

Bekir Kagan Yildiz

Instructor's Name : Selim Temizer

23 April 2017

## Part 4 Report :

In this part I have created a CSVSolver script that creates a new data file from old data . In order to make it more flexible I have added window size and shift as argvs and script invoked by using the invocation below :

— python CSVSolver.py <filename> <window size> <shift>

By using CSVSolver script I have created two files one OHLC and the other is Volume and In order to make prediction more accurate I have set window size 4 and shift as 1 .

By doing so I have two files (GOOG-2016-OHLC.csv and GOOG-2016-Volume.csv) and I have used those files in order to get predictions from other parts .

Part1 :

$$w_0(1.0) = 25.0180484716$$

$$w_1(\text{OHLC1}) = 0.0217625751833$$

$$w_2(\text{OHLC2}) = 0.0435864447894$$

$$w_3(\text{OHLC3}) = -0.38267669434$$

$$w_4(\text{OHLC4}) = 1.28346748979$$

$$\text{RMS Error} = 7.40935990126$$

Those values are that I found using the Analytical Solver for OHLC's .

$$w_0(1.0) = 459328.740498$$

w1 (Volume1)=-0.0224868087849

w2 (Volume2)=0.084772740245

w3 (Volume3)=0.062841322304

w4 (Volume4)=0.627078775299

RMS Error = 652029.219218

Those values are that I found using the Analytical Solver for Volume's .

Part 2 :

In this part I had some difficulties because if I use the same mu and same continuation constant It only took one iteration to stop and it gives the values below :

Iteration 1 : [0.0014858002786295543, 1.107086566822945, 1.1068460693922595, 1.1066100099240437, 1.10638778149028], RMS Error = 2550.8988834

w0 (1.0) = 0.00148580027863

w1 (OHLC1)=1.10708656682

w2 (OHLC2)=1.10684606939

w3 (OHLC3)=1.10661000992

w4 (OHLC4)=1.10638778149

RMS Error = 2550.8988834

It's most probably because of the mu value because if it's too much Iteration goes to inf and cause problem to execution that's why I decreased the value of mu.

If I use the 0.0000001 0.000001 respectively for mu and continuation constant It takes too much iteration I have seen 10k iteration and stopped to process using keyboard Interrupt.

I increased to values of continuation constant in order to decrease the iteration but still has the same problem continuation constant 0.00001 and 0.0001.

So I increased the value to 0.001 and I get this output for OHLC

Iteration 0 : [0.0, 0.0, 0.0, 0.0, 0.0], RMS Error = 743.692493064

Iteration 1 : [0.00014858002786295545, 0.11070865668229452, 0.11068460693922594, 0.11066100099240436, 0.110638778149028], RMS Error = 414.330573586

Iteration 2 : [0.00023133940520913873, 0.17236296703527704, 0.17232724237641647, 0.17229470517039183, 0.17226650351010273], RMS Error = 230.995534732

Iteration 3 : [0.0002774399000011423, 0.2066964062908519, 0.2066558988004148, 0.20662260327400883, 0.20659746980645943], RMS Error = 129.071865595

Iteration 4 : [0.0003031231904594481, 0.22581345699555222, 0.22577200513807655, 0.22574250307408777, 0.22572547642263327], RMS Error = 72.6332837959

Iteration 5 : [0.00031743509684283017, 0.236455723800998, 0.23641546532424118, 0.23639229169870707, 0.23638617801162035], RMS Error = 41.7654303573

Iteration 6 : [0.0003254136985322025, 0.242377950712478, 0.24234007614814917, 0.2423246426744037, 0.24233100472054805], RMS Error = 25.4750052353

Iteration 7 : [0.0003298649592900854, 0.24567134646645356, 0.24563651879326184, 0.24562961158226515, 0.24564931962899939], RMS Error = 17.5838937306

Iteration 8 : [0.0003323516636701335, 0.24750061203142856, 0.2474692004068718, 0.24747125713966922, 0.24750479576489307], RMS Error = 14.2755544229

Iteration 9 : [0.0003337442062204748, 0.2485144288720399, 0.24848663880957725, 0.24849790310880035, 0.2485455420802551], RMS Error = 13.0802598625

Iteration 10 : [0.0003345273541264769, 0.2490740808319005, 0.24905002670326967, 0.24907063411051075, 0.24913252355719057], RMS Error = 12.6864968502

Iteration 11 : [0.00033497109899656023, 0.24938078533862307, 0.2493605307540843, 0.24939055661508816, 0.24946678003306066], RMS Error = 12.5617193902

Iteration 12 : [0.0003352258129491723, 0.24954661065623338,  
0.24953019095438997, 0.24956967708876332, 0.2496602808628187], RMS Error =  
12.522634656

Iteration 13 : [0.0003353752458983891, 0.249633973325636, 0.2496214080991643,  
0.24967037765409916, 0.24977538749986455], RMS Error = 12.5103590476

Iteration 14 : [0.0003354660424494751, 0.2496776363068188, 0.24966893637868698,  
0.2497274021017581, 0.24984683222315104], RMS Error = 12.5064218144

Iteration 15 : [0.0003355241813658034, 0.24969696086795629,  
0.24969213217683695, 0.24976010102354526, 0.24989395921302726], RMS Error =  
12.5050732893

Iteration 16 : [0.00033556413154774476, 0.24970273028830245,  
0.24970177605255747, 0.24977925175250637, 0.24992754223167468], RMS Error =  
12.5045279911

w0 (1.0) = 0.000335564131548  
w1 (OHLC1)=0.249702730288  
w2 (OHLC2)=0.249701776053  
w3 (OHLC3)=0.249779251753  
w4 (OHLC4)=0.249927542232

RMS Error = 12.5045279911

Part 3 :

In part 3 I have used 2016 GOOG datas for training data and the first row of the 2017 data for prediction

```
python WeightedKNNsolver.py GOOG-2016-OHLC1.csv 5 4 829.38999975  
831.9199982499999 827.8887482499999 820.3367310000001
```

And I found this solutions :

k = 5

Attribute 1 (OHLC1) = 829.38999975

Attribute 2 (OHLC2) = 831.91999825

Attribute 3 (OHLC3) = 827.88874825

Attribute 4 (OHLC4) = 820.336731

Instance 1 : Index = 45 , [802.17750525, 811.542496, 809.50250225, 796.96749875] ,  
Class Label = 800.06748975 , Distance = 45.1656674122 , Weight = 0.022140711237

Instance 2 : Index = 46 , [811.542496, 809.50250225, 796.96749875,  
800.0674897499999] , Class Label = 800.59249875 , Distance = 46.7765254001 , Weight =  
0.021378244567

Instance 3 : Index = 44 , [797.8349915, 802.17750525, 811.542496, 809.50250225] ,  
Class Label = 796.96749875 , Distance = 47.5911221841 , Weight = 0.0210123223431

Instance 4 : Index = 43 , [803.1999969999999, 797.8349915, 802.17750525,  
811.542496] , Class Label = 809.50250225 , Distance = 50.8538151473 , Weight =  
0.0196642080266

Instance 5 : Index = 47 , [809.50250225, 796.96749875, 800.0674897499999,  
800.59249875] , Class Label = 792.57124325 , Distance = 52.7356322592 , Weight =  
0.0189625108709

Weighted class label (Label) = 799.965414808

Prediction is 799.96 but the real value is 816.3325044999999 it may caused because of the unstable stock and recently trending and we could get better prediction using last 3 months datas which could give us better prediction.