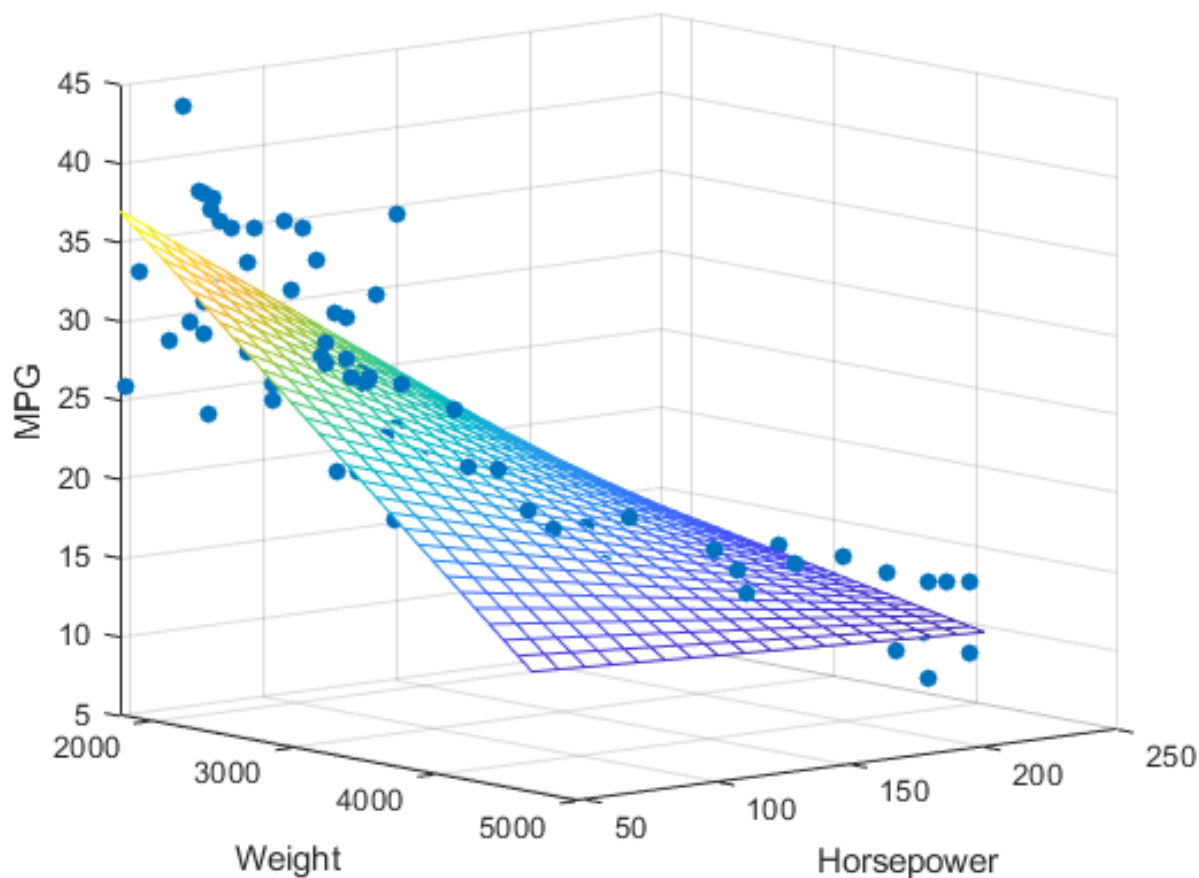


# ИЗВЕШТАЈ ЗАДАТКА 2

*Имплементација модела вишеструке регресије*



**Душан Бркић, Филип Живанац**

30. април 2022.

Софтверско инжењерство и информационе технологије

Факултет техничких наука

Универзитет у Новом Саду

## ЗАДАТАК

Предиктовати плату (колона плата (долари)) наставног особља у САД на основу више атрибута (звање, област, година\_доктор, година\_искуства, пол). Задатак се сматра успешно урађеним уколико се на комплетном тестном скупу података добије RMSE (Root mean squared error) мањи од 28500. Потребно је наћи најбољи модел вишеструке регресије уз помоћ алгоритама регресије и одабира обележја.

## ПРИСТУП ПРОБЛЕМУ

За решење проблема најпре смо имплементирали *Ridge* регресију и на добијених резултата вршили одабир обележја уз помоћу кога смо уклонили обележје “пол” јер смо установили да оно не доприноси квалитету решења. Остале манипулације обележјима нису помогле у побољшању датог решења. Затим смо пробали регресију вишег степена и дошли до закључка да је најбоља она са првим степеном. Пошто обична *Ridge* регресија није давала задовољавајућа решења, употребили смо кернелизовану *Ridge* регресију.

## ИСПРОБАНИ АЛГОРИТМИ

### RIDGE (L2 Regularization)

Ова регресија је уснови регресија најмањих квадрата с тим да има у себи и параметар  $\alpha$  (алфа). Уколико  $\alpha=0$  овај модел има исто решење као и регресија најмањих квадрата. Повећавањем  $\alpha$  постижемо то да се наш модел слабије уклапа у тренинг скуп података али се зато смањује варијанса са различитим тестним скуповима података што нам за овај проблем одговара јер нам тестни скуп није потпуно познат.

### KRR (Kernel Ridge Regression - L2 Regularization), са коришћењем параметризованог Гаусовог кернела

Ова регресија је уснови *Ridge* регресија која је кернелизована. Кернелизована регресија за параметре узима функцију кернела и ламбда параметар, који служи као *noise* параметар, чијим повећавањем повећавамо робусност модела, као код *Ridge* регресије

За кернел је изабран параметризован Гаусов кернел, који користи параметре *sigma* и *length*. Када су ови параметри постављени на “1”, добија се Гаусов кернел. *Sigma* је пропорционална коначној вредности кернела, а *length* обрнуто пропорционалан. Манипулацијом ових параметара можемо такође да утичемо на робусност кернела. Повећавањем *sigma*, а смањивањем *length* параметра

повећавамо робусност решења, и обрнуто.

## РЕЗУЛТАТИ

KRR (Kernel Ridge Regression - L2 Regularization), са коришћењем параметризованог Гаусовог кернала

Ламбда	Сигма	Length	RMSE
0.01	1	1	79142.9849615965
0.01	1	5	24803.435801745407
0.01	5	1	78795.82853649929
0.01	5	5	30323.191981805507
0.01	5	10	<b>15957.632028958116</b>
0.01	5	20	17372.388871926116
0.01	10	10	16164.26237898722
0.01	10	5	33974.846285795466
0.01	10	20	17879.20595590022

## ОДАБРАНО РЕШЕЊЕ

Као одабрано решење одлучили смо се за кернелизовану *Ridge* регресију параметара вредности Ламбда=0.01 Сигма=5 и Length=10.