

Regression Variables Selector

This document shows a step-by-step procedure of how the variables were selected for modeling

Loading the packages

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --

## v ggplot2 3.3.0      v purrr  0.3.4
## v tibble  3.0.1      v dplyr  0.8.5
## v tidyr   1.0.3      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.5.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lasso2)
```

```
## R Package to solve regression problems while imposing
##   an L1 constraint on the parameters. Based on S-plus Release 2.1
## Copyright (C) 1998, 1999
## Justin Lokhorst <jlokhors@stats.adelaide.edu.au>
## Berwin A. Turlach <bturlach@stats.adelaide.edu.au>
## Bill Venables <wvenable@stats.adelaide.edu.au>
##
## Copyright (C) 2002
## Martin Maechler <maechler@stat.math.ethz.ch>
```

Loading the datasets

```
s17_1 <- read.csv("~/DSI-SRP1/season2017.csv", encoding = "UTF-8")
s17_2 <- read.csv("~/DSI-SRP1/FPL_2016_17_new.csv")
s18_1 <- read.csv("~/DSI-SRP1/season2018.csv", encoding = "UTF-8")
s18_2 <- read.csv("~/DSI-SRP1/FPL_2017_18_new.csv")
s19_1 <- read.csv("~/DSI-SRP1/season2019.csv", encoding = "UTF-8")
s19_2 <- read.csv("~/DSI-SRP1/FPL_2018_19_new.csv")
```

Variable selector

```
var_sel <- function(df) {  
  df %>%  
    select(goals_scored,assists,total_points,minutes.played,bonus,  
           bps,goals_conceded,clean_sheets,ict_index,position_index,  
           selected_by_percent)  
}
```

Part I

This is done using the seasonal datasets that is going to be used for modeling

```
s17_1n <- var_sel(s17_1)  
s18_1n <- var_sel(s18_1)  
s19_1n <- var_sel(s19_1)  
s_combined <- rbind(s17_1n, s18_1n, s19_1n)  
lasso.s <- l1ce(total_points ~., data = s_combined)  
summary(lasso.s)$coefficients
```

##		Value	Std. Error	Z score	Pr(> Z)
##	(Intercept)	16.46621655	1.438894135	11.443661	0.000000e+00
##	goals_scored	0.00000000	0.313436323	0.000000	1.000000e+00
##	assists	0.00000000	0.325963467	0.000000	1.000000e+00
##	minutes.played	0.00000000	0.003115412	0.000000	1.000000e+00
##	bonus	0.65961380	0.187782364	3.512650	4.436607e-04
##	bps	0.08088562	0.010440822	7.747055	9.325873e-15
##	goals_conceded	0.00000000	0.104186649	0.000000	1.000000e+00
##	clean_sheets	0.67163353	0.346257177	1.939696	5.241669e-02
##	ict_index	0.16194721	0.024962065	6.487733	8.713763e-11
##	position_index	0.00000000	0.445673259	0.000000	1.000000e+00
##	selected_by_percent	0.00000000	0.139350609	0.000000	1.000000e+00

```
lasso.s <- l1ce(total_points ~ ict_index + bps + clean_sheets + bonus,  
               data = s_combined)  
summary(lasso.s)$coefficients
```

##		Value	Std. Error	Z score	Pr(> Z)
##	(Intercept)	25.47307218	0.814252771	31.283986	0.000000e+00
##	ict_index	0.12862290	0.017572139	7.319707	2.484679e-13
##	bps	0.07502564	0.007530691	9.962651	0.000000e+00
##	clean_sheets	0.00000000	0.334994853	0.000000	1.000000e+00
##	bonus	0.03378626	0.204580419	0.165149	8.688267e-01

Part II

This is done by combining all three weekly datasets

```
s_week <- rbind(s17_2, s18_2, s19_2)
s_week <- s_week %>%
  select(-player_name:-position, -X, -season, -GW)
sample.index <- sample(1:nrow(s_week), nrow(s_week)*0.85, replace = FALSE)
s_week1 <- s_week[-sample.index,]
lasso.week <- l1ce(total_points ~., data = s_week1)
summary(lasso.week)$coefficients
```

```
##              Value Std. Error  Z score  Pr(>|Z|)
## (Intercept)  13.00911259 0.624490792 20.8315523 0.000000e+00
## position_index 0.00000000 0.184557211 0.0000000 1.000000e+00
## goals_scored  0.00000000 0.152634678 0.0000000 1.000000e+00
## assists       0.00000000 0.155038199 0.0000000 1.000000e+00
## ict_index      0.18247190 0.012564276 14.5230734 0.000000e+00
## goals_conceded 0.00000000 0.049189494 0.0000000 1.000000e+00
## minutes       0.00000000 0.001514983 0.0000000 1.000000e+00
## own_goals     0.00000000 0.770282447 0.0000000 1.000000e+00
## bps           0.08757608 0.005396739 16.2275921 0.000000e+00
## bonus         0.52957001 0.090157267 5.8738472 4.257958e-09
## clean_sheets  0.15558794 0.160064255 0.9720342 3.310335e-01
```

```
lasso.week <- l1ce(total_points ~ ict_index + bps + bonus + clean_sheets,
  data = s_week1)
summary(lasso.week)$coefficients
```

```
##              Value Std. Error  Z score Pr(>|Z|)
## (Intercept)  18.84367749 0.298272063 63.17614      0
## ict_index     0.14721372 0.008969937 16.41190      0
## bps           0.07436583 0.003702662 20.08442      0
## bonus         0.00000000 0.095315680 0.00000      1
## clean_sheets  0.00000000 0.150588470 0.00000      1
```

Part III

This is done using the cumulative data over the three seasons

```
scom <- read.csv("~/DSI-SRP1/FPL_ULTIMATE.csv")
scom <- scom %>%
  select(-player_name:-position, -season, -bonus, - bps, -total_points, -GW, -X)
sample_s <- sample(1:nrow(scom), nrow(scom)*0.9, replace = FALSE)
scom1 <- scom[-sample_s,]
model <- l1ce(ict_index ~ ., data = scom1)
summary(model)$coefficients
```

```
##              Value Std. Error  Z score  Pr(>|Z|)
## (Intercept)  50.85753103 3.057375797 16.634374 0.00000000
## position_index 0.00000000 0.918202600 0.0000000 1.00000000
## goals_scored   3.73671607 0.268423683 13.920963 0.00000000
## assists        6.17467998 0.316738261 19.494582 0.00000000
```

```
## goals_conceded  0.00000000 0.182813614  0.000000  1.00000000
## minutes         0.01247876 0.004722484  2.642415  0.00823172
## own_goals       0.00000000 2.801216373  0.000000  1.00000000
## clean_sheets    0.77773186 0.635279071  1.224237  0.22086301
```

```
model <- l1ce(ict_index ~ goals_scored + assists + minutes + clean_sheets,
              data = scom1)
summary(model)$coefficients
```

```
##               Value Std. Error  Z score  Pr(>|Z|)
## (Intercept)  52.47947785 1.178142166 44.544266 0.00000000
## goals_scored   3.58463217 0.250823019 14.291480 0.00000000
## assists        6.05238126 0.319274174 18.956689 0.00000000
## minutes        0.01203361 0.001250045  9.626542 0.00000000
## clean_sheets   0.76210499 0.338800373  2.249422 0.02448566
```