**Homework 4**

**Cynthia Khan**

**10/16/2018**

Section A: Total points based on average market value of each team

1. LR model using top 10 teams as training and bottom 10 as test – team.py

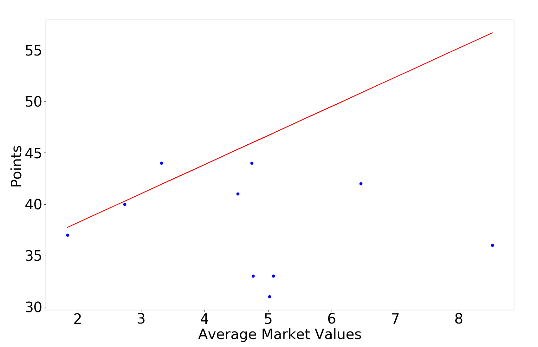


Fig: 1

(Sorry for the extra small font)

The Red line is the Linear Regression line plotted against the last 10 rows of Average Market value and Total Points.

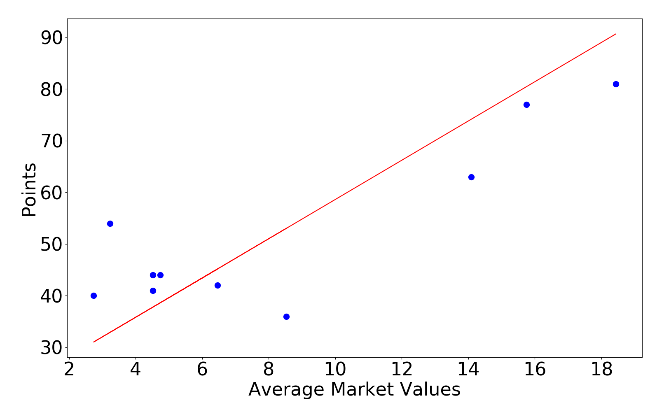
1. LR using randomly selected 10 teams as training and the remaining as test – RandomTeam.py

Fig : 2

Here the Red line of Linear Regression is plotted against a scatter plot of Average Market Values and Total Points of 10 teams picked randomly.

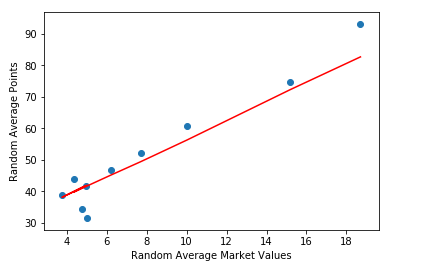
1. LR using average from random selection 20 times – avgRandTeam.py

Fig: 3

The results from plotting randomly selected 10 teams and experiment repeated 20 times.

1. Which method gives more accurate result?

From the three figures, it is very evident that the results from the average of the experiment being conducted 20 times using randomly selected teams (Fig: 3) is much more accurate than the first two.

Section B: Total points based on average age and number of foreign players in each team

1. LR model using top 10 teams as training and bottom 10 as test – multiLR.py

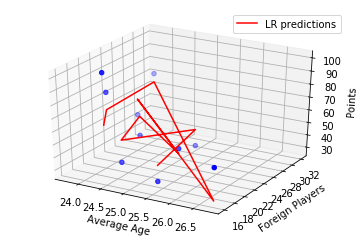


Fig: 4

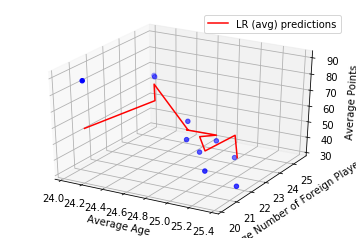
1. LR using randomly selected 10 teams as training and the remaining as test – multiRandLR.py

Fig: 5

1. LR using average from random selection 20 times – multiAvgRandLR.py

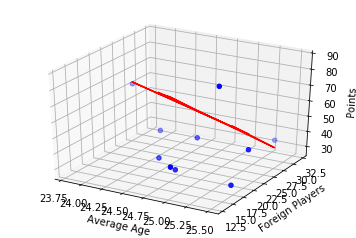
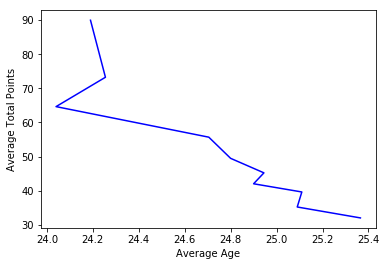


Fig: 6

1. Between A and B which gives better prediction?

Usually Multiple Linear Regression produces better results compared to Simple Regression. However, that heavily depends on a case-by-case scenario. In our case, multiple linear regression does not produce a good prediction model since the independent attributes (Age and No. of Foreign Players) do not have a strong correlation with the dependent attribute (Total Points). This is shown by the two independent plots below:

