DYF Strategy

Donglin, Yining, Francis

Agenda

1. Strategies Introduction
    a. Weighted Moving Average
    b. MLP
        i. Data Processing for MLP
        ii. MLP Model Training and Evaluation
    c. Combination
2. Performance Measurement

Weighted Moving Average (Week 1)

- **Weight Assignment:** We assigned increasing weights to the data, prioritizing recent market information. (e.g., Linear Function)
- **Calculating Rate of Change (ROC):** For each asset, we calculated the rate of change, which reflects the percentage change in price compared to the previous period. (Within a window of 20 days)
- **Applying Weights to ROC:** We then apply the assigned weights to these ROC values, highlighting the importance of recent changes.
- **Final WMA ROC Calculation:** Finally, We compute the Weighted Moving Average Rate of Change for each asset.

Data Processing for MLP (Week 3)

- **Data Segmentation**: We segmented the data into two groups - equal and non-equal columns based on initial values
- **Rate of Change Calculation:** We calculate the rate of change for each asset to measure stock price momentum.
- **Sliding Window:** We used a sliding window technique for dataset preparation. We created the Input and the corresponding output dataset in a window of 30 days
- **Data Augmentation (Week 5):** Add Gaussian noise to the raw data to get a more robust representation.
- **Data Filtering (Week 7):** Using changepoint detection to filter out data under regime shifts.

MLP Model Training and Evaluation

- **Model Architecture:** Our MLP model consists of three layers with 28, 64, and 16 nodes respectively, and a final output layer with one neuron.
- **Training Process:** We use the Adam optimizer and Mean Squared Error loss function for training over 100 epochs.
- **Evaluation:** Post-training, the model is evaluated with the test dataset using standard regression metrics
- **Final Output:** The model predicts the rate of change in stock prices.

Final Combined Ranking

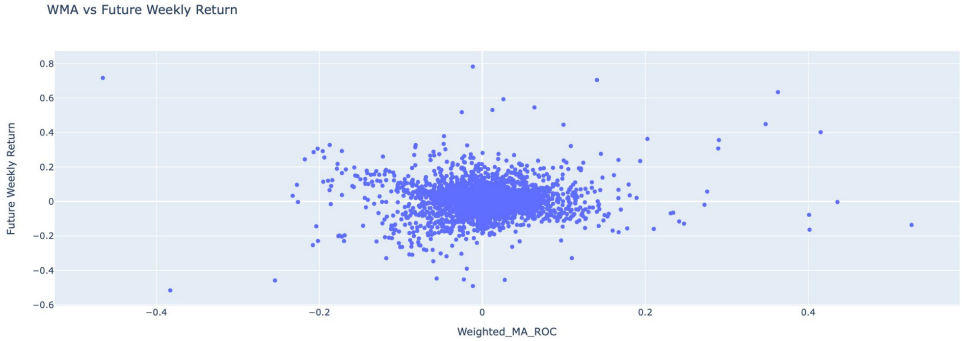
- **Normalization:** All Two metrics (ROC, MLP outputs) are normalized using the Z-score method, ensuring comparability across different scales.
- **Weight Assignment:** Priority on MLP based on their relative predictive power evaluation - relative lower weights on WMA

Ranking Prediction & Decision Making

- **SoftMax:** Map the final score to 5-class classification logits.
- **Quantile Threshold for Candidates:** Rank all the assets according to final combined score, only keep the top k1 assets and last k2 assets. (k1, k2 decided by estimators).
- **Weight Assignment:** Normalize the selected asset score to get the weight for decision.

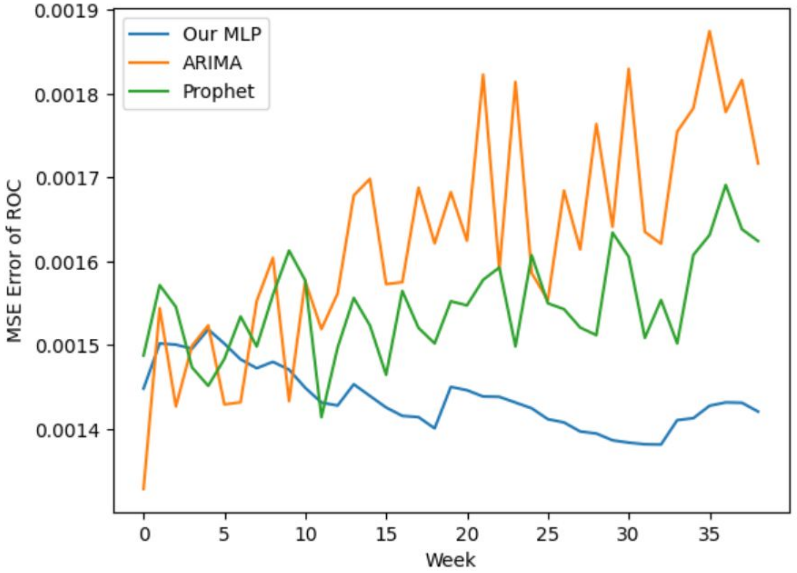
Performance Measurement - Weighted Moving Average

- **Backtesting with rolling in sample windows**
  - In-sample: 2018-01-01 to 2020-12-31
  - Validation: 2021-01-01 to 2022-12-31
  - Testing: 2023-01-01 onwards
  - Window size: 52; 1
- **Robustness Check: what we learned**
  - Training: Cluster around 0
  - Validation: Variability
  - Testing: Stable in unseen market conditions



Performance Measurement - MLP

- **Data Feeding Strategy**
  - Data Training: 52 weeks data
  - Testing: previous 4 weeks data to predict next week's stock returns
- **Comparison - Time-Series Forecasting**
  - ARIMA
  - Prophet
- **Model Accuracy**
  - Consistently low MSE indicates high predictive accuracy
  - Ability to learn complex non-linear patterns in the data
  - Strength in short-term forecasting
- **Model Stability**
  - Relatively stable MSE across weeks suggests that the model has learned a consistent pattern.
  - No significant spikes in error - model is not overfitting to noise in the data.



Insights

- **Regime Shifts**
  - When WMA/Linear Model fail
  - Possible Solution: CPD/Transfer Learning/SSL/DA/DG
- **Limited Data**
  - Unified Model: Heterogeneity (across section)
  - Data Augmentation
- **Model Selection**
  - Transformer: overfit
  - MLP: agile
  - Linear: explainable
- **Source (Unfinished)**
  - News/Twitter => Event Modeling (Event-driven shift)
  - Multimodality