Quiz 1 - Introduction to Systematic Trading Strategy with Machine Learning Algorithms

Name *

Matthew Durai

Email *

matthew.durai24@imperial.ac.uk

Labeling Methods

In fixed-time horizon labeling, why is using a fixed threshold τ for all samples problematic? 1 point

$$y_i = \begin{cases} -1 & \text{if } r_{t_{i,0},t_{i,0}+h} < -\tau \\ 0 & \text{if } |r_{t_{i,0},t_{i,0}+h}| \leq \tau \\ 1 & \text{if } r_{t_{i,0},t_{i,0}+h} > \tau \end{cases} \quad \text{with} \quad r_{t_{i,0},t_{i,0}+h} = \frac{p_{t_{i,0}+h}}{p_{t_{i,0}}} - 1$$

- (●) It ignores changes in volatility across regimes
- ( ) It creates label imbalance
- ( ) It introduces data leakage

In the triple barrier method, when is a trade labeled as –1? 1 point

- ○ The trade hits the profit-taking barrier
- ○ The trade reaches the time-out horizon without hitting any barrier
- ◉ The trade hits the stop-loss barrier

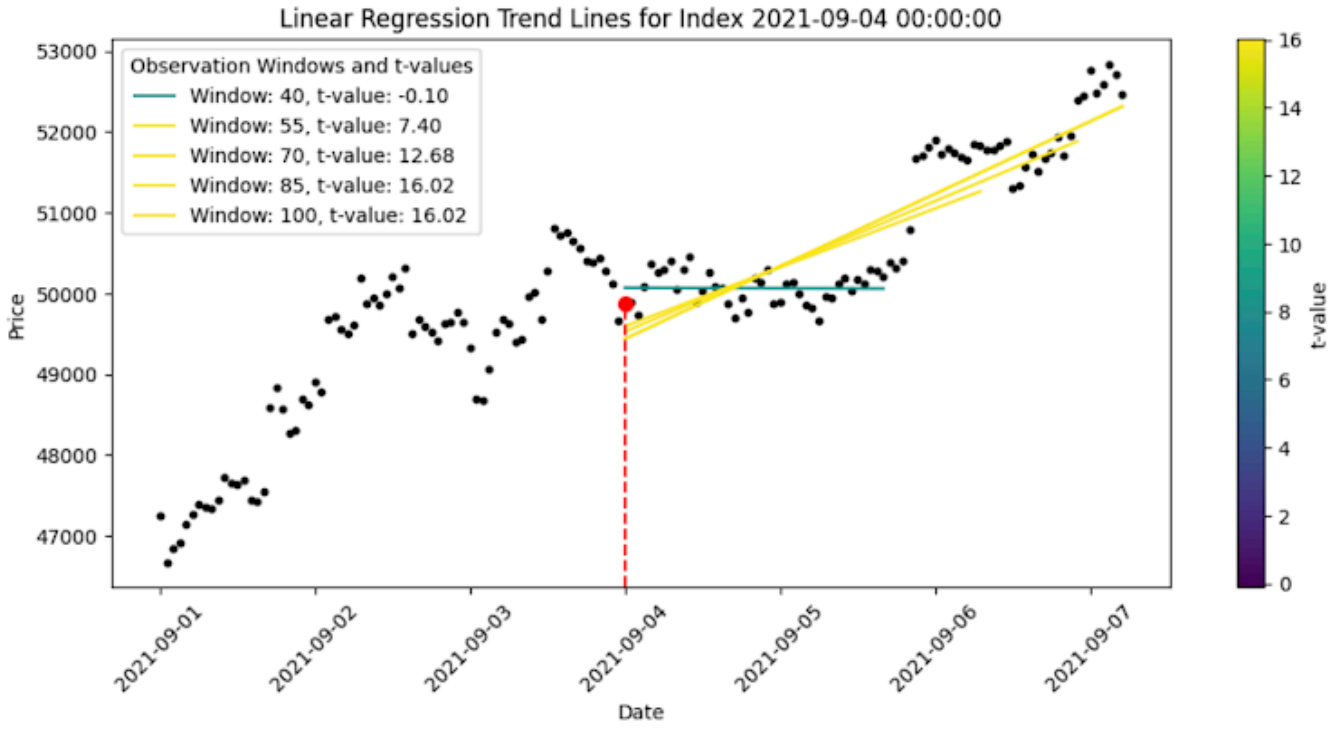
What is the key advantage of triple barrier labeling over fixed-time horizon labeling? 1 point

- ○ It doesn't require asset prices
- ◉ It captures path-dependence and risk management
- ○ It avoids overlapping samples

Trend scanning performs regression over a forward-looking window. What statistic does it use to determine if a trend exists? 1 point

- ○ R^2 of the fit
- ○ Standard deviation of returns
- ◉ t-statistic of the slope

In the figure below, several linear regression trend lines are fitted starting from September 4, 2021, using different forward-looking windows. Each window corresponds to a different t-statistic of the slope. Based on the trend scanning method, which observation window will be selected as the most informative? 1 point



- ○ Window: 40
- ○ Window: 70
- ◉ Window: 85

Given that the trend scanning method selects the window with the highest absolute t-statistic, what label will be assigned to the sample in the previous question ? 1 point

- ○ −1 (Downtrend)
- ○ 0 (No clear trend)
- ◉ +1 (Uptrend)

Evaluation Metrics

What is the interpretation of AUC (Area Under ROC Curve) in binary classification? 1 point

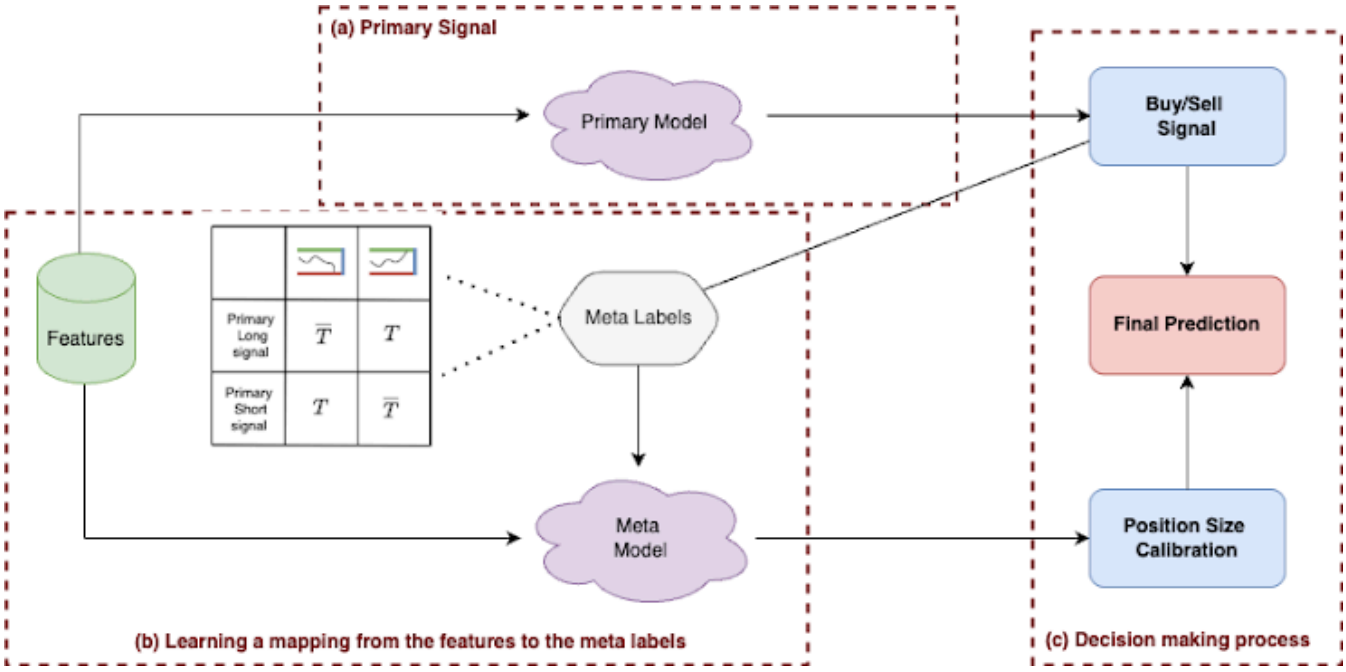
- ○ Probability the model outputs calibrated probabilities
- ◉ Probability a random positive is ranked above a random negative
- ○ Average value of true positives

Which metric penalizes confident wrong predictions the most? 1 point

- ○ Accuracy
- ○ F1 Score
- ◉ Log Loss

The Metamodel approach

In the Meta-Labeling framework depicted below, what is the main role of the Meta-Model? 1 point



- [ ] To generate new buy/sell signals directly from raw features
- [x] To predict whether the primary model's signal should be acted upon
- [ ] Rebalancing the portfolio across assets

A regime-related feature has high feature importance in the meta-model but low importance in the primary model. How do you explain this? 1 point

- ( ) The feature is noisy and should be removed from both models
- ( ) The meta-model is overfitting on features that are irrelevant for price prediction
- (•) The feature is predictive of when the primary model performs well, but not of asset returns directly

Questions ?

Any comments or questions you'd like to share about this quiz or the lecture content?

NIL