CLASSIFICATION BASED ON HISTORY PERFORMANCE

ZHOU ZHIHAO

2014-4-1

BASIC STRUCTURE

Response:
whether Actual Total Goal over market expectation
2 class: 0 and 1

Predictors:
1. 2 dummy variables to divide a match into 4 types:
    - Home – Away
    - 1 – 1 : Strong vs. Not Strong
    - 1 – 0 : Strong vs. Not Strong
    - 0 – 1 : Not Strong vs. Strong
    - 0 – 0 : Not Strong vs. Not Strong
2. K previous matches performance on actual total goal over market expectation

Total no. of predictors: **2 + 2 * k**

DATA SETS

Raw data set:

- 8 seasons BPL Match data
- From season 2005-06 to season 2012-2013
- Total 380* 8 = 3040 matches

	Date	HomeTeam	AwayTeam	FTHG	FTAG	Lambda.H.	Lambda.A.
1	18/8/12	Arsenal	Sunderland	0	0	1.892	0.556
2	18/8/12	Fulham	Norwich	5	0	1.523	0.832
3	18/8/12	Newcastle	Tottenham	2	1	1.192	1.139
4	18/8/12	QPR	Swansea	0	5	1.363	0.894
5	18/8/12	Reading	Stoke	1	1	1.219	1.030
6	18/8/12	West Brom	Liverpool	3	0	0.866	1.415
7	18/8/12	West Ham	Aston Villa	1	0	1.282	1.000
8	19/8/12	Man City	Southampton	3	2	2.437	0.382
9	19/8/12	Wigan	Chelsea	0	2	0.651	1.641
10	20/8/12	Everton	Man United	1	0	0.842	1.442
11	22/8/12	Chelsea	Reading	4	2	2.226	0.468
12	25/8/12	Aston Villa	Everton	1	3	1.066	1.211
13	25/8/12	Chelsea	Newcastle	2	0	1.814	0.675

DATA SETS

Methods to process raw data sets:

1. process_data = function(data)

2. classify.strong = function(data,k=1.3)
   - k is the threshold to divide teams into Strong and Not Stong
   - Strong ⇔ team's average goal per match > average goal for all team * k

3. add_previous_performance = function(data,lag=3)
   - lag specifies no. of previous matches involved

DATA SETS

After processing use **k = 1.3, lag = 3**:
No. of observations = 2808

	row.names	HomeTeam	AwayTeam	H_level	A_level	y	Home_p1	Home_p2	Home_p3	Away_p1	Away_p2	Away_p3
1	30	Arsenal	Southampton	1	0	1	-0.278	-2.251	-2.448	2.474	-0.289	2.181
2	31	Aston Villa	Swansea	0	0	0	-0.332	1.723	-1.282	1.753	0.788	2.743
3	32	Fulham	West Brom	0	0	1	0.655	2.353	2.645	-0.278	-0.440	0.719
4	33	Man United	Wigan	1	0	1	2.474	2.353	-1.284	1.816	-0.289	-0.292
5	34	Norwich	West Ham	0	0	0	-0.723	-0.306	2.645	0.655	0.788	-1.282
6	35	QPR	Chelsea	0	1	0	1.033	-0.306	2.743	-0.489	3.306	-0.292
7	36	Stoke	Man City	0	0	0	1.816	-2.251	-0.249	1.033	1.675	2.181
8	37	Sunderland	Liverpool	0	1	0	0.000	0.000	0.000	-0.278	1.675	0.719
9	38	Reading	Tottenham	0	0	1	0.000	0.000	0.000	-0.723	-0.440	0.669
10	39	Everton	Newcastle	0	0	1	-0.278	1.723	-1.284	-0.332	-0.489	0.669
11	40	Chelsea	Stoke	1	0	0	-2.243	-0.489	3.306	-0.209	1.816	-2.251
12	41	Southampton	Aston Villa	0	0	1	4.361	2.474	-0.289	-0.236	-0.332	1.723
13	42	Swansea	Everton	0	0	1	-0.236	1.753	0.788	1.768	-0.278	1.723
14	43	West Brom	Reading	0	0	0	0.797	-0.278	-0.440	1.711	3.306	-0.249
15	44	West Ham	Sunderland	0	0	0	-2.270	0.655	0.788	-0.278	1.753	-2.448
16	45	Wigan	Fulham	0	0	1	1.269	1.816	-0.289	0.797	0.655	2.353

ANN RESULT

Size =5, lag = 3, k = 1.3

Training Result: 59.2%

p1	0	1
0	661	464
1	400	593

Testing Result: 52.2%

	0	1
0	176	166
1	164	184

Try different size, but testing result almost only around 50%, seems not significant.

May try to:

- Adjust k
- Adjust lag
- Adjust ANN parameters

ANN RESULT

Size =5, lag = 5, k = 1.2

Training Result: 67.2%

p5	0	1
0	762	433
1	286	637

Testing Result: 52.2%

	0	1
0	176	166
1	164	184