Analysis Summary

Details

Data:

- BTC Bitfinex 2019-2023

Strategies:

-Exponential Moving Average

-Simple Moving Average - Relative Strength Index - Bollinger Bands

Models:(multivariate)

-Support Vector Classification (SVC)

-Random Forest Classifier (RFC)

-k Nearest Neighbor (kNN)

Analysis Process

Data Collection:

-BTC/USD 1hr collected from Bitfinex

- data download in csv format from CryptoDataDownloads.com

Data Cleaning:

-Raw Data was collected in REPO *Main/Data/RawData*

-files for cleaning and formatting data in REPO *Main/Data/RawData*

-data saved in csv format in REPO *Main/Data/FormattedData*

JupiterLab Notebook files:

-General layout of ML_Strategy files *Main*

*Data import, sliced and formatted

*Trading strategy conditions coded

*Strategy signals established for buy/sell actions

*Actual & Strategy returns calculated and plotted

*Data formatted into training and testing sets

- 10/90, *20/80 & 40/60 all trialed

*Data scaled, and resampled

—-------

*Model instantiated, fit with scaled or resampled data[1]

- target variable predicted

*Evaluation metrics run on model performance

- Classification report (primary) R2, and Mean Squared Error

*Model predicted returns calculated and graphed compared to actual returns

*Repeat of model runs with different parameters for optimization of model for best returns

[1] SVC & kNN performed best with resampled data, RFR performed best with scaled but not resampled data

Analysis Findings

Strategy 1: EMA Crossover

The EMA crossover strategy returns outperformed the actual returns, producing 25% greater cumulative return over the 2 year 6 month period (6 months was used as training data).

1- The SVC model enhanced this strategy performance to produce a 37% cumulative return, but only when the training period was 6 months. Model performance was highly biased with only 3 months training and overly fit with 9 months. The 6 month training produced a fairly balanced classification report. Within this dataset, the low recall on the negative (0-sell) class worked to its advantage producing a prediction biased toward (the recall of the positive 1-buy) class improving the cumulative returns in this particular dataset. However, when SVC buy/sell predictions were tested and visualized within a portfolio total, we see cumulative returns at 53%! Only, caveat, is all the buy sell action happened while the market was stable. Once increased volatility entered the market the model did not produce signals for that time period. This indicates SVC is not a good model to handle times of high market volatility.

2- The RFC had the most balanced classification report when not using the resampled data. The depth of the model parameter worked best at a depth=6. With depth=3, the model was not picking up on enough buy signals to produce a good return, depth=12 didn't improve the results from depth=6, but added computation time. The classification report was well balanced, with recall higher on the positive class, and well distributed across the time frame. The returns predictions produced an outcome slightly below that of the strategy alone, -1%.

3- The kNN results n_neighbors parameter was trialed from 3-50, with 3 producing a well balanced classification report with recall low on the sell, very high on the buy, this should have produced the best results of the 3 models. However, the distribution of the buy/sell signals was not well fit and the model stopped producing signals once market volatility began to increase. The portfolio results produce a -25% loss in this case.

Strategy 2: Simple Moving Average-Relative Strength Indicator-Bollinger Bands

The SMA-RI-BB strategy produced very poor results for this data and timeframe, because the RSI and Bollinger bands did not match with usual settings, they were offset never triggering the buy signal. Thus all signals in the strategy are sell, and hold. hold=33432, buy= 1789. With the first return being 1 when entering the market on day 1, the Strategy managed to return a positive number, but 89% below the actual return.

1-The SVC was computed with 3 kernels, two non-linear 'rbf' & 'poly', and one linear, 'clf'. With the poor data from our strategy, this model with all of its kernels only amplified the losses. The linear 'clf' kernel, stopped predicting any sell signals

The big question is could the models turn this poor performer into a profitable performer?
In general this model with all kernels did not do well, ‘clf’ produced the least negative total return.

2. The RFC had slightly better returns than the strategy by about 1%. The biased precision to hold(the negative class in this case =0) allowed this strategy to produce a 5% gain on portfolio returns.

3.The kNN wat the top performer with a more balanced recall of negative to positive class, but ALSO better placement of those sell and hold transactions. This key factor set it apart from the RFC for this strategy. This allowed the model to have a total portfolio return of 10%.

Conclusion:
Best Strategy: EMA

Best model for EMA strategy: SVC when in a stable market with general uptrends, RFC if accounting for volatility performance.

Best model for SMA-RI-BB strategy: kNN