Lecture 9 (April 1)

April 29 - final project presentations

April 29 – May 6 (Mon – Mon) Final

Check class discussion section:

Linear models

Linear regression

Homework 9:

1. Fix window size at 10 days. Re-compute your results using explicit Python code (you can use code on page 16 – do not use sklearn or any other libraries.
2. Estimate slope and intercept using gradient descent method (page 19, add some value for L learning rate).

You will investigate accuracy for different L and different “epochs”

Plot the following graphs:

Consider 5 learning rates L=0.01, 0.02, 0.03, 0.04, 0.05, take the number epochs = 100

Given L = 0.02 and epochs=100, use gradient descent to estimate slope and intercept

Estimate P11 -----------🡪 return for day 11

LOGISTIC REGRESSION

2 years of data 2017 and 2018

(x, y, label)

X – average of 5 daily returns

Y – standard deviation of 5 returns

Assignment for logistic regression:

Part 1:

1. Take a linearly-separable dataset that you created for 2017-2018
2. Apply logistic regression classifier

Look up at the documentation for logistic regression in sklearn and find out the equation that separates your points

On the same graph, plot your points (colored), plot the line that you used to remove extra points and plot the equation

given by the logistic regression classifier

Part 2: look at the original dataset

use points from 2017 to train a logistic regression classifier and apply this classifier to predict the labels for each week in 2018

Compute your accuracy (% of weeks that you correctly predicted the label)

Summary: homework consists of linear regression and logistic regression

Linear regression:

1. Write formulas for slope and intercept (w/o using sklearn) and recompute your results for W=10
2. W=10 and use gradient descent to compute slope and intercept (for different learning rates L=0.01, 0.02, 0.03, 0.04 and 0.05) and then estimate accuracy

Table: L accuracy for 2018

* 1. 48%
  2. 39%

Logistic regression

1. Compute equation for “reduced” data set and plot that together with points and the line that you chose to remove extra points
2. Original data set 2017, logistic regression to predict 2018

What is the next topic(s)

Support Vector machines (SVM)

Decision Trees

Clustering

Ensemble Learning: combining classifiers together

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