Solution

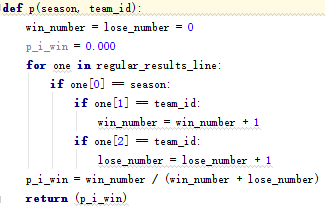
Question one:



Using this formula, which year’s tournament results had the highest probability?

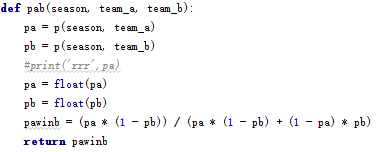
Solution:

I use the data of regular\_season\_results.csv to calculate the probability that A wins in a matchup against B.

The code to calculating the winning percentage of each team :

The formula as follows:

Then we can use the P(A wins |A and B play) formula calculate the P(A wins |A and B play)(using “pawinb” named this value) , and the code as follows:



We had known the data of tourney results in tourney\_results.csv. So if team A play with B in tournament, we can calculate the “pawinb”, if pawinb > 0.5, we can come to conclusion that people may think A had the high probability to win. And if the true tourney result is same as the prediction, we record the pre\_maybe\_true\_num to record how many tournament games the people predict true.

We use this formula:

score=pre\_maybe\_true\_numfloat(tourney\_num)

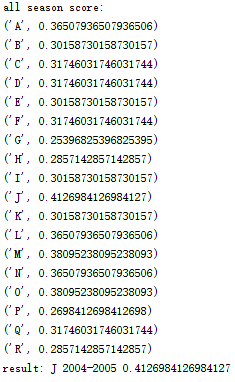
tips:

pre\_maybe\_true\_num : The number of people had a true prediction.

toutney\_num : The match number of tourney.

So the more higher the score is, the more greater the chance of people have an right prediction.

Use the code had get the score of each season:



In conclusion:

we can see the season J(2004-2005) had the highest probability.

Question two:

According to the problems, we get people to guess the result for a game, the largest probability of one person can guess right one of the game:

Thus we can calculate the probability of one person can guess right all:

According to the data in tourney\_slots. CSV we can see that: match\_number = 63

So:

=

The above results is the probability of get a million dollar prize.

People\_num: The number of people who join this game.

One\_price: The money people should charge to join the game.

Assumption:

people\_num=

One\_price=2

So：

5=2.955110687

So AIG need not charge Sanguine Bank for the insurance contract.