Lecture 6:

* Prepare your labeled dataset (2017 and 2018)

For every week, you compute 2 numbers (from 5 daily values):

Average daily return, average standard deviation

Week Daily\_return Average\_STD Label

39 2.36 1.5 green

For every day, you have the following:

Open, high, low, (adj close)

You have daily returns

1. Compute an overnight return:

(Open – prev\_close)/prev\_close

1. We have daily returns;

Trading strategy: overnight return has “inertia” for the whole day

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Open | Close | overnight | DEC. | P/L |
| Mon | 100 | 100 |  |  |  |
| Tue | 110 | 95 | 10% | BUY | -15 |
| Wed | 92 | 90 | -3% | SS | 2 |
| Thu | 88 | 85 |  | SS | 3 |
| Fri | 90 | 95 |  | BUY | 5 |

Profit:

Long (buy at open, sell at close) –

Profit = ( close – open) \* shares

Short Position (sell short at open, buy at close) –

Profit = (open – close) \* shares

For each trade, you are given $100.

Shares = 100/open\_price

If open price == close of previous day, you do nothing

Specialists:

Tasks:

What do you need to do with this trading strategy (2018 only)

1. Implement as described without any preliminary analysis
2. Examine your data and modify your algorithm – generate the graph: on y axis you have daily returns, on x axis you have overnight returns
3. Plot the following result:

Suppose you are willing to trade only if overnight change is at least some value R-min

For R\_min = [-10, -9, …, -2, -1, 0, 1, …, 10]

Compute your average profit and plot it.

In addition, we will examine a non-financial data set ‘Tips”

Matplotlib

Seaborn (sns) - datasets

Task #2: load “tips” dataset and answer questions below:

Think of 10 questions that you want to ask (and code):

1. Are tips higher for lunch or dinner?
2. When are tips highest (which day and time)?
3. Is there any relationship between price and tipping percentage?
4. Any relationship between tip (as a percentage) and size of the group
5. What percentage of people are smoking?
6. Assume that rows are arranged in time

Are tips increasing with time?

1. Any correlation between gender and time (assume that each meal is split into two equal times)
2. Correlation between tip amounts from smokers and non-smokers
3. Average tip for each day of the week
4. Which gender smokes more?

Write python code to answer these questions

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Seaborn has many advanced functions for visualization (uses matplotlib and adds more)

kNN – use a library to do this

--- do kNN yourself with three distances:

Euclidean, manhattan and Minkowski

P = 2 p =1 p =1.5