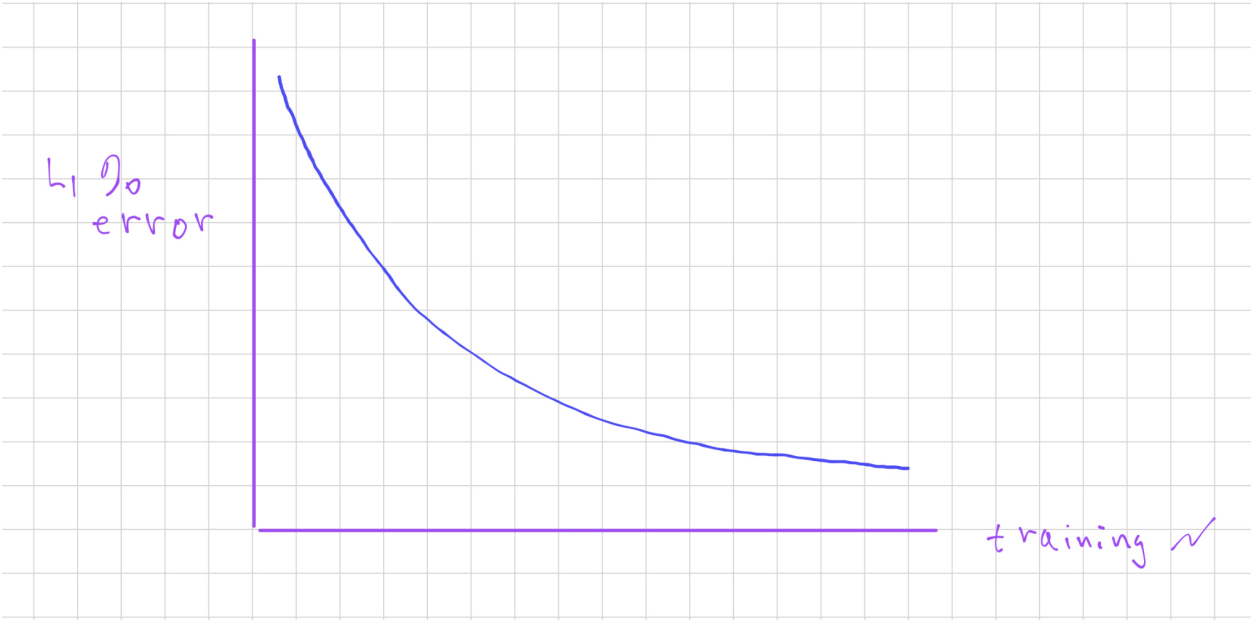
Machine Learning project 02

Using Pandas, Numpy, and Sciklearn build a linear regression model that takes features from the financial files provided and predicts the closing price of the Dow for the next day (data provided in Dow.csv) . The files provided are

File	Description
VIX.csv	The Cboe Options Exchange calculates a real-time index to show the expected level of price fluctuation in the S&P 500 Index options over the next 12 months. Officially called the Cboe Volatility Index and listed under the ticker symbol VIX, investors and analysts sometimes refer to it by its unofficial nickname: the fear index. (investopedia)
CornFutures.csv	Corn futures are **CFTC-regulated, exchange-traded contracts on the Chicago Board of Trade** (CBOT) and are one of the top 5 most-commonly traded commodity futures. Corn is the most widely grown crop across the United States, and corn futures contracts are the most active market in grains and oilseeds. (Charles Scwab)
CrudeOil.csv	Cost per barrel of oil
Ford.csv	Ford stock price per share
Gold.csv	Stock price of **Barrick Gold Corporation** . a mining company that produces gold and copper with 16 operating sites in 13 countries.[3] It is headquartered in Toronto, Ontario, Canada.[4] It has mining operations in Argentina, Canada, Chile, Côte d'Ivoire, Democratic Republic of the Congo, Dominican Republic, Mali, Papua New Guinea, Saudi Arabia, Tanzania, the United States and Zambia.[5] In 2019, it produced 5.5 million ounces of gold at all-in sustaining costs of $894/ounce[6] and 432 million pounds of copper at all-in sustaining costs of $2.52/pound.[7] As of December 31, 2019, the company had 71 million ounces of proven and probable gold reserves.[8] (wikipedia)
USsteel.csv	Stock price of US steel

For this assignment we will only use the closing price of each file which gives six features. Please note a couple of things: the dates may be ascending or descending, the date formats might be in YYYY-MM-DD or MMM DD, YYYY format. Some files have data gaps so you must make sure that you have a complete pairing of current day features paired with the next day Dow.

#1) Create a csv file of clean data that contains the six features with no data gaps.
#2) Read this file in with Pandas
#3) MaxMin normalize the data
#4) Write an algorithm that uses the 80/20 rule and produces a learning curve like the one shown



To help compare each others work lets have the features be in the following order:
[VIX, Ford, CrudeOil, Gold, CornFutures, USsteel].

Once trained your program should ask for the days closing numbers and provide a prediction for the next days Dow. I will follow the markets for three test days. The best algorithm will receive a prize (of inconsequential worth).

What is turned in:
1) the file that you compiled for training
2) The code file in proper format