CS 4803/7646 – MLT (Machine Learning for Trading)

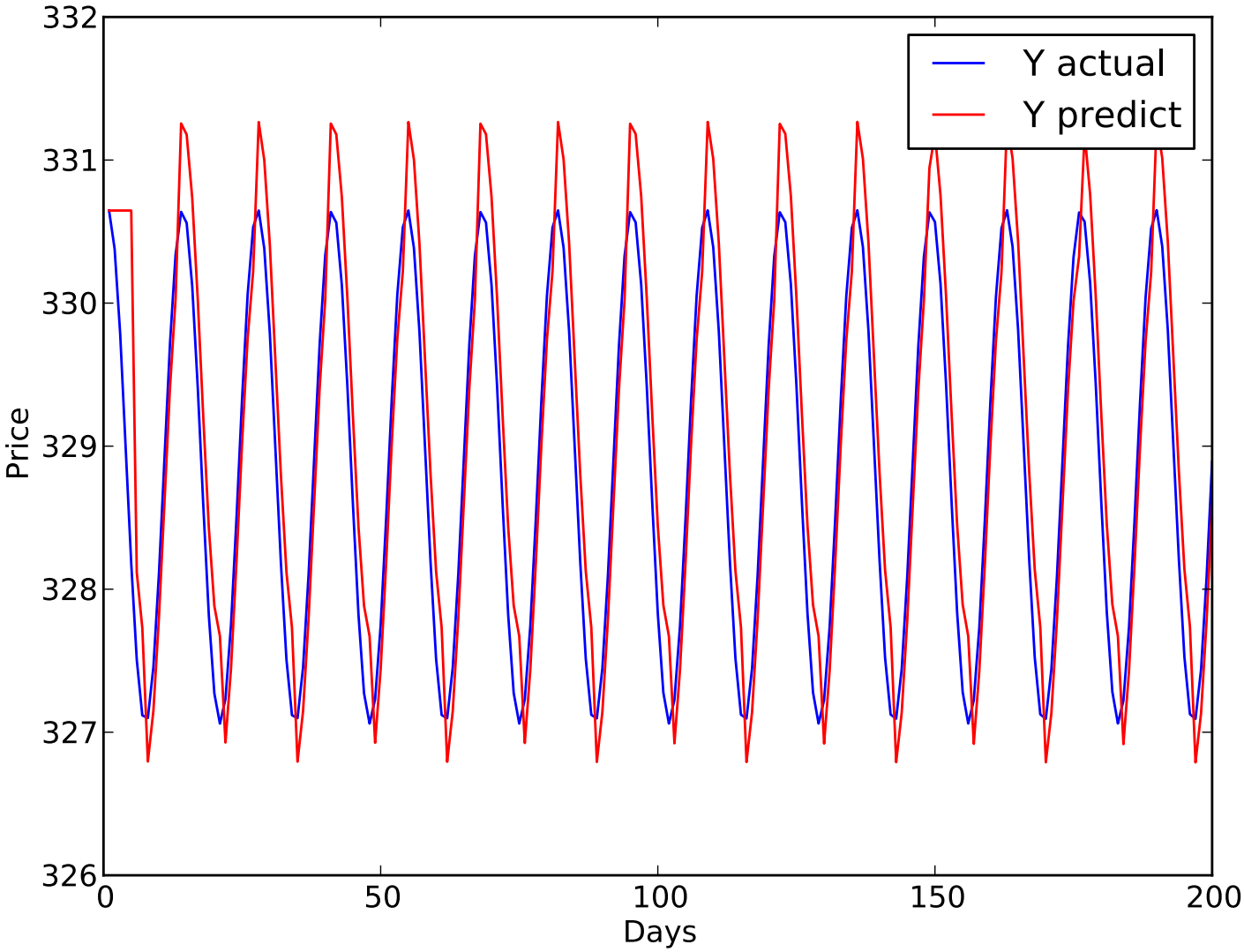
Project Name: Project 4

Student Name: Utkarsh Garg

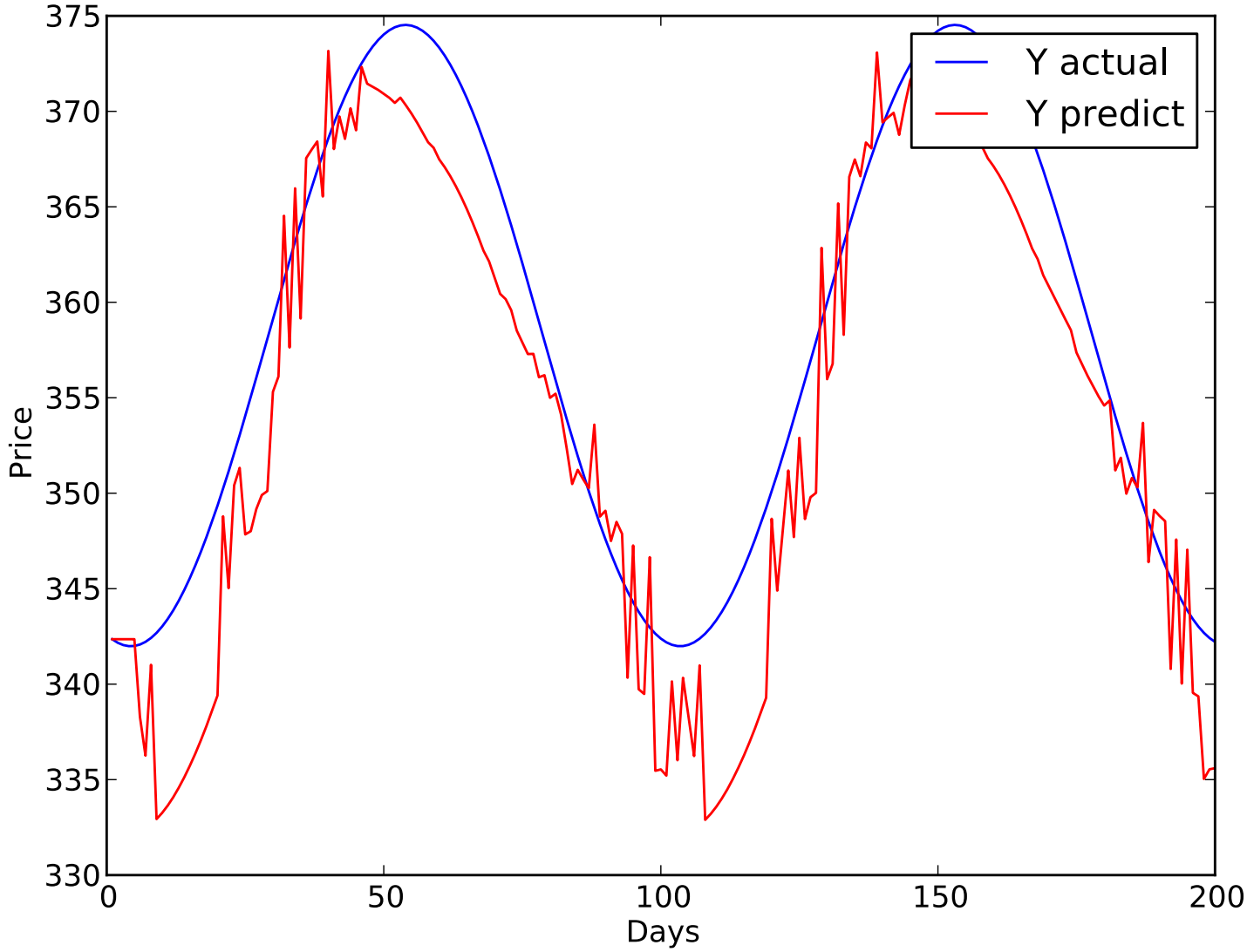
GTID: 902904045

My data sets – ML4T-292.cv and ML4T-328.cv (U+G = 28)

Time series plot of Yactual vs Ypredicted for the first 200 days

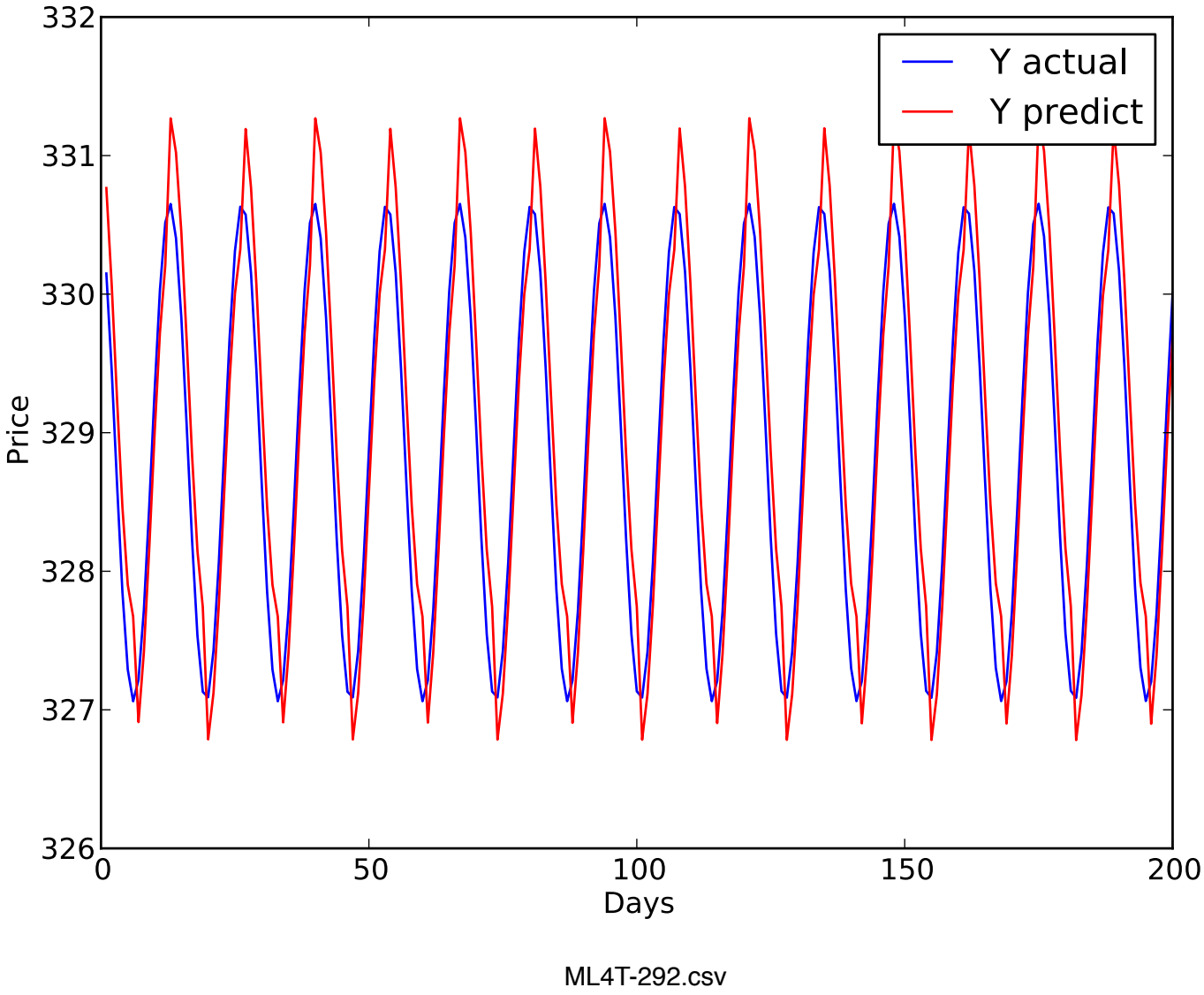


ML4T-292.csv

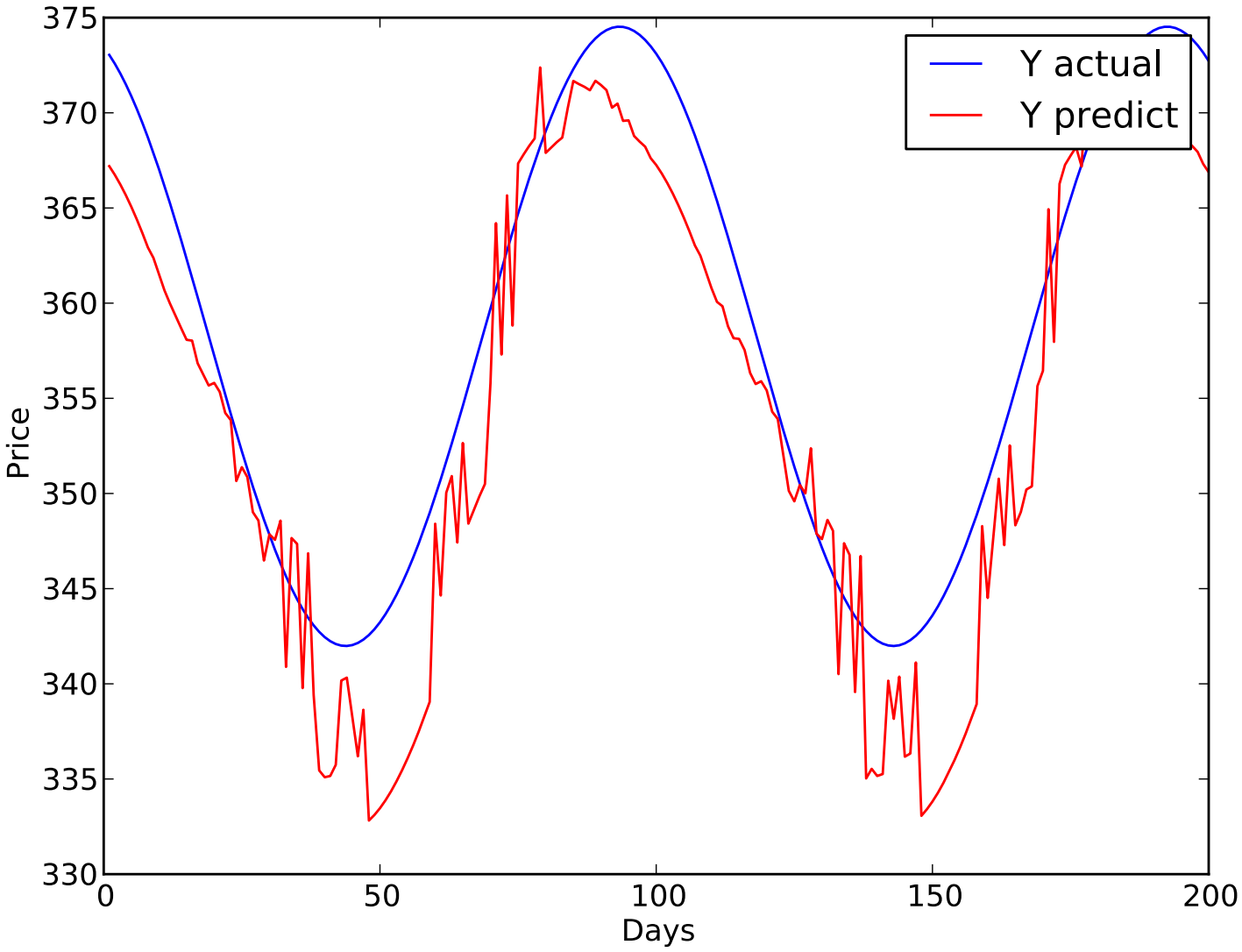

Y actual
Y predict
Price
Days
375
370
365
360
355
350
345
340
335
330
0
50
100
150
200

ML4T-328.csv

Time series plot of the last 200 days of Yactual and Ypredicted

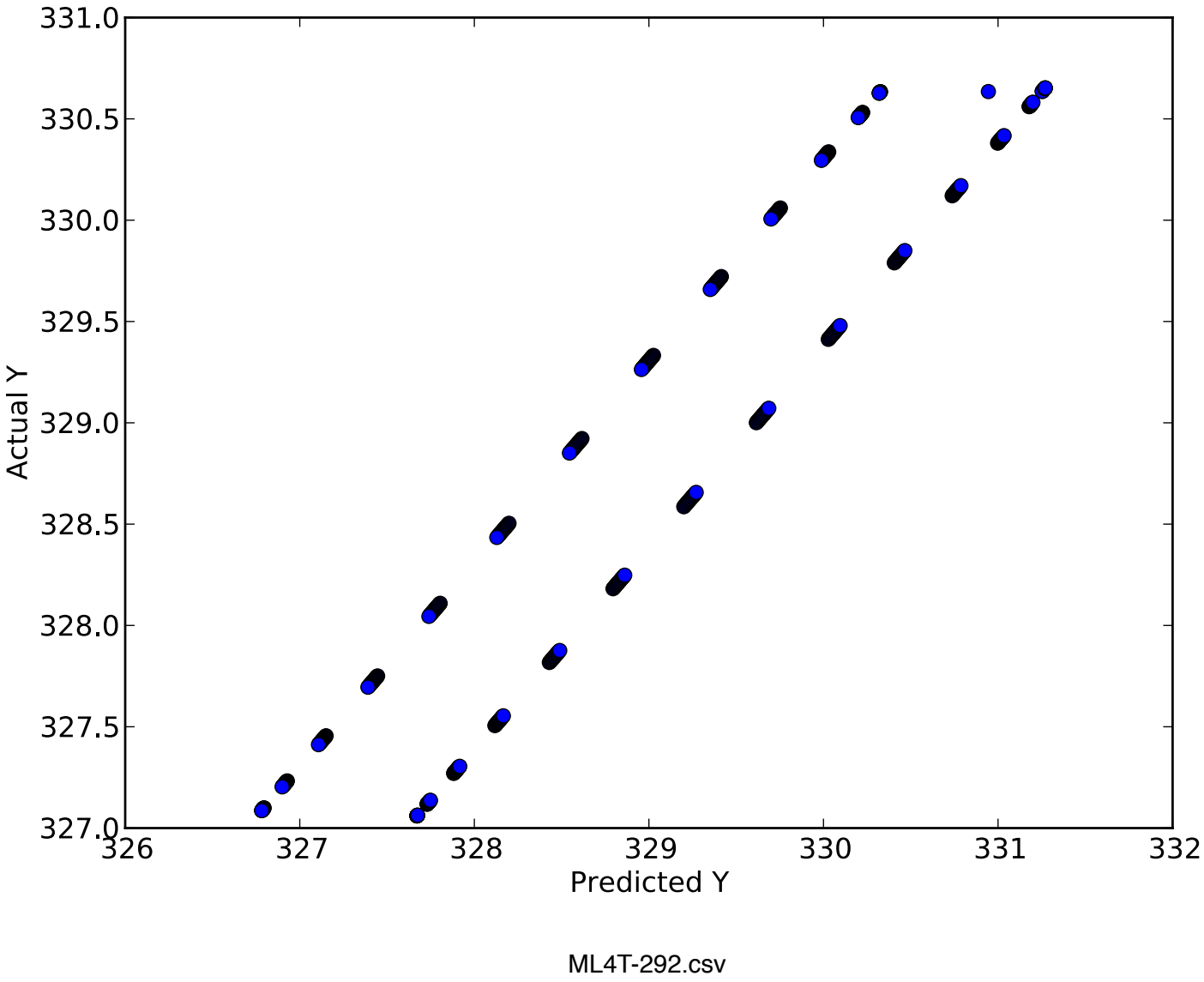


ML4T-292.csv

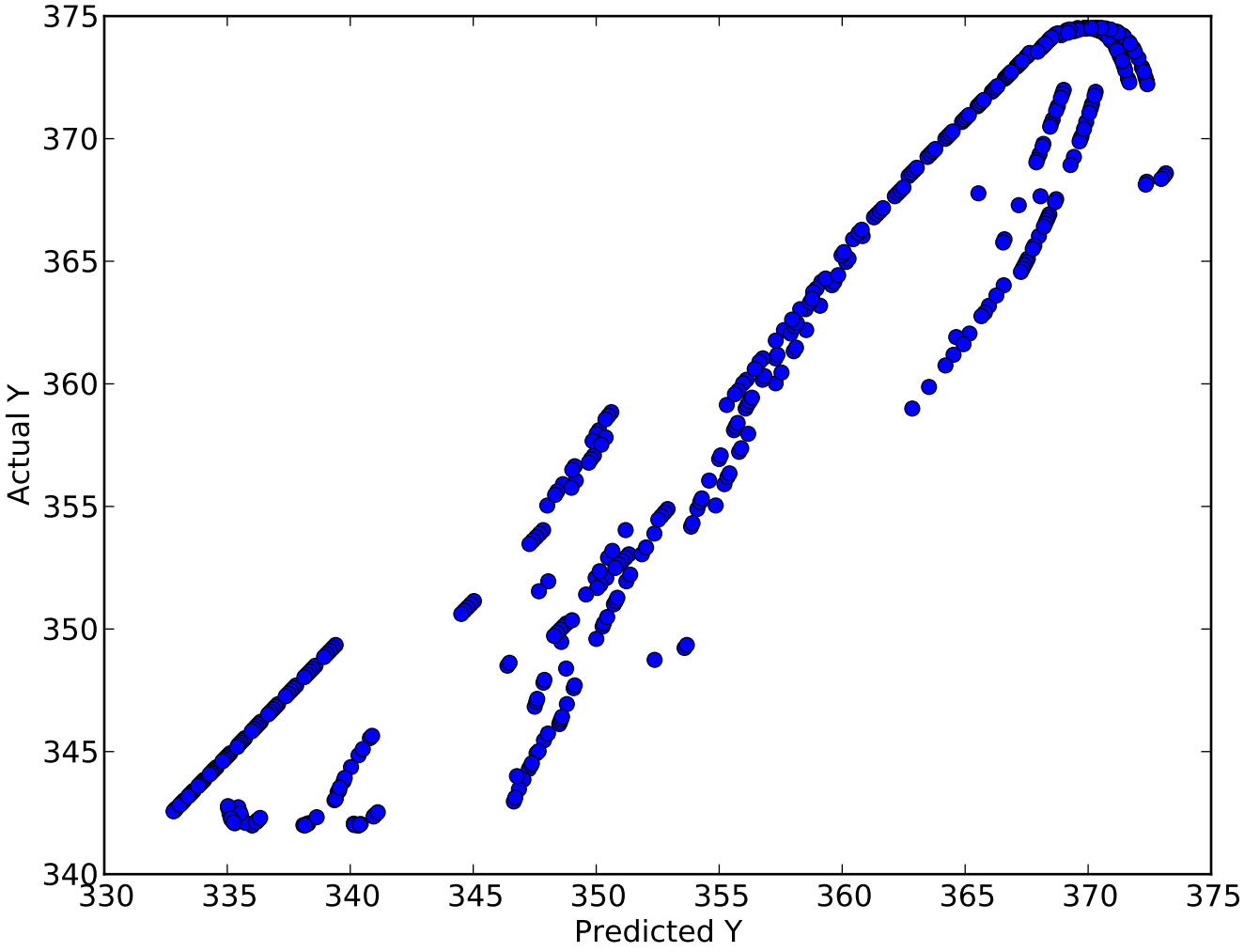

Y actual
Y predict
Price
Days
375
370
365
360
355
350
345
340
335
330
0
50
100
150
200

ML4T-328.csv

Scatterplots of Ypredict versus Yactual

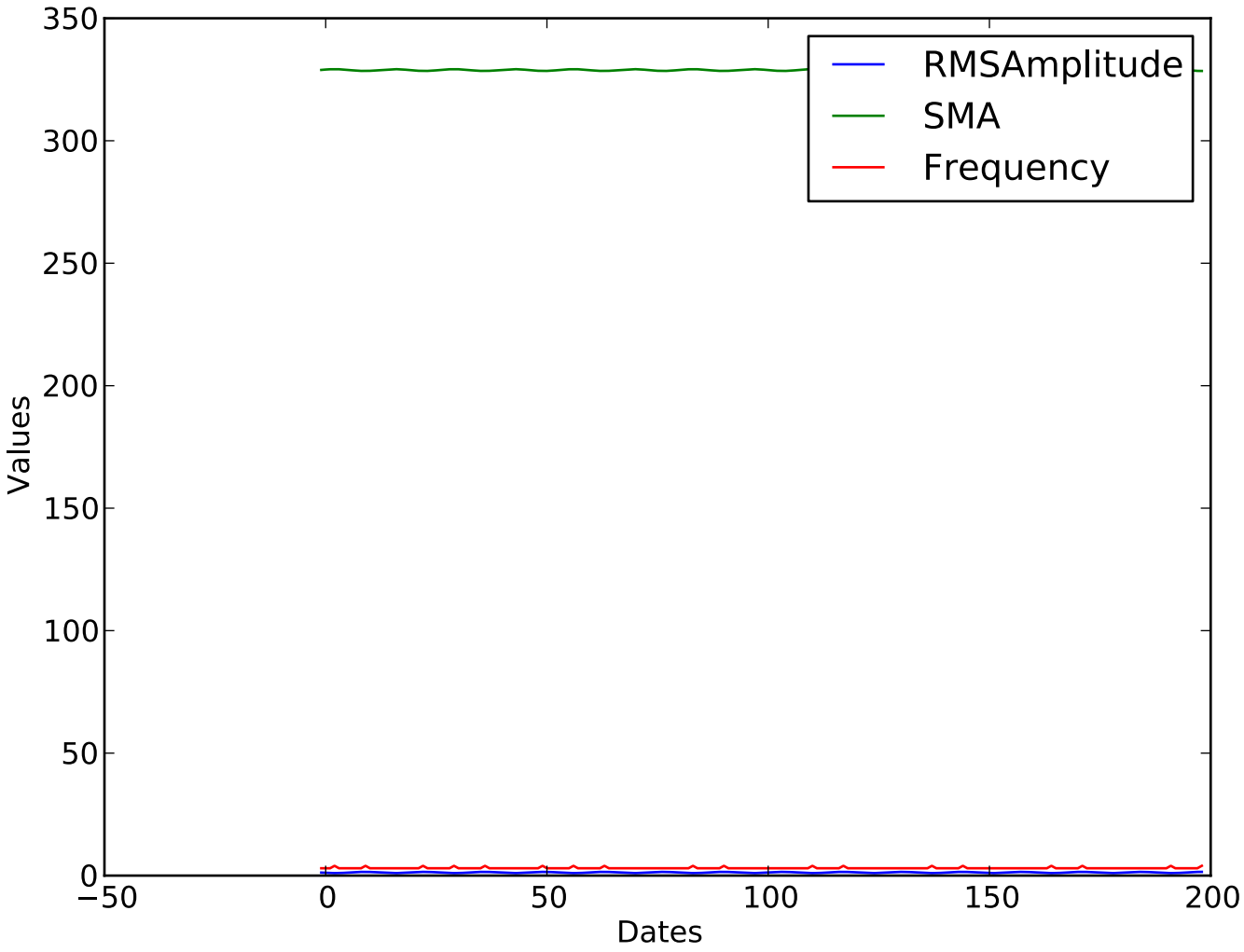


ML4T-292.csv

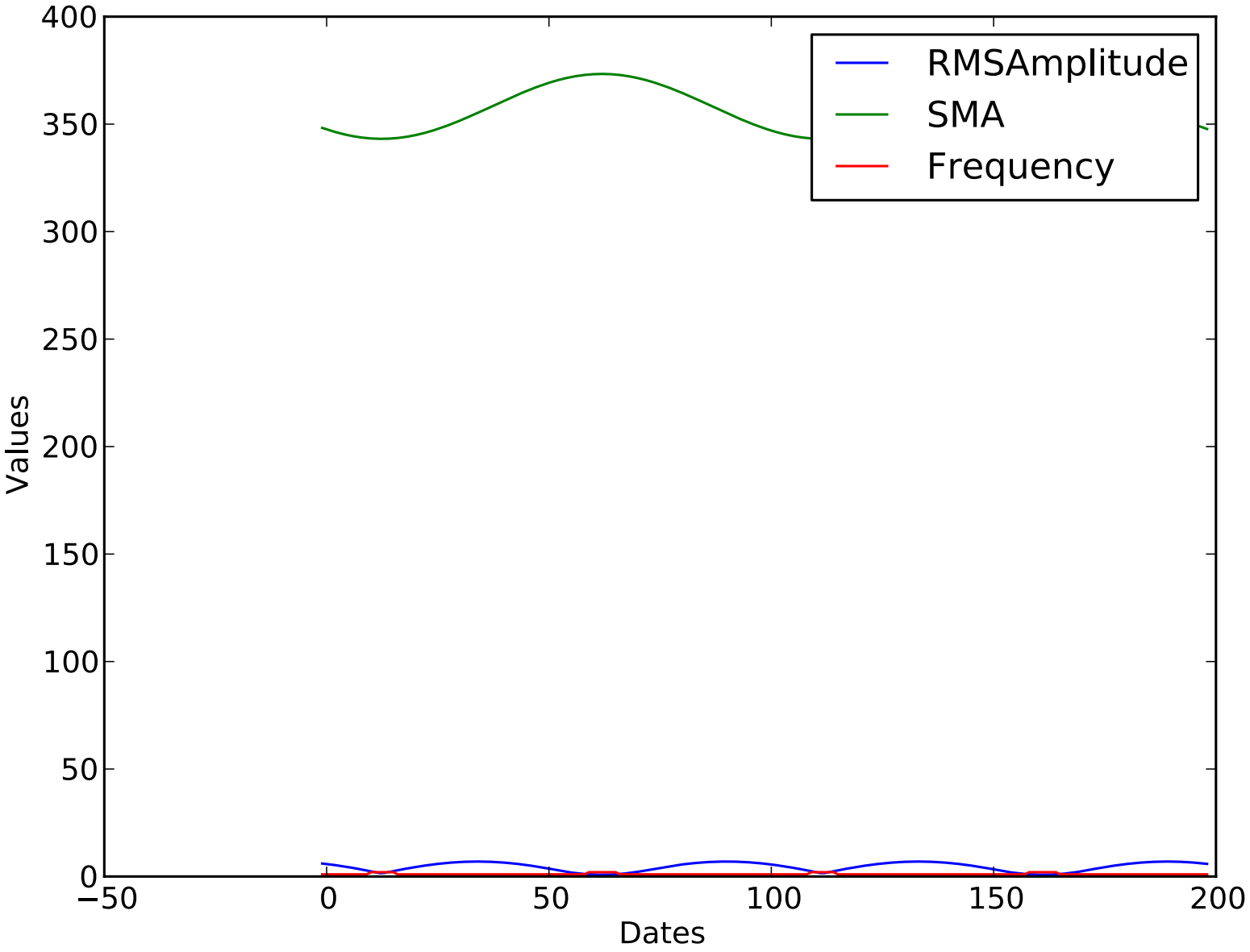

375
370
365
360
355
350
345
340
Actual Y
330
335
340
345
350
355
360
365
370
375
Predicted Y

ML4T-328.csv

Values of features for the first 200 days. Features are Root Mean Square Amplitude, Simple Moving Average and Frequency.



ML4T-292.csv



ML4T-328.csv

```
ML4T-292.csv:
    Correlation Coefficient: 0.941522373
    RMS Error: 0.494273962

ML4T-328.csv:
    Correlation Coefficient: 0.959590258
    RMS Error: 5.285795985
```

Learning Method and Indicators Used:

Indicators Used: I tried different combinations of RMS Amplitude, Peak to Peak Amplitude, Frequency, SMA, Peak Amplitude and Exponential Moving Average. Finally, used only 3, the frequency, the simple moving average and the rms amplitude.

Learning Methods: Used KNN with the three features mentioned above. My own KNN implementation was too slow, so used the sklearn KNN Regressor on the TA's recommendation.

Why the learning method works:

If you look at the graphs of r both my files, it is clear that the sine waves are very different and a hacky way for one of the data sets would be disastrous for the other. I think the sma is just a general indicator. Out of the various amplitudes, I think the rms amplitude is most relevant because the height of the wave is relevant from the X axis, which in this case is the mean. As for frequency, I think it helped because it could predict the shape of the curve formed by the data, whether it was spread out or increasing and decreasing rapidly.