starts at 31:25

Project Overview

Use knn learner to predict stock prices. This is what the proj is about

step 1. massage the data ab it so that you’re predicting the right thing

ML4T399.csv is a sine wave – there’s a clear pattern in it so that youre learner should easily be able to forecast it

debug your learner here, and then trivially change the filename to IBM to try it on that.

there’s different examples of sine waves

I suggest you apply your learner to several of these sine waves

step 1 – build your Y data to train on.

stockcharts.com – can learn about 100’s of technical indicators. but pick 3. eg. Bollinger value, momemtum

**then take the historical prices and map them to new data frame with x1,x2,x3, arranged by time**

y – don’t want to predict price b/c want to generalize

rather predict change in price

can fill in the column initially with what the real prices were, but you want to end up with how much did the price change 5 days into the future

the most important step is to try to predict the future by going back in time and seeing what happened

calculate close to close 5 days in the future

use for loop to iterate down the prices to calculate the 5 day change

Why 5 days? need to commit to a distance number to train the forecaster

our predictability seriously degrades 20 days into the future. Best success is in 1 day

you’re going to have to develop your own policy based on predicted change. eg. when forecasted change is >2% buy, and then hold for 5 days

or something fancier: hold till price drops

RL will tell you when to get in or out, but regression can’t

bagging will be applicable to these models

market inefficiencies are nowadays hard to find, but can maybe find them with judicial combinations of indicators

Plot:

plot the data and then superimpose on it the forecasted change

scale the y so that it matches the original figure and plot it

Remaining things to do:

on the real stock data, it will have bad results

but eventually you can find success, but it just takes persistence.

Lin reg has a good chance of doing better b/c …

we’re essentially creating non-linear indicators … taking this non-linear problem and essentially linearizing it

the market notices what you’re doing and tries to defeat you