

METIS

Predicting points in soccer leagues

Ming Tang

10/06/2017

Background

- Spanish Soccer League (La Liga)

– I am a Real Madrid fan ~~~



- Questions

- 1) **correlate** the points by the linear regression?
- 2) **predict** the points by the model?

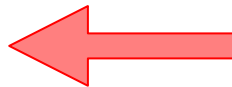


Look at the data



	Season	Rank	Team	Points	Goal_for	Goal_against	Goal_diff	Goals	Assist	Shots_On_Goal	...	Red	Fouls	...
3	2014-2015	1	Barcelona	94	110	21	89	110	80	291	...	3	369	120
74	2016-2017	1	Real Madrid	93	106	41	65	106	81	282	...	2	422	136
22	2013-2014	1	Atletico Madrid	90	77	26	51	77	56	212	...	2	523	130
42	2015-2016	1	Barcelona	91	112	29	83	112	79	277	...	1	385	120
23	2013-2014	2	Barcelona	87	100	33	67	100	72	290	...	1	425	110

1 target:
Points



9 features:
Goals scored, Goals lost, Assists...

Intro

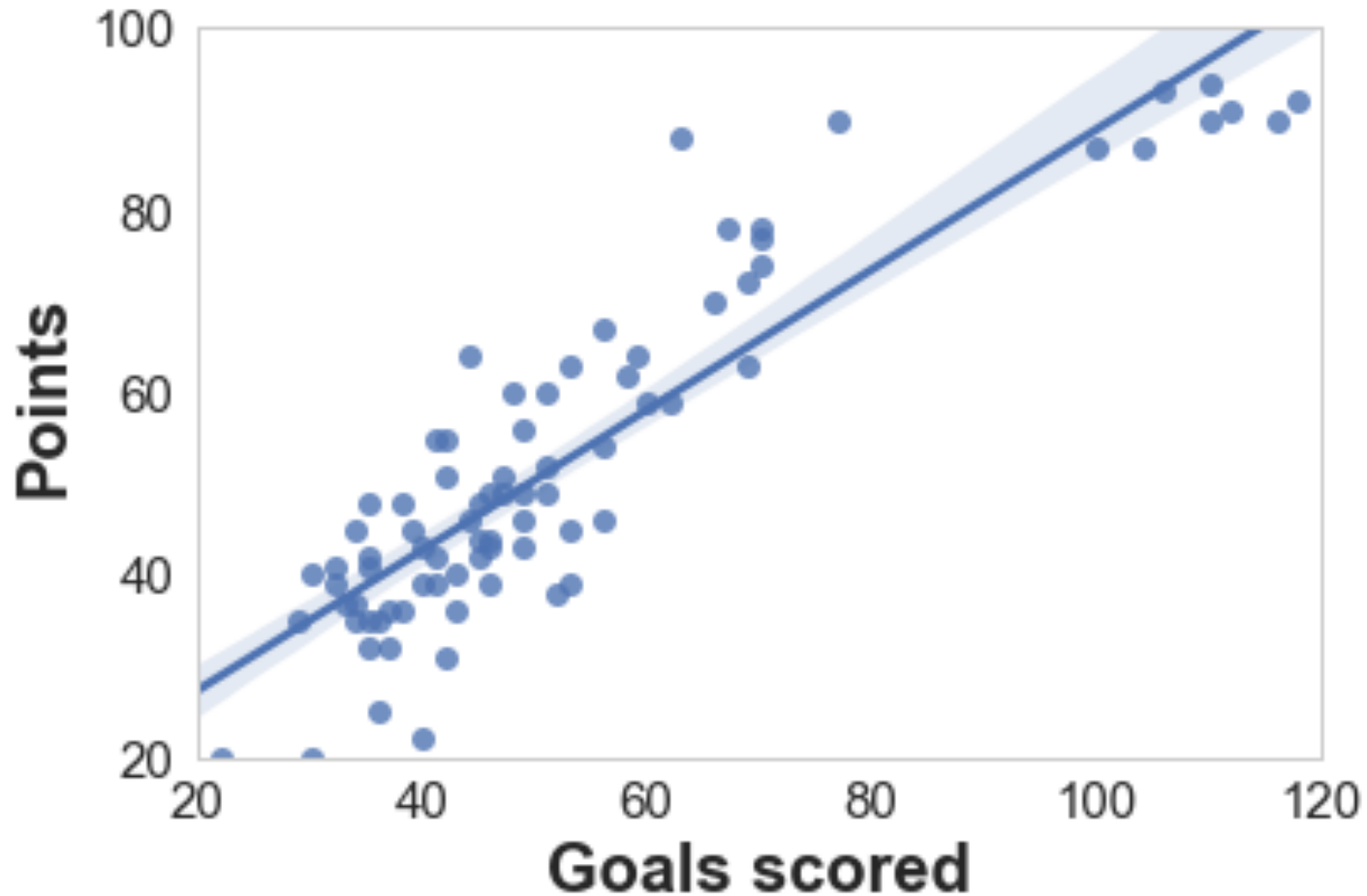
- 20 teams
- 38 games for each team
- 3 - win, 1 - draw, 0 - loss

Data

- 4 seasons (2013-2017)
- *2013-2014 – test data*

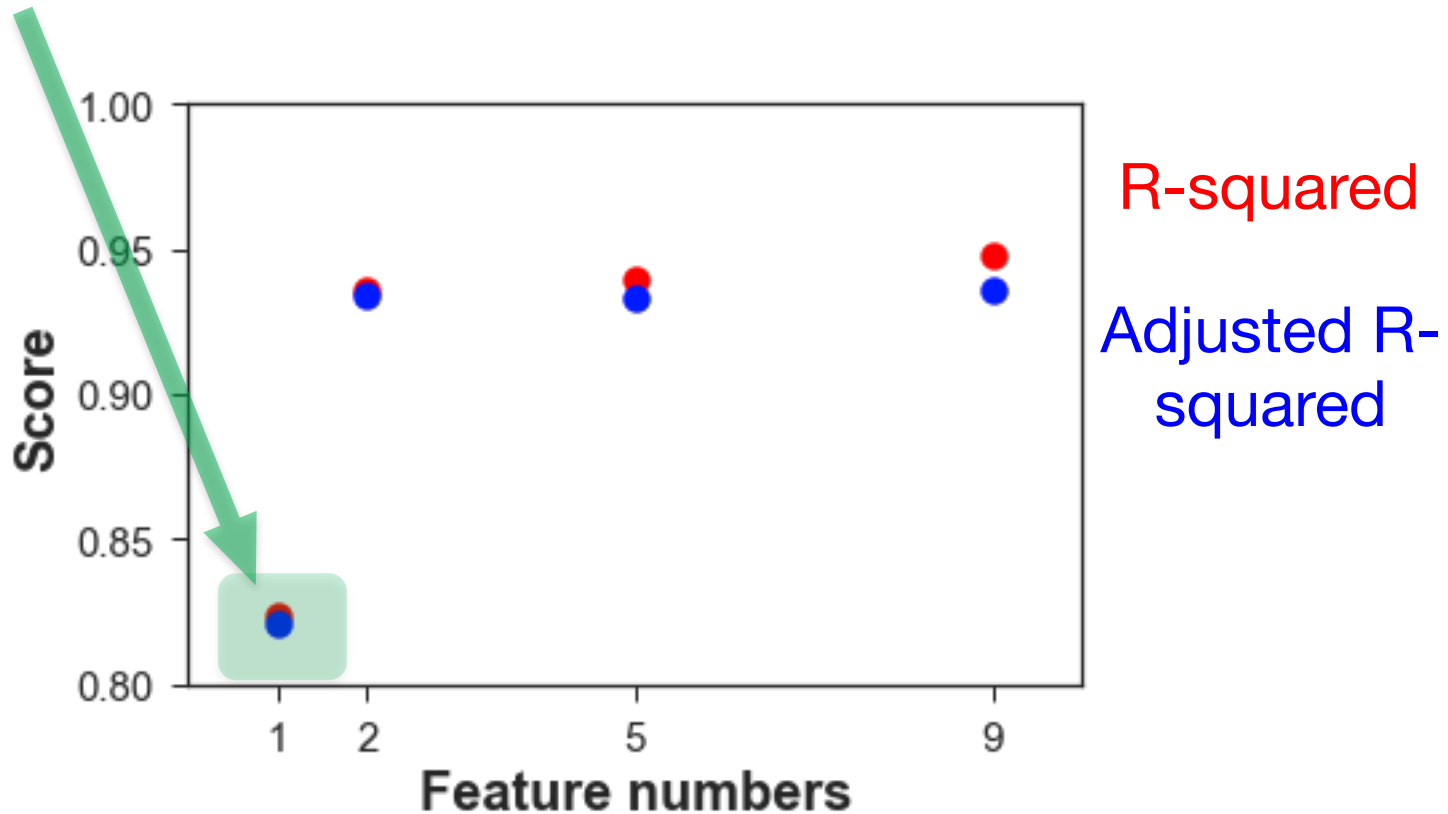


Start from one feature



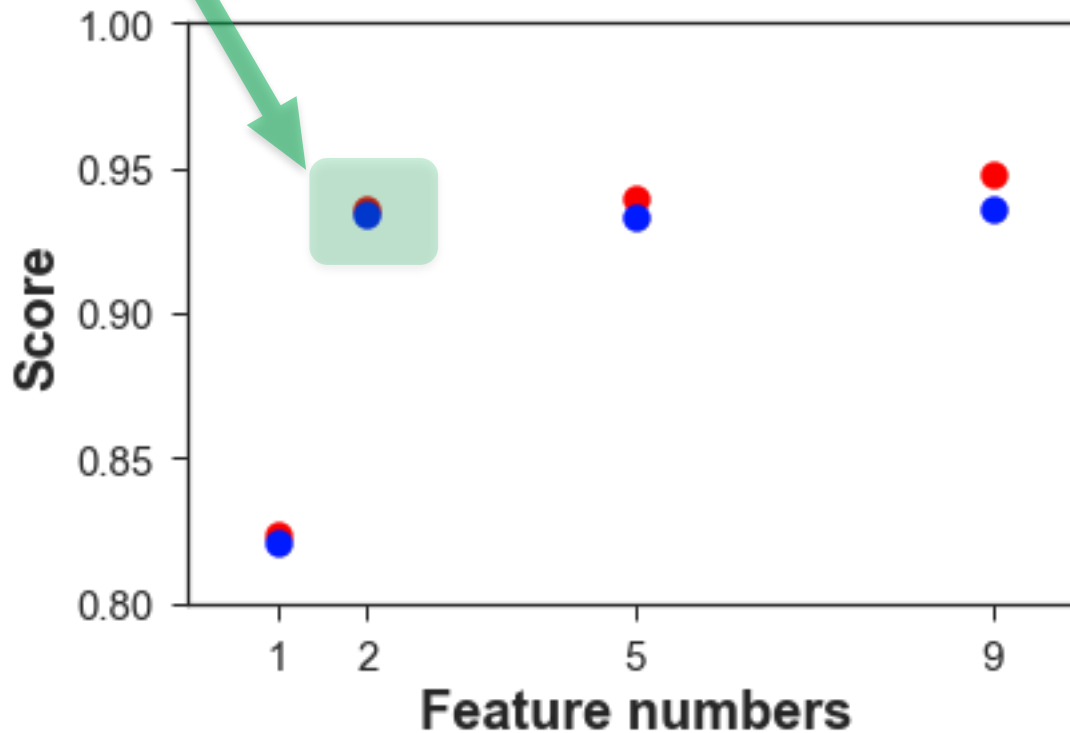
Add more features

Target	Features											
Points	Goal_for	Goal_against	Goal_diff	Assist	Shots On_Goal	Shots	Goals_Header	Goals_Kicked	Cards_Yellow	Cards_Red	Fouls	Offsides



Add more features

Target	Features											
Points	Goal_for	Goal_against	Goal_diff	Assist	Shots On_Goal	Shots	Goals_Header	Goals_Kicked	Cards_Yellow	Cards_Red	Fouls	Offsides



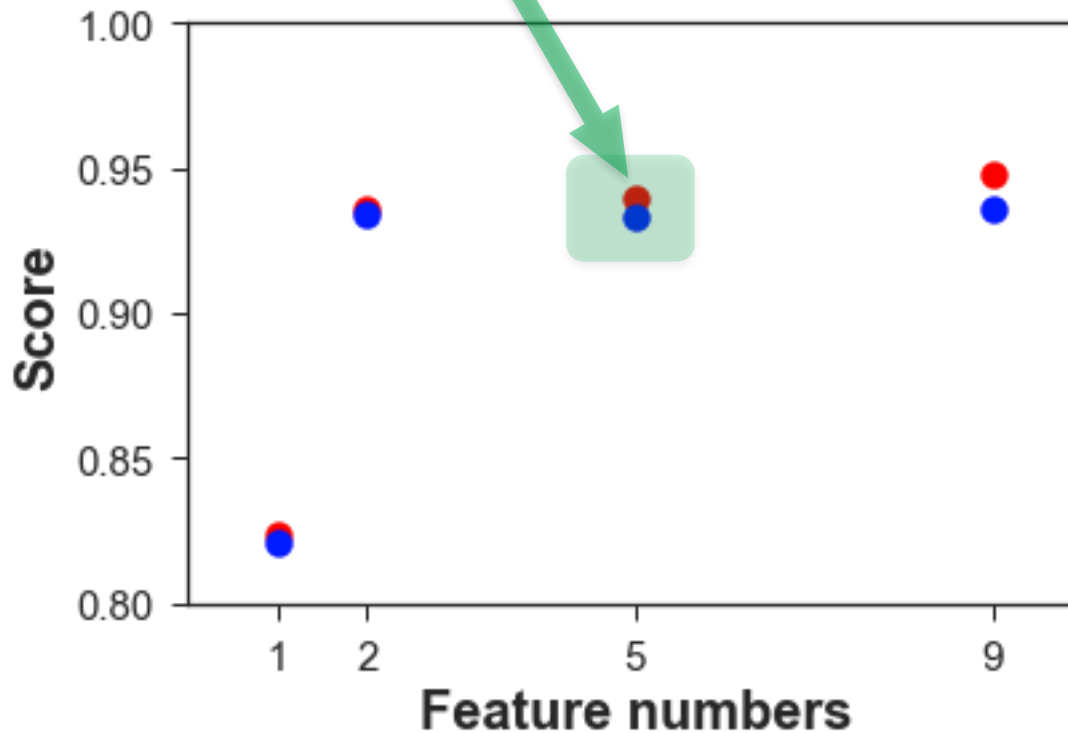
R-squared

Adjusted R-squared



Add more features

Target	Features											
Points	Goal_for	Goal_against	Goal_diff	Assist	Shots On_Goal	Shots	Goals_Header	Goals_Kicked	Cards_Yellow	Cards_Red	Fouls	Offsides



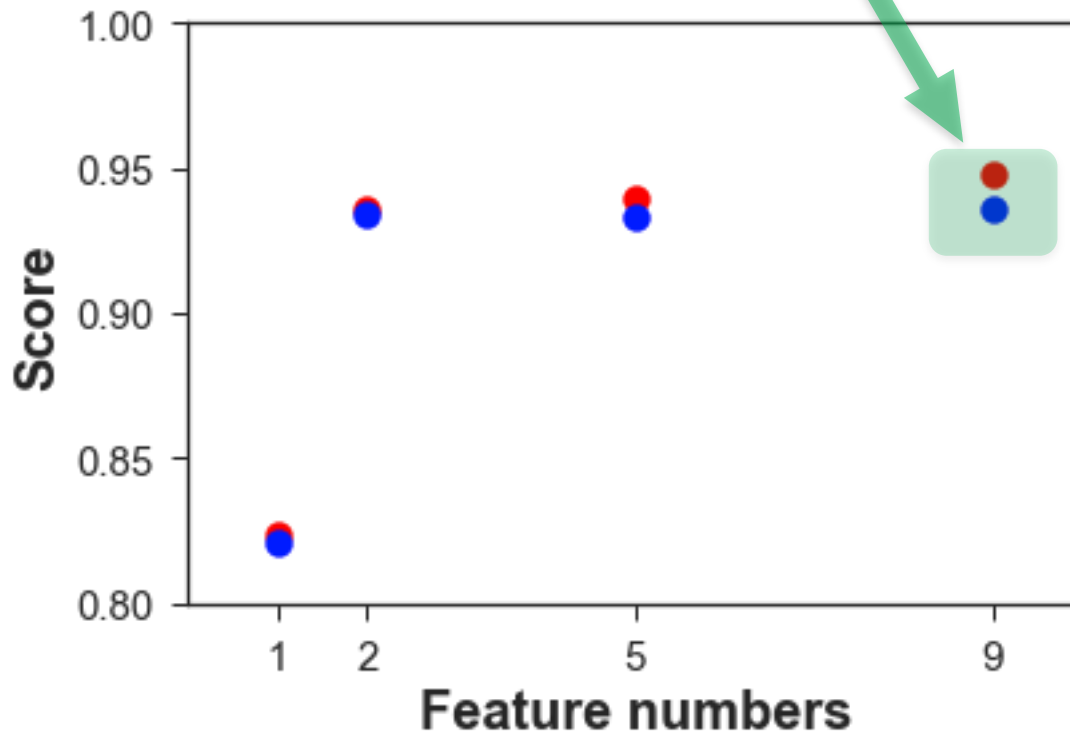
R-squared

Adjusted R-squared



Add more features

Target	Features											
Points	Goal_for	Goal_against	Goal_diff	Assist	Shots On_Goal	Shots	Goals_Header	Goals_Kicked	Cards_Yellow	Cards_Red	Fouls	Offsides



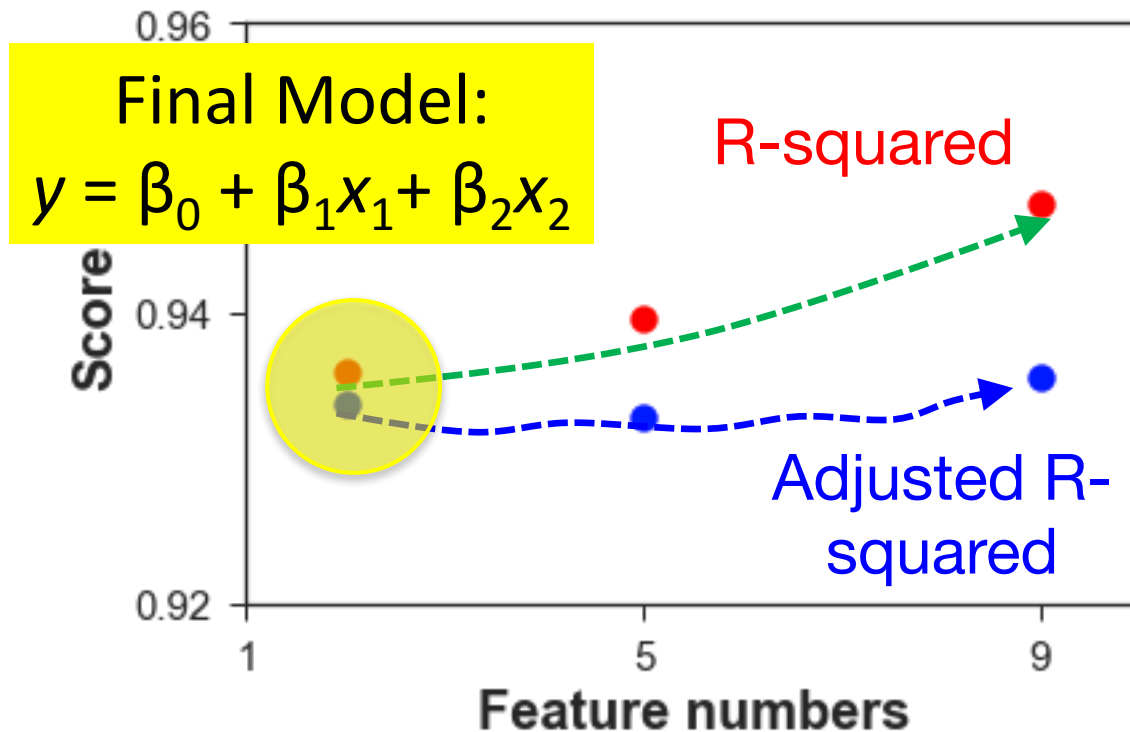
R-squared

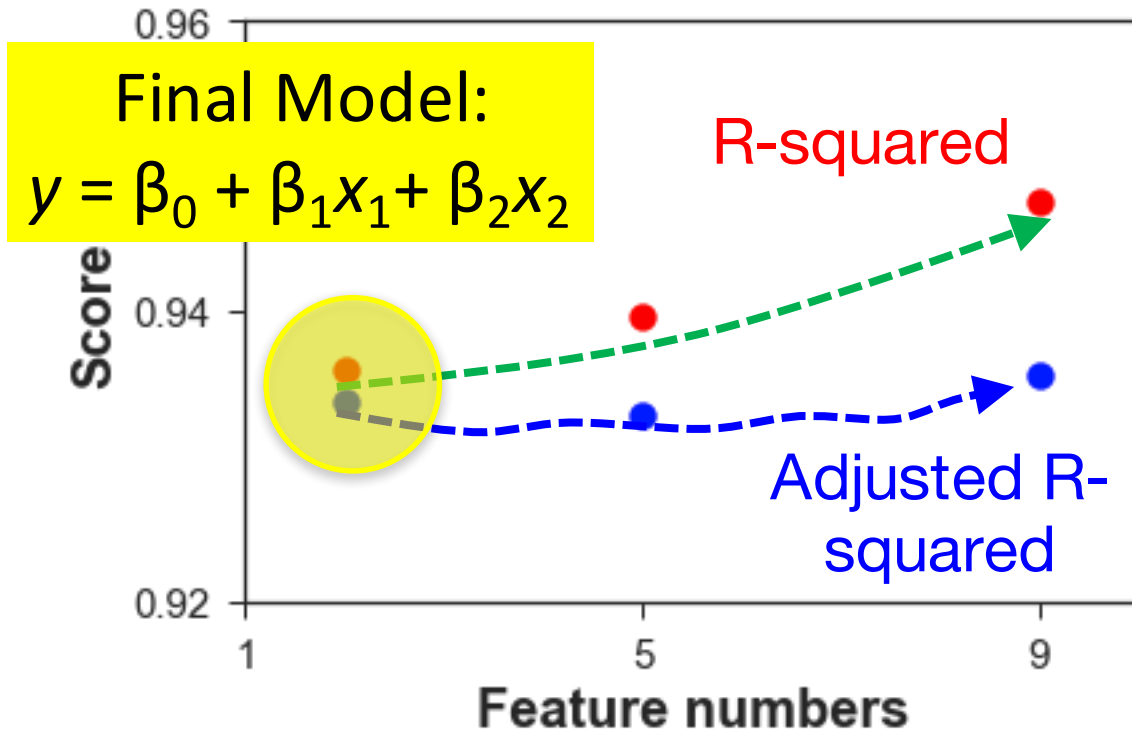
Adjusted R-squared



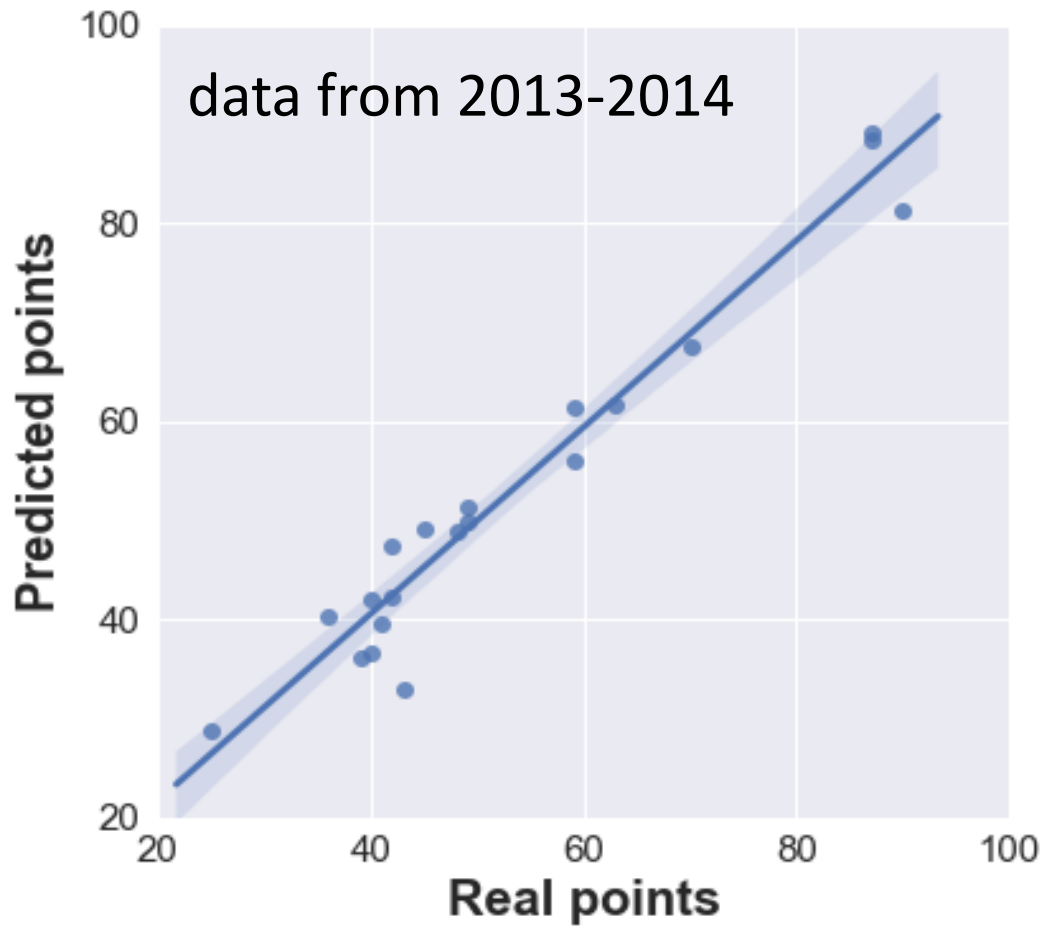
R-squared can be misleading

Target	Features											
Points	Goal_for	Goal_against	Goal_diff	Assist	Shots On_Goal	Shots	Goals_Header	Goals_Kicked	Cards_Yellow	Cards_Red	Fouls	Offsides

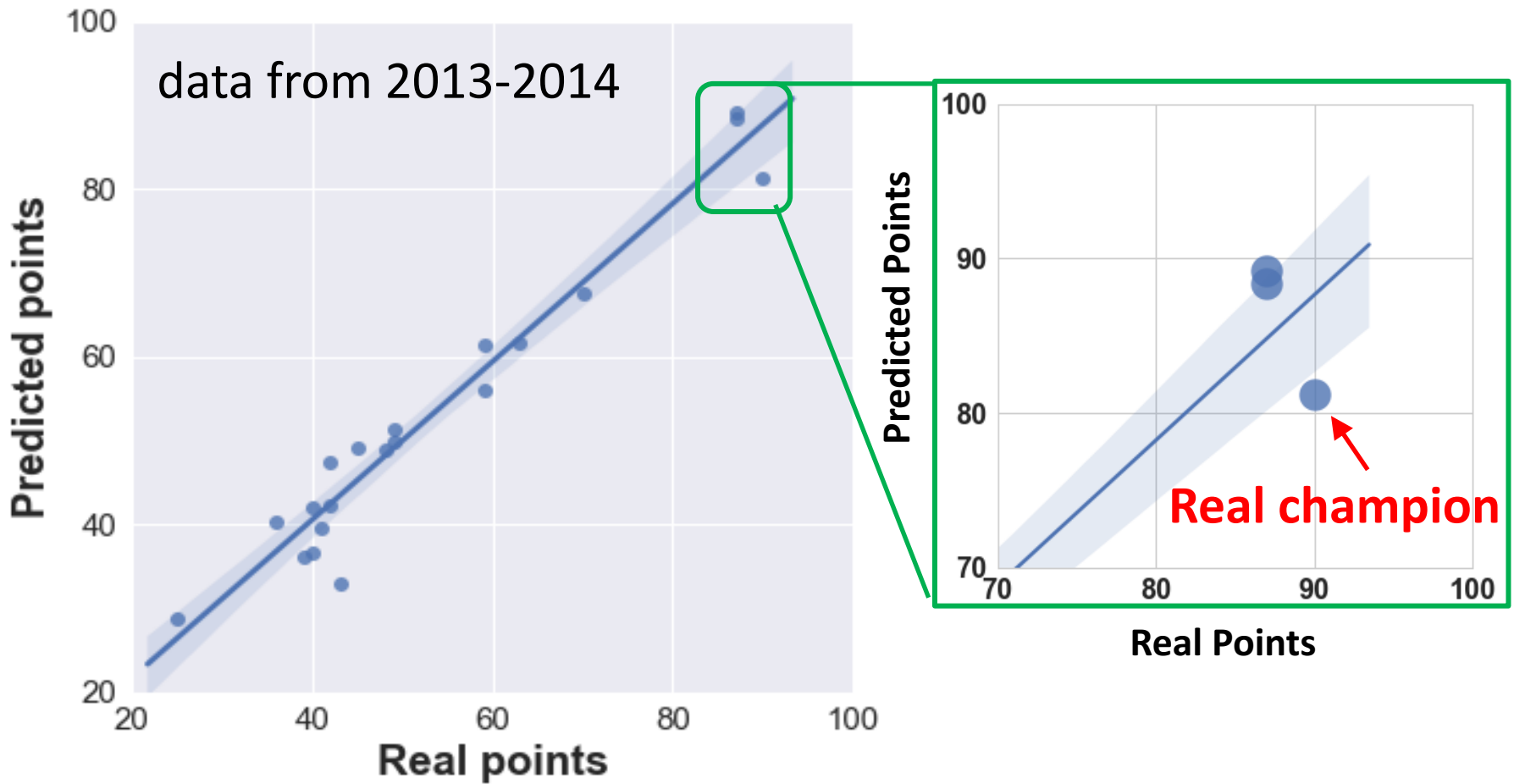




Validate the model with test data



The model is not perfect



Conclusions so far

- Q1: is it possible to **correlate** the points?
 - *Yes*
 - Linear regression gives a reasonable first-approximation
- Q2: is it possible to **predict** the points?
 - *Yes*
 - More in-depth analysis is necessary



Thank you for your attention!

