Chapter 1: Meta-labeling

Guillermo Creus

HKUST - UPC

6th November 2020

Idea

Imagine you were dealing with a financial time series and you wanted to predict the **side** (long/short or pass) and **size** of your investment. If the explanatory variables of **side** are different, then meta-labeling is useful:

Primary Model (M1)

- Labels: $y_{\text{M1}} \in \{-1, 1\}$. A positive is understood as the model telling you to open a position.

Secondary Model (M2)

Its main function is to decide if a positive from the primary model is a true positive (i.e. whether to trade with the side of M1)

- Features: $\hat{y}_{\text{M1}}$ and same or different set of features than model 1.
- Labels: $y_{\text{M2}} = 1$ if $\hat{y}_{\text{M1}} = y_{\text{M1}},\ 0$ otherwise.
- **TP:** profitable investment, **FP:** losing investment, **TN:** M2 predicts correctly that M1 was wrong, **FN:** M2 predicts M1 was wrong when the latter was right.

Toy Project

The main idea of this project was to determine how meta-labeling works with synthetic data. In this case, 5 features have been used: $\mathbf{X}_{\cdot,1},\ \mathbf{X}_{\cdot,2}$ (M1) and $\mathbf{X}_{\cdot,3},\ \mathbf{X}_{\cdot,4},\ \mathbf{X}_{\cdot,5}$ (M2).

- $\mathbf{X}_{k,i} \sim N(\mu_i,\ \sigma^2)$
- $\omega_k = \textbf{sigmoid}\left(\alpha + \sum_{i=1}^{5} \mathbf{X}_{k,i} \cdot \beta_i + \epsilon_k\right)$ where $\epsilon_k \sim N(0,\ \sigma_\epsilon^2)$ and $\textbf{sigmoid}(z) = \frac{1}{1+e^{-z}}$

The labels are defined as:

$$y_k^{\text{M1}} = \begin{cases} -1 & \text{if } \omega_k < 0.5 \\ 1 & \text{otherwise} \end{cases}$$

Toy Project

Definition of Models

Primary Model

- **Features:** $\mathbf{X}_1, \ \mathbf{X}_2$
- **Labels:** y^{M1}
- **Model:** Neural Network (1 layer) for Binary Classification

Secondary Model

- **Features:** $\mathbf{X}_3, \ \mathbf{X}_4, \ \mathbf{X}_5, \ \hat{y}^{\text{M1}}$
- **Labels:** y^{M2}

$$y_k^{\text{M2}} = \begin{cases} 1 & \text{if } \hat{y}_k^{\text{M1}} = y^{\text{M1}} = -1 \text{ or } \hat{y}_k^{\text{M1}} = y^{\text{M1}} = 1 \\ 0 & \text{otherwise} \end{cases}$$

- **Model:** Neural Network (1 hidden layer - Leaky ReLU) for Binary Classification

Primary Model

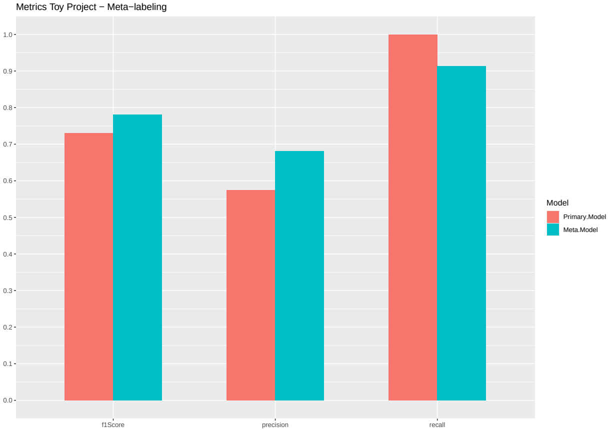
		Prediction outcome 1	Prediction outcome 0	Total
Actual value	**1**	TP 115	FN 0	115
	0	FP 85	TN 0	85
	Total	200	0	

Secondary Model

		Prediction outcome 1	Prediction outcome 0	Total
Actual value	**1**	TP 105	FN 10	115
	0	FP 49	TN 36	85
	Total	154	46	

Remember that Actual Value $= 1 \Rightarrow$ Trade and Actual Value $= 0 \Rightarrow$ Don't trade

Toy Project
Metrics



The high recall of the Primary Model brought many False Positives which have been corrected in the Meta Model (Primary + Secondary Model). That is why the F1-Score of the Meta Model is higher.

- **Horizontal barriers:** Dynamic levels that depend on the 10 day rolling volatility and the price at the first day considered. They can be symmetric or not.
- **Vertical barrier:** Set as a fixed time horizon. In this case, 10 days.

Figure: Symmetrical horizontal barriers

Figure: Non-symmetrical horizontal barriers (Side $= -1$)

Triple Barrier Method

Noting $t_{\text{hit}},\ t_{\text{start}}$ as the time the barriers start and get hit, respectively, the realized return can be computed as:

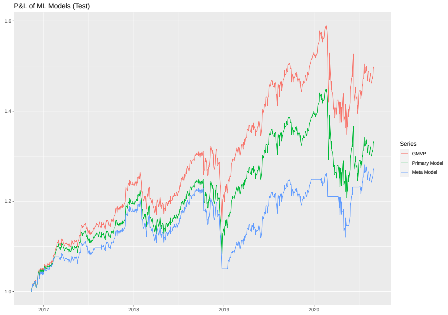
$$r_k = \left(\frac{p_{t_{\text{hit}}}}{p_{t_{\text{start}}}} - 1 \right) \cdot \textbf{signal}_{t_{\text{start}}}$$

Combining everything, the observations are labeled as:

$$y_k = \begin{cases} 1 & \text{if } r_k > 0 \\ 0 & \text{otherwise} \end{cases}$$

Results

Moving Average (MA) & ML Models



Model	Max. Drawdown (Test)	SR (Test)
B&H	15.20%	0.9682
M1	34.41%	-0.7467
MM	5.02%	0.4887

Model	Max. Drawdown (Test)	SR (Test)
B&H	15.20%	0.9682
M1	16.66%	0.7069
MM	14.57%	0.7302