Lecture 7 (March 18, 2019)

Recall:

1. Prepare a labelled data set
2. Implement Day Trading Strategy
3. Analyze tips dataset

Let r be the percent difference between today’s open and yesterday close

r > [0, 1, 2, 3, 4, 5]

r > 0 buy at the open sell at the close

r >1 – buy only if open is higher than yesterday close by at least 1%

r > 2

For each of these values, indicate average p/l

r < 0, -1, -2, … - 5 ----🡪 short positions

For r = -5, -4, -3, -2, -1, 1, 2, 3, 4, 5:

Compute (and plot on two graphs)

1 graph: r on x –axis, P/L on Y axis

2 graph: r on x-axis, %profitable trades on Y axis

What is (if any) a good trading strategy?

Tips data set: finish it!!!!

* kNN - take 2017 as a training set and for each k in [1,3,5,7,9,11] you trained your classifier and applied it to 2018

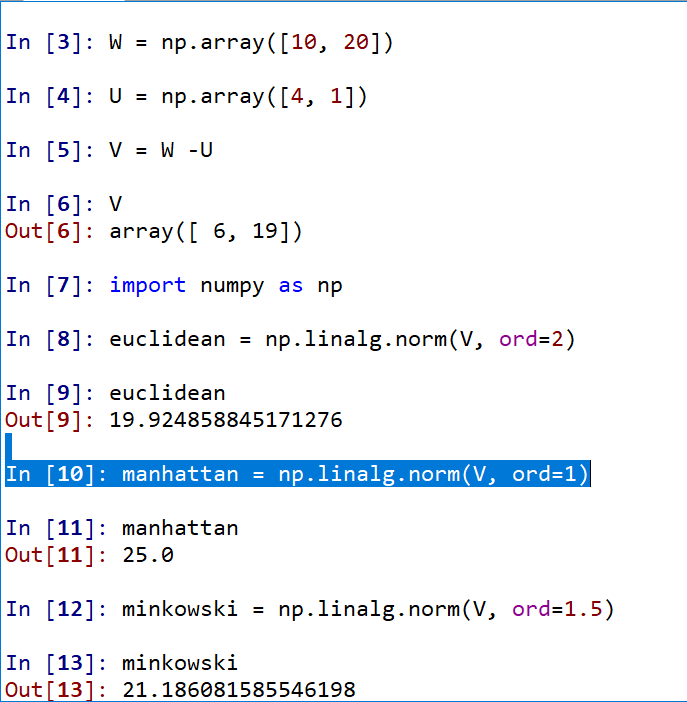
do kNN “manually” (without using sklearn library) and the following 3 distance metrics:

1. Euclidean p =2
2. Manhattan (street) p=1
3. Minkowski distance p =1.5

In “Data\_science\_module\_numpy.pdf”, take a look at pages 27 – 30

W and U are two numpy vectors

V = W - U



“Re-do” your nearest neighbor classification using 3 distances without using sklearn

Summarize your assignment:

1. Finish labels
2. Day trading
3. Implement kNN for 3 distances without sklearn

Q: can we use sklearn to standardize

Or do it yourselves

In any ML algorithm that deals with distances, you need to standardize your data

We will not take off points if you don’t

Naïve Bayesian

Regression (ordinary and logistic)

Decision Trees

Random forests

Clustering