

## 6. Aproksimacija funkcija

### Predviđanje ishoda utakmica

Tim koji stoji iza sajta <http://www.football-data.co.uk/> razvio je model za kvote, za kvote predstojećih utakmica. Definirali su indikator kvaliteta nekog tima:

$$g = g_{\text{scored}} - g_{\text{conceded}}$$

, gde je  $g$  razlika postignutih golova  $g_{\text{scored}}$  i primljenih golova  $g_{\text{conceded}}$ . Pri predviđanju rezultata predstojeće utakmice, formira se ukupna razlika za oba tima koji učestvuju u poslednjih  $n$  utakmica pred predstojećom (preporučeno je  $n=6$ ). Definirali su i *match rating*:

$$MR = g_{\text{home}} - g_{\text{away}}$$

, gde je  $g_{\text{home}}$  razlika golova domaćina (za poslednjih  $n$  odigranih utakmica pred predstojećom), a  $g_{\text{away}}$  razlika golova gostujućeg tima (za poslednjih  $n$  odigranih utakmica pred predstojećom). *Match rating* predstavlja ulazni parametar u predviđanju ishoda.

Pre toga se moraju formirati aproksimirati 3 funkcije koje na osnovu izračunatog  $MR$  za predstojećom utakmicu predviđaju verovatnoću sva 3 moguća ishoda (svaka po jedan): pobeda domaćina, nerešen ishod i pobeda gostujućeg tima.

Za svaku prethodno odigranu utakmicu (svih timova) iz perioda tokom više godina potrebno je za svaku pojavu istog *match rating*-a sumirati ukupan broj pobeda domaćina, nerešenog ishoda i pobeda gostujućeg tima, a zatim predstaviti ove odnose procentualno:

<i>Match rating</i>	<i>Number of home wins</i>	<i>Number of draws</i>	<i>Number of away wins</i>	<i>% of home wins</i>	<i>% of home draws</i>	<i>% of away wins</i>
-26	0	1	1	0.0%	50.0%	50.0%
-23	0	0	2	0.0%	0.0%	100.0%
-22	0	0	3	0.0%	0.0%	100.0%
-21	0	2	4	0.0%	33.3%	66.7%
-20	2	2	7	18.2%	18.2%	63.6%
-19	1	1	3	20.0%	20.0%	60.0%
-18	5	7	9	23.8%	33.3%	42.9%
-17	7	9	12	25.0%	32.1%	42.9%
-16	6	14	21	14.6%	34.1%	51.2%
⋮						

Slika 1. *Match rating* utakmica tokom više godina

Na osnovu 1. i 5. kolone tabele formira se aproksimacioni polinom  $p_{\text{home}}(MR)$  koji predviđa pobedu domaćina na osnovu izračunatog  $MR$  za predstojećom utakmicu.

Na osnovu 1. i 6. kolone tabele formira se aproksimacioni polinom  $p_{\text{draw}}(MR)$  koji predviđa nerešen ishod na osnovu izračunatog  $MR$  za predstojećom utakmicu.

Na osnovu 1. i 7. kolone tabele formira se aproksimacioni polinom  $p_{\text{away}}(MR)$  koji predviđa pobedu gosta na osnovu izračunatog  $MR$  za predstojećom utakmicu.

Na osnovu predviđene verovatnoće sva 3 ishoda za poznati  $MR$ , moguće je odrediti kvote:

$$Q_{home} = \frac{100}{p_{home}(MR)}$$

$$Q_{draw} = \frac{100}{p_{draw}(MR)}$$

$$Q_{away} = \frac{100}{p_{away}(MR)}$$

Za svaki aproksimacioni polinom  $p(x)$  izračunat na osnovu parova  $(x_i, y_i)$ , potrebno je utvrditi **pouzdanost aproksimacije** zavisno promenljive  $\tilde{y}_i$  od nezavisno promenljive  $x_i$ :

$$\tilde{y}_i = p(x_i)$$

Srednja vrednost zavisno promenljive na osnovu koje je formiran aproksimacioni polinom je:

$$\bar{y} = \frac{\sum_i y_i}{i}$$

, gde su  $y_i$  vrednosti zavisno promenljive korišćenje u izračunavanju aproksimacionog polinoma.

Definišu se 3 mere:

a) suma kvadrata grešaka regresije:

$$SSR = \sum_i (\tilde{y}_i - \bar{y})^2$$

b) suma kvadrata grešaka:

$$SSE = \sum_i (y_i - \tilde{y}_i)^2$$

c) ukupna suma kvadrata:

$$SST = \sum_i (y_i - \bar{y})^2$$

$$SST = SSR + SSE$$

Pouzdanost aproksimacije je određena **koeficijentom determinacije**:

$$r^2 = \frac{SSR}{SST}$$

Njegova vrednost se kreće u rasponu  $[0,1]$ . Više vrednosti bi ukazale na pouzdano predviđanje.

### Zadatak 1

Izračunati kvote za sva 3 moguća ishoda predstojeće utakmice na osnovu poznate fiktivne tabele odigranih utakmica u proteklom periodu, kao i poznatih rezultata prethodnih 6 utakmica za domaći i gostujući tim. (instalirati biblioteku **Pandas** i **xlrd** za učitavanje xls fajla – “pip install pandas xlrd”)

a) Učitati tabelu *games.xls* i izolovati prvu i poslednje 3 kolone:

```
import pandas as pd
```

```
games = pd.read_excel('games.xls')
```

```
MR = games.iloc[:,0]
```

```
home = games.iloc[:,4]
```

```
draw = games.iloc[:,5]
```

```
away = games.iloc[:,6]
```

**b)** Naći aproksimacione polinome 3. stepena za sva 3 ishoda:

.  
.  
.

**Rezultat:**

```
pH =  
    0.0000    -0.0000    0.0099    0.4705
```

```
pD =  
   -0.0000   -0.0003    0.0002    0.2938
```

```
pA =  
   -0.0000    0.0003   -0.0101    0.2356
```

**c)** Naći aproksimirane vrednosti polinoma za svaki ulazni MR:

```
pHome = np.polyval(pH, MR)  
pDraw = np.polyval(pD, MR)  
pAway = np.polyval(pA, MR)
```

**d)** Naći koeficijente determinacije za sva 3 polinoma:

.  
.  
.

**Rezultat:**

```
r2Home =  
    0.8898
```

```
r2Draw =  
    0.4306
```

```
r2Away =  
    0.7604
```

`r2Home` i `r2Away` daju dosta pouzdanu aproksimaciju za razliku od `r2Draw`.

**e)** Definisati funkciju `g` koja izračunava meru kvaliteta tima na osnovu rezultata poslednjih  $n$  utakmica učitanih iz `.xls` datoteke :

```
def g(file):  
    games = pd.read_excel(file)
```

.  
.  
.

- f) Uz pomoć definisane funkcije `g` izračunati meru kvaliteta tima domaćina i gostujućeg tima na osnovu rezultata učitanih iz `.xls` datoteka, a zatim izračunati  $MR$  za predstojeću utakmicu i upotrebiti ga za određivanje kvota za sva 3 njena ishoda:

```
gHome1 = g('home.xls')
gAway1 = g('away.xls')
MR1 = ...
pHome1 = ...
pDraw1 = ...
pAway1 = ...
QHome1 = round(100/pHome1)/100
QDraw1 = round(100/pDraw1)/100
QAway1 = round(100/pAway1)/100
```

**Rezultat:**

```
QHome1 =
    2

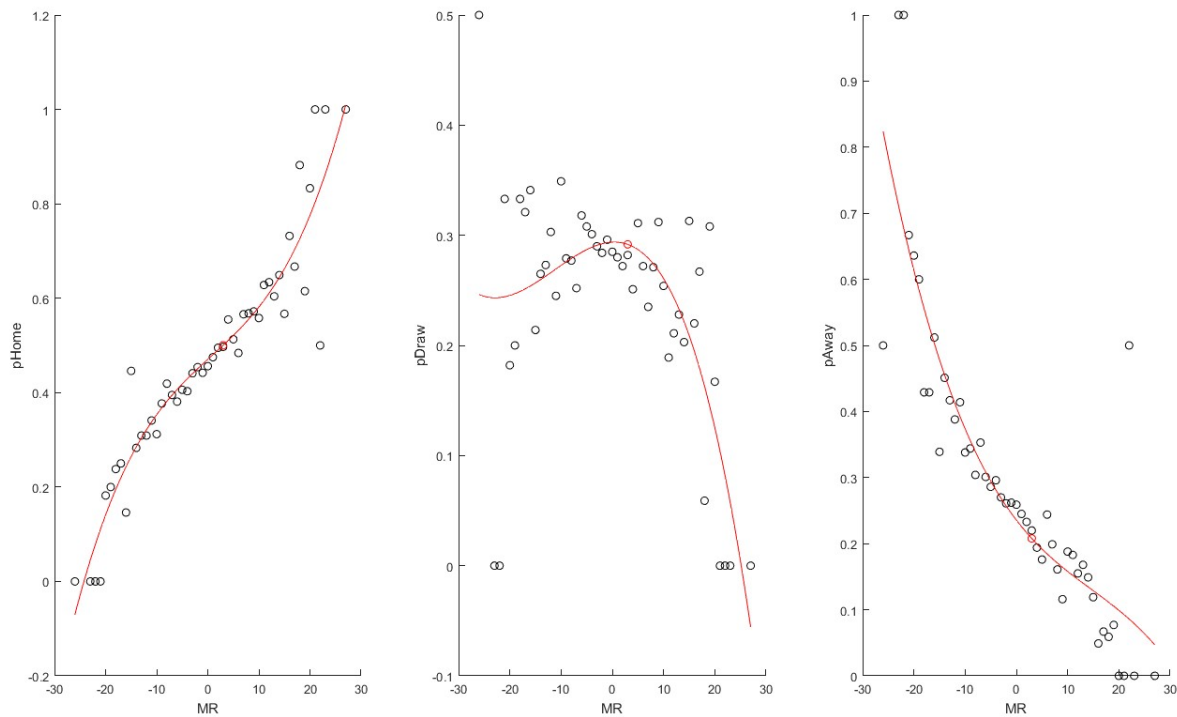
QDraw1 =
    3.4300

QAway1 =
    4.8100
```

- g) Nacrtati sva 3 polinoma, tačke uz pomoć kojih su dobijene i prikazati aproksimirane tačke za izračunati  $MR$ :

```
x = np.linspace(np.min(MR), np.max(MR), 100)
plt.subplot(1, 3, 1)
plt.scatter(MR, home, c='black')
plt.plot(x, np.polyval(pH, x), 'red')
plt.scatter(MR1, pHome1, 'red')
plt.xlabel('MR')
plt.ylabel('pHome')
plt.subplot(1, 3, 2)
plt.scatter(MR, draw, 'black')
plt.plot(x, np.polyval(pD, x), 'red')
plt.scatter(MR1, pDraw1, c='red')
plt.xlabel('MR')
plt.ylabel('pDraw')
plt.subplot(1, 3, 3)
plt.scatter(MR, away, c='black')
plt.plot(x, np.polyval(pA, x), 'red')
plt.scatter(MR1, pAway1, c='red')
plt.xlabel('MR')
plt.ylabel('pAway')
```

Rezultat:



Slika 1. Polinomi 3. Stepena

- h) Povećavati stepen sva 3 polinoma polinoma sve dok su problemi nalaženja polinoma dobro uslovljeni, a zatim izračunati njihov koeficijent determinacije i ponoviti iscrtavanje pod g):

rešenje:

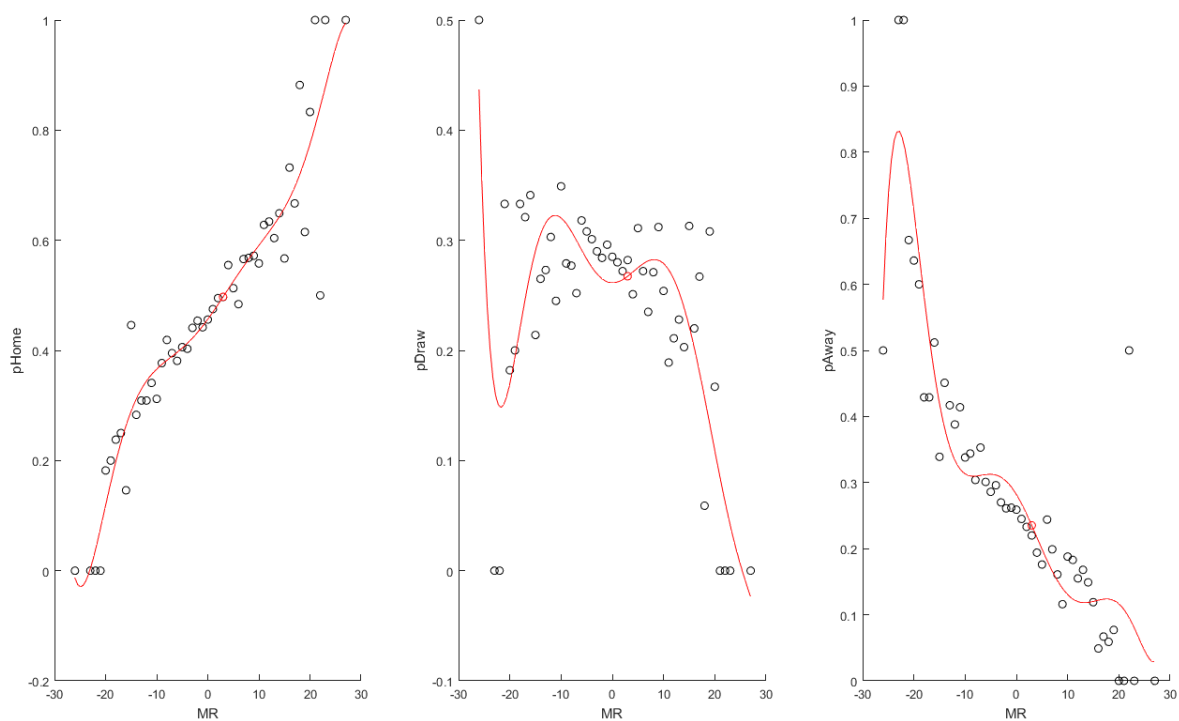
*stepen=7*

Rezultat:

r2Home =  
0.8952

r2Draw =  
0.6025

r2Away =  
0.8359



Slika 1. Polinomi 7. Stepena