

## 1. domača naloga

Namen te domače naloge je raziskovati linearno regresijo v nadzorovanem okolju. S tem namenom definirajte (ne)linearne funkcije več spremenljivk, s pomočjo katerih generirajte učne in testne podatke za strojno učenje. Vaša analiza naj obsega vsaj: ocenjevanje koeficientov linearnega modela, računanje intervalov zaupanja, skaliranje in standardizacijo vrednosti atributov. Eksperimentirajte z različno veliki vzorci učnih podatkov, opazujte konvergenco ocen koeficientov, vizualizirajte rezultate.

Ocenjuje se obširnost in poglobljenost vaše analize (ne dolžina poročila! Ena tabela lahko prikaže zelo obširne rezultate.); za pozitiven izkupiček točk,  $\geq 8$  (od 15 možnih), je potrebno izbrati vsaj eno linearno funkcijo z  $> 5$  spremenljivkami na različnih skalah, generirati podatke z vsaj 5 stopnjami šuma in za vsako od teh množic podatkov primerjati razliko med modeli brez skaliranja, s skaliranjem in standardizacijo.

Svojo študijo regresijskih algoritmov lahko poglobite z:

- analizo več linearnih funkcij z različnim številom spremenljivk (kako na učenje vpliva število dimenzij),
- analizo nelinearnih funkcij; uporabite lokalno uteženo regresijo, polinomsko regresijo, regresijo na  $k$  najbližjih sosedih (z različnimi metrikami)
- analizo Ridge in Lasso regresije, oz. primerjavo z drugimi regresijskimi algoritmi po lastni izbiri.

Poročilo (v formatu pdf) oddajte v spletni učilnici:  
<https://e.fe.uni-lj.si/mod/assign/view.php?id=6273>  
do 12.5.2020, 17:00.