

# Klasifikacija 3

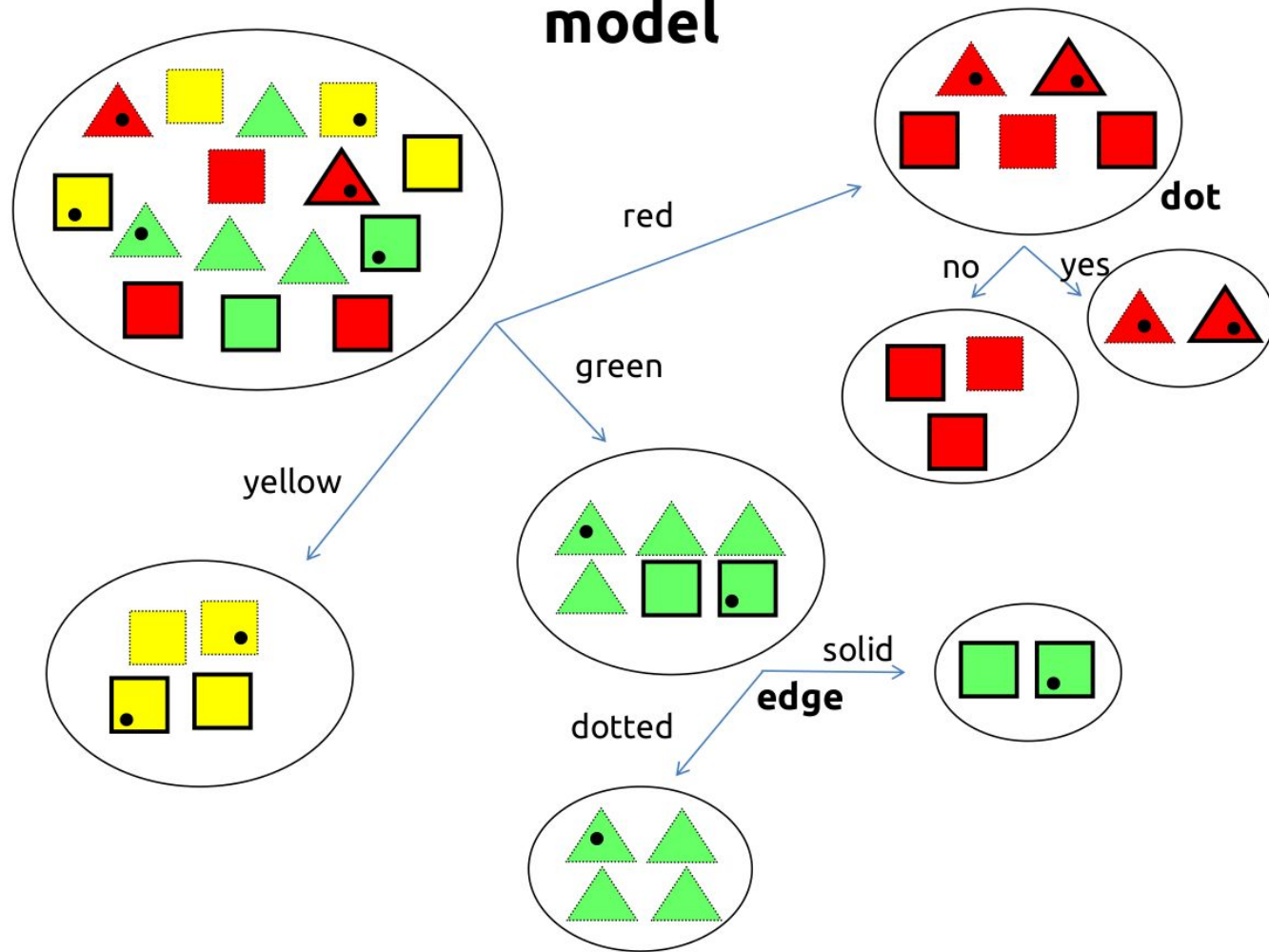
21.5.2020

Uvod v strojno učenje  
jure.zabkar@fri.uni-lj.si

# Vsebina

- Ocenjevanje točnosti in rezanje
- Vaje
- Naivni Bayesov klasifikator
- Vaje
- LDA, QDA
- Prečno preverjanje

# model



# Ocenjevanje verjetnosti

Točnost  $T$  = verjetnost pravilne klasifikacije.

Napaka =  $1 - T$

$N$  ... število vseh primerov,  $n$  ... število uspešnih poskusov

- relativna frekvenca:  $p = n/N$

- m-ocena:  $p = (n + p_a * m) / (N + m)$

ekspert zaupa v  $p_a \Rightarrow$  velik  $m$ , sicer majhen  $m$  (tipično  $m=2$ )

- Laplace:  $p = (n+1)/(N+k)$

# Rezanje dreves

- Nepopolni podatki, (merske) napake v podatkih
- Učenje šuma, namesto učenja dejanske funkcije, ki generira podatke
- Slaba razumljivost dreves
- pretirano prilagajanje => nižja klasifikacijska točnost na testnih podatkih

# Rezanje naprej

- omejevanje št. primerov v vozlišču
- ustavljanje gradnje pri doseženi želeni točnosti v vozlišču

# Rezanje nazaj

Postopek MEP (Minimal Error Pruning)

Cilj: poreži drevo tako, da bo ocenjena klasifikacijska točnost maksimalna

Za vsako vozlišče  $v$  izračunamo:

- statično napako (napaka v vozlišču, če odrežemo oba naslednika)
- vzvratno napako (napaka v vozlišču, če ohranimo njegova naslednika)

Režemo pod  $v$ , če je statična napaka manjša od vzvratne.

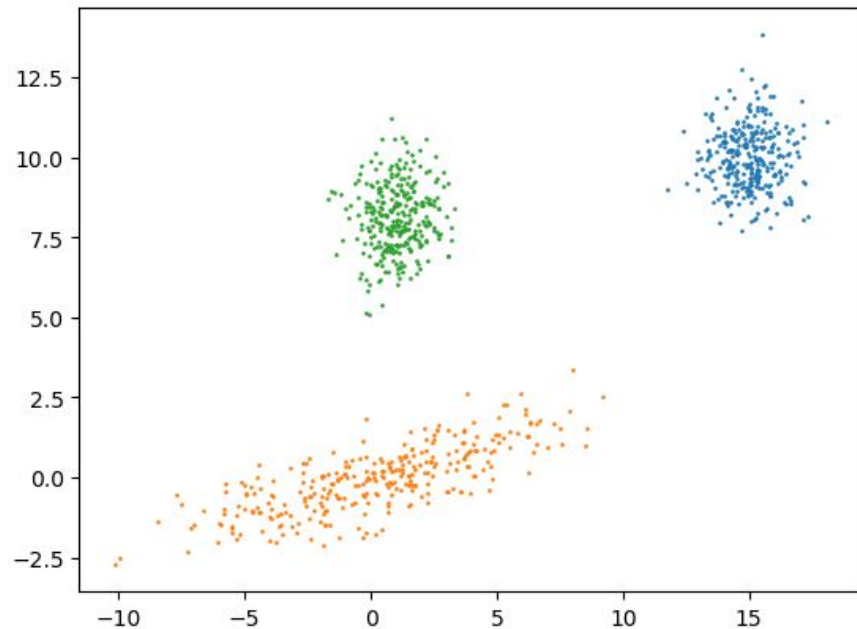
# Naivni Bayesov klasifikator

- Verjetnosti klasifikator,
- Naivnost: v predpostavki, da so atributi med seboj pogojno neodvisni,
- Algoritem se v praksi pogosto izkaže kot dober,
- Priljubljen zaradi svoje preprostosti in hitrosti,
- Vizualizacija z nomogrami



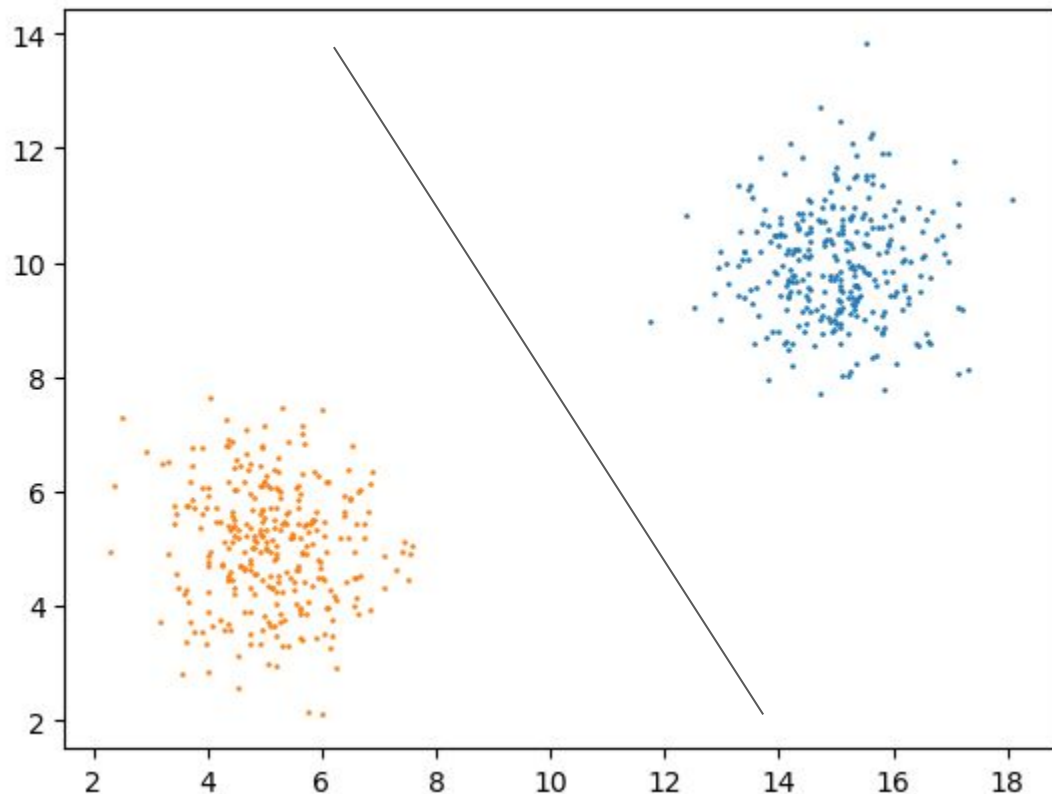
# Diskriminantna analiza

- Alternativa logistični regresiji
- modeliramo porazdelitve posameznih razredov in uporabimo Bayesov izrek za izračun  $\Pr(Y = k|X = x)$
- Log. reg. ima probleme pri majhnem vzorcu in ko so razredi precej ločeni.



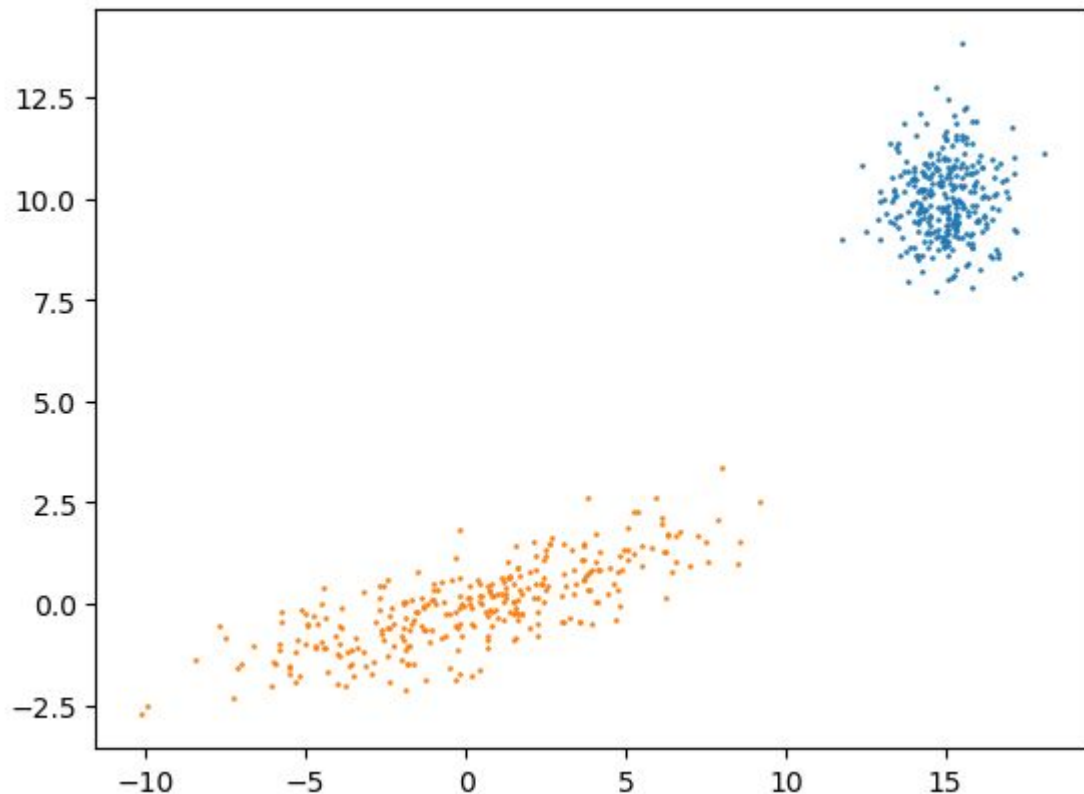
# LDA (linearna diskriminantna analiza)

Predpostavlja, da so učni primeri posameznega razreda iz normalne porazdelitve s svojim povprečjem in isto kovariančno matriko (za vse razrede).

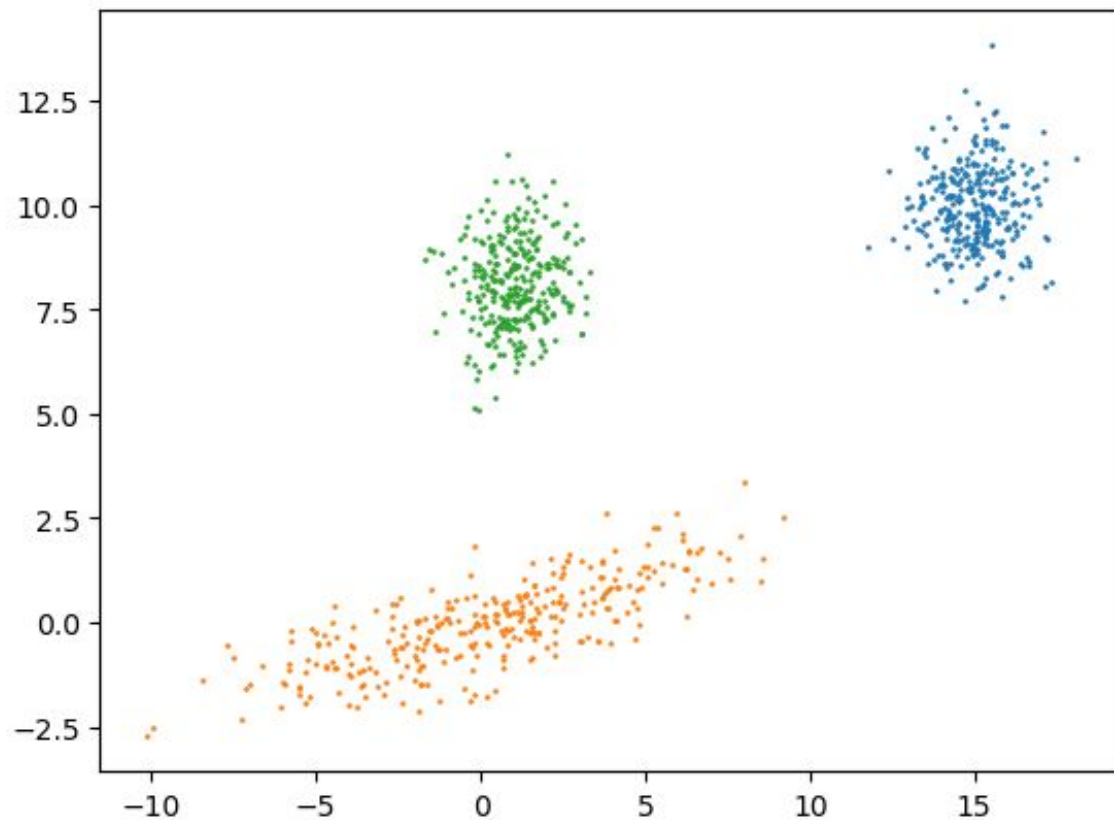


# QDA (Kvadratna diskriminantna analiza)

Podobna predpostavka kot pri LDA; razlika: vsak razred ima svojo kovariančno matriko.



# Diskriminantna analiza



# Prečno preverjanje (cross-validation)

