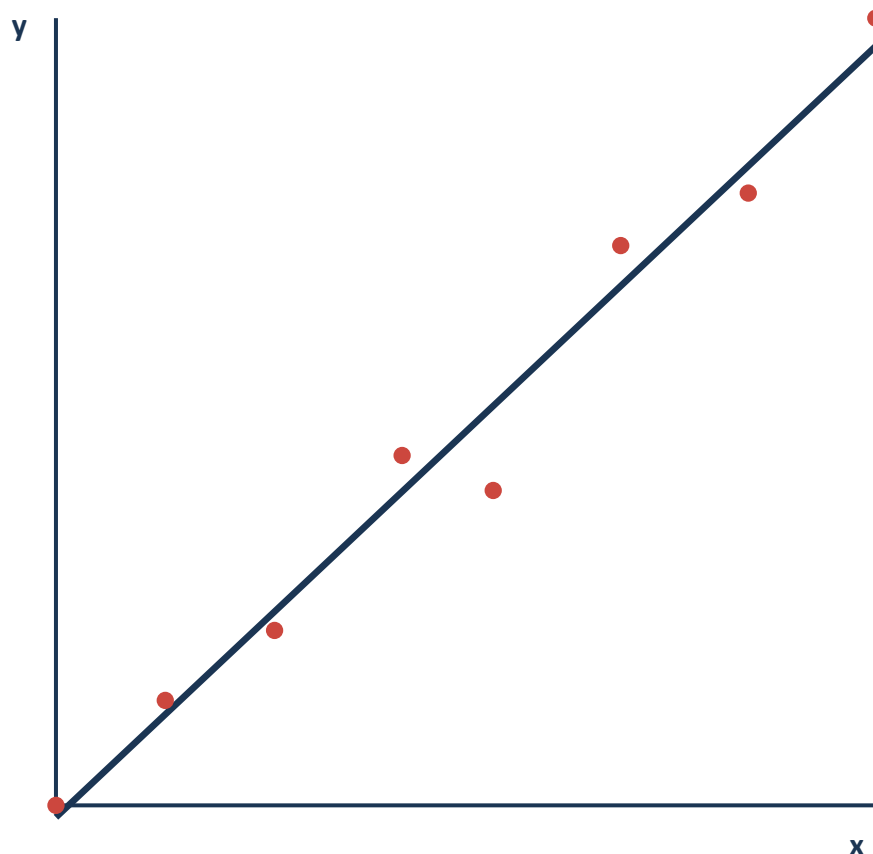
En vzorec poveže vhodne podatke z znanim pravilnim odgovorom.

Iz več takih vzorcev se algoritem nauči modela.



Model linearne regresije

Premica, ki povezuje vhod in napoved



$$\hat{y} = b_0 + b_1x$$

b_0 izhodišče premice

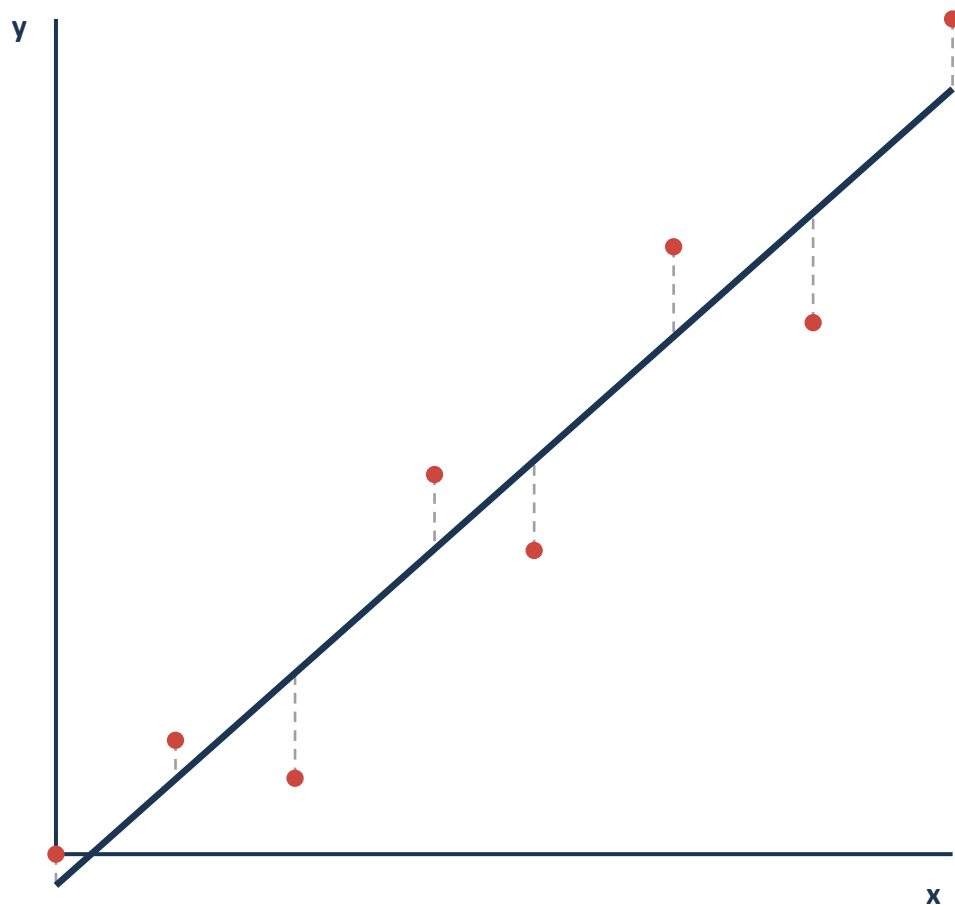
b_1 naklon – koliko se napoved spremeni, ko se x spremeni za 1

Model uporabimo za napoved pri novih vrednostih x .



Kako se model nauči?

Iščemo parametre, pri katerih so napovedi čim bližje podatkom



Napaka

razlika med pravo in
napovedano vrednostjo

Učenje modela

algoritem poišče b_0 in b_1
tako, da je skupna napaka
čim manjša

To je optimizacijski problem.

Gradientni spust je ena od
možnih metod optimizacije.



Od ene do več značilk

Multipla linearna regresija

Ena značilka

$$\hat{y} = b_0 + b_1x$$

npr. površina → cena

Več značilk

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p$$

izolacija

temperatura

št. stanovalcev

povpr. starost

velikost hiše

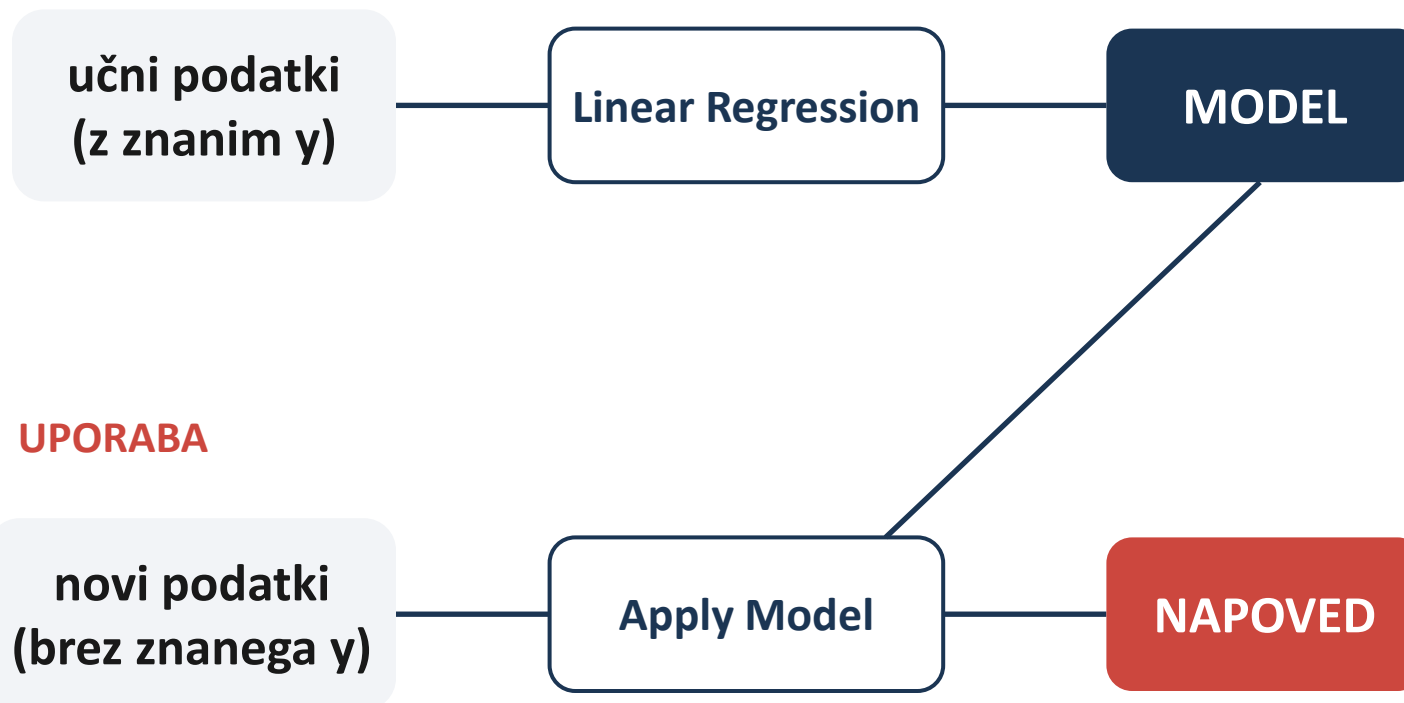
**napoved
porabe
kurilnega
olja**



Učenje in uporaba modela

Učna in testna množica imata različni vlogi

UČENJE



Model se uči samo na učni množici. Testni oziroma novi podatki so namenjeni uporabi in preverjanju modela.



Preverimo območje novih podatkov

Ekstrapolacija je lahko nezanesljiva

Primer: povprečna starost stanovalcev (Avg_Age)



Pri novih podatkih preverimo, ali so značilke podobne tistim, na katerih se je model učil.

V vaji bomo vzorce izven območja učne množice filtrirali.



Vaja: napoved porabe kurilnega olja

Multipla linearna regresija na podatkih strank

Vhodne značilke

Insulation debelina izolacije

Temperature zunanja
temperatura

Num_Occupants št. stanovalcev

Avg_Age povpr. starost

Home_Size velikost hiše

**LINEARNA
REGRESIJA**

**Heating_Oil
napoved
porabe**

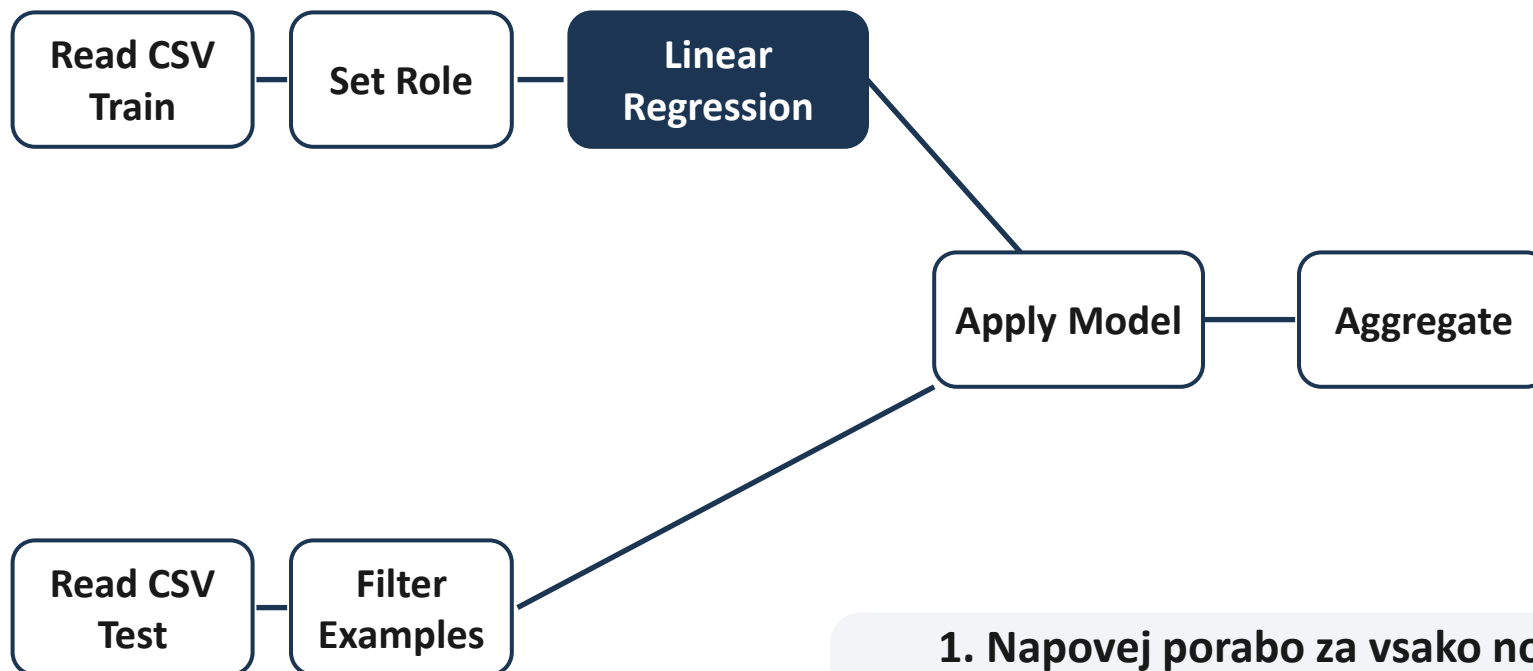
Učna množica: 1.218 obstoječih strank

Nato model uporabimo za nove stranke in
izračunamo skupno ter povprečno
napovedano porabo.



Potek v RapidMinerju

Od podatkov do napovedi



1. Napovej porabo za vsako novo stranko.
2. Izračunaj skupno porabo.
3. Izračunaj povprečno porabo na hišo.