



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





Klasifikacija in Naivni Bayes



UNIVERZA
V LJUBLJANI

FE

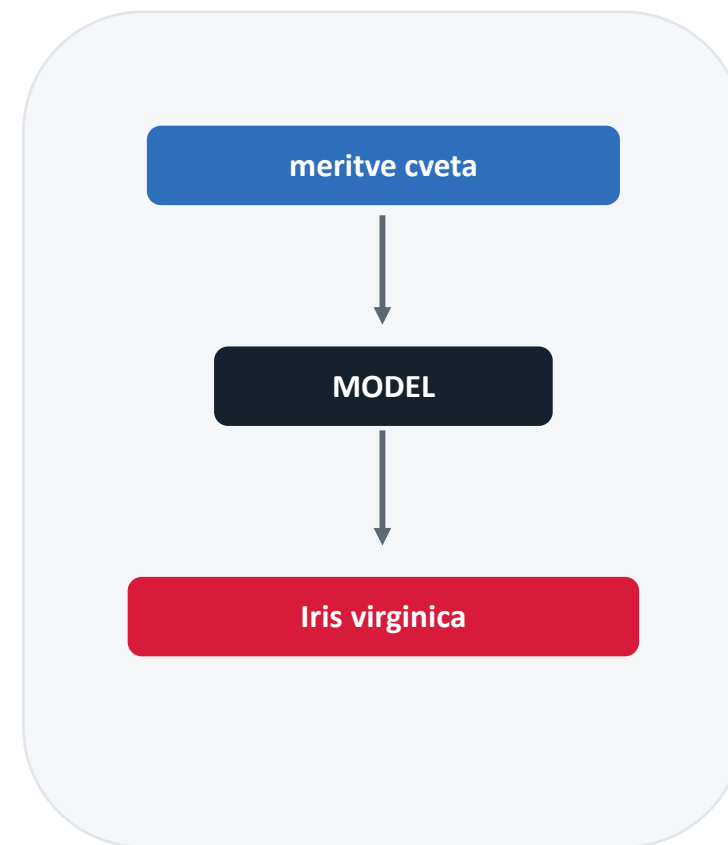
Fakulteta
za elektrotehniko



Klasifikacija in Naivni Bayes

3. sklop • Uvod v strojno učenje

Od značilk do napovedanega razreda





Regresija ali klasifikacija?

V obeh primerih se učimo iz označenih podatkov – razlikuje se tip napovedi.

REGRESIJA

Napovedujemo številsko vrednost.

Primer:

Površina, starost in lokacija hiše

→ napovedana cena: 325.000 €

število

KLASIFIKACIJA

Napovedujemo enega izmed razredov.

Primer:

Meritve cveta

→ Iris setosa / versicolor / virginica

razred





Kako se klasifikator uči?

UČENJE



UPORABA



Pri nadzorovanem učenju med učenjem poznamo pravilen razred.



Primer: podatkovni nabor Iris

150 cvetov, 4 numerične značilke in 3 vrste perunik.

Značilke

sepal length

sepal width

petal length

petal width

MODEL

Razred (species)

setosa

versicolor

virginica

Vprašanje: ali lahko iz štirih meritev pravilno določimo vrsto novega cveta?

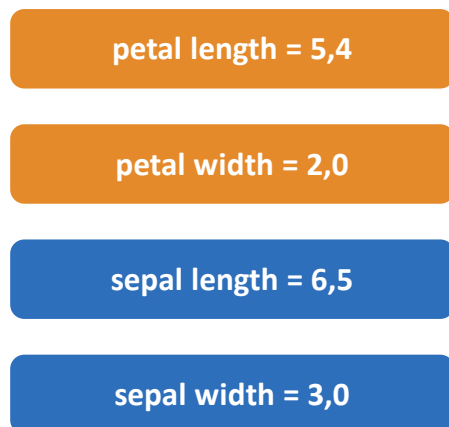




Ideja Naivnega Bayesa

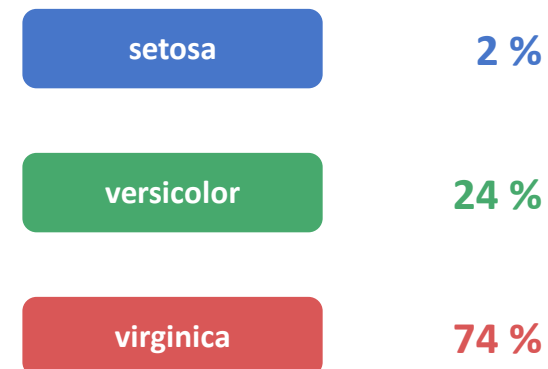
Za vsak razred ocenimo, kako verjetne so izmerjene značilke.

Nov cvet



Naivni Bayes

ocena verjetnosti



Napoved: Iris virginica

Model izbere razred z največjo ocenjeno verjetnostjo.



Zakaj »naivni« Bayes?

Poenostavljena predpostavka: če poznamo razred, obravnavamo značilke kot medsebojno neodvisne.



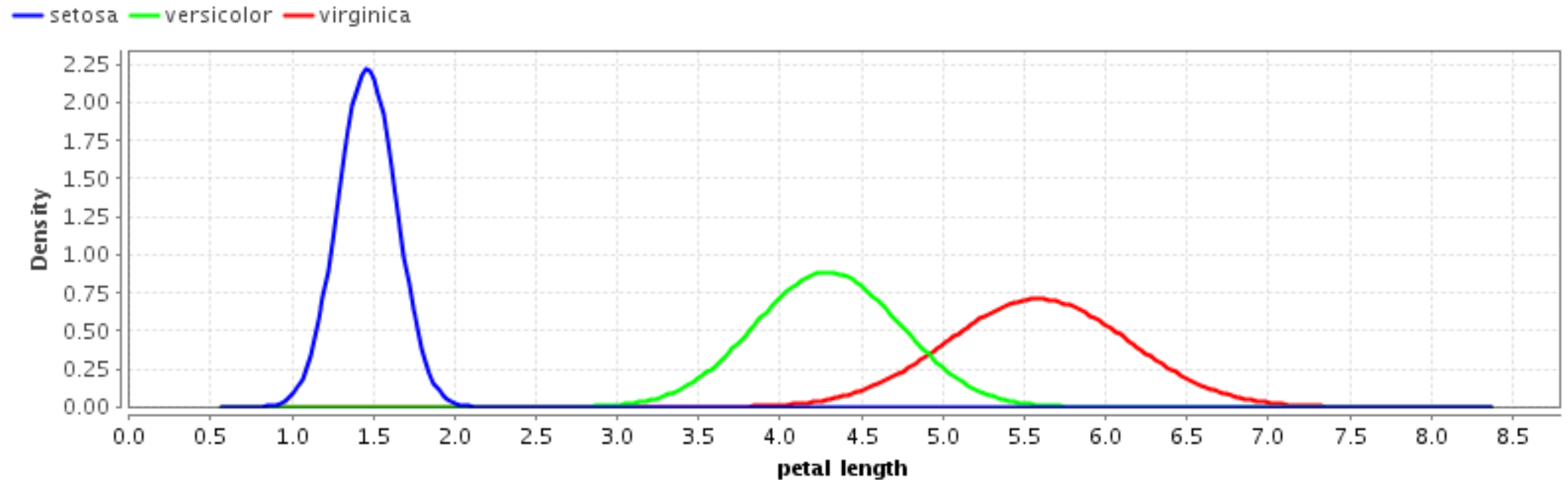
V resničnih podatkih značilke pogosto niso povsem neodvisne.

Kljub poenostavitvi je metoda hitra, pregledna in pogosto presenetljivo učinkovita.



Značilke pomagajo ločiti razrede

Primer: porazdelitev dolžine venčnega lista (petal length).

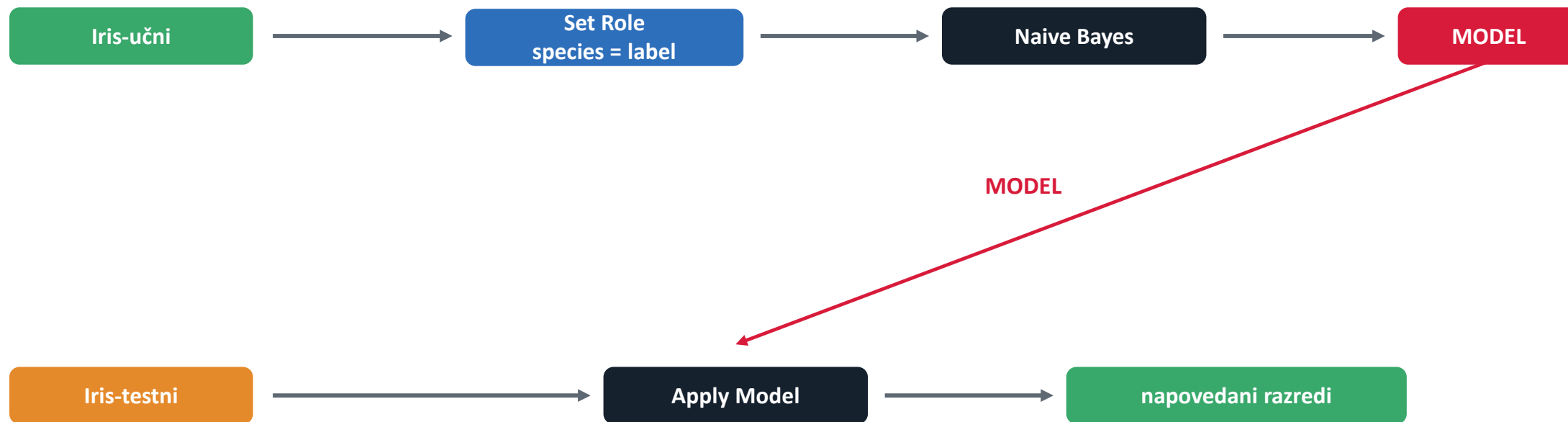


Bolj kot se porazdelitve razredov razlikujejo, lažje je značilko uporabiti za klasifikacijo.



Učna in testna množica

Model učimo na znanih primerih in nato preverimo napovedi na novih primerih.





Praktična vaja: Naivni Bayes v RapidMinerju

- 1 Naložimo učne in testne podatke Iris.
- 2 Pregledamo značilke in raztresne grafe.
- 3 species nastavimo kot label.
- 4 Zgradimo model z operatorjem Naive Bayes.
- 5 Model uporabimo na testnih podatkih.
- 6 Napovedi primerjamo s kontrolnimi vrednostmi.

Cilj: razumeti pot od podatkov do napovedanega razreda.