**ステップ1**

それぞれのデータには、ビットコインのdate, open, high, low, close, volumeの5分刻みデータが入っています。

trainのデータ期間は2013-04-01 00:30:00～2022-12-31 23:55:00、validのデータ期間は2023-01-01 00:00:00～2023-12-14 23:55:00。データがない時間帯が存在する可能性があります。

私は投資で儲けるために予測モデルを構築しようと考えています。

trainのデータを学習データ、validは検証データです。

closeの「極小点、極大点、その他」を予測するモデルを構築します。極小点の定義は「その点のlowが前後k個のlowより小さい、大きい」です。

極大点の定義は「その点のhighが前後k個のhighより小さい、大きい」です。

以下はデータの性質です。

・2020年以前はvolumeとhigh-low差に周期性は見られなかったが、2020年以降は1日周期性が見れた。

・closeの差分は両側指数～両側ベキの分布となった。

・価格の大きさ,volumeは昔と最近でスケールが全く異なる。

以下の特徴量とモデル、出力、を学習・検証方法もとに処理を実現するpythonコードをステップごとに出力してください。モデルはpytorchで構築してください。

【特徴量】

予測時点をふくむ過去1日分の以下の時系列

・close（予測時点のcloseで割ってスケールをそろえる）

・high（予測時点のcloseで割ってスケールをそろえる）

・low（予測時点のcloseで割ってスケールをそろえる）

・volume（予測時点で割ってスケールをそろえる）

・特徴量においてより良い候補や、修正すべき点があれば教えてください。

【モデル】

Transformerエンコーダー

【出力】

・極小値、極大値、その他（ただし極小値、極大値は周りと比べてどれくらい小さいか大きいかを踏まえてsoftlabelingによってラベル付けしてください）

・回帰による次ステップのcloseの現時点のcloseに対する割合（このタスクは補助タスク）

【損失】

・分類は極小値、極大値のprecisionが高くなるように損失関数を設計してください。

【学習・検証方法】

・2020年以降のデータだけでまずは学習させてください。

・テストデータで検証し混合行列,classification\_reportを出力するコードを実装してください。

・validデータはランダムに検証データとテストデータに分けしてください。

・エポックごとにいくつかのの検証データに対して、Transformer層後と全結合層後の出力ベクトルをヒートマップで可視化してください。

・エポックごとに検証データに対してclassification\_reportを出力してください

**ステップ2**

「前後 k個の high/low」についてはk=6としてください

5分刻みで欠損するタイムスタンプがある場合、線形補間してください

ソフトラベリングの詳細については具体的な基準はありません。候補があれば教えてください

Transformerのハイパーパラメータについては、ひとまず4層/6ヘッド/128次元にしてください。

バッチサイズ、学習率、エポック数は任せます。ただし学習率はタイムスケジュールしてください

valid データを検証用とテスト用に分ける際の比率は5:5にしてください

特に追加したい特徴量、入れたくない特徴量はありません

極小、極大と予測する際の閾値を0.25, 0,5, 0.75と変化させて各指標を出してください。

混同行列や classification\_report に加えて、ROC-AUC, PR-AUC,F1などできる限り多くの指標を出力してください。

極小、極大と予測された際、それぞれにおいて将来のウィンドウにおける最安値と最高値の、現在の終値からの変動率（%）も評価指標として出力してください

**ステップ3**

ステップ1のコード→詳しい説明→ステップ2のコード→詳しい説明という流れで出力してください

また、横軸エポック数でtrainデータとvalidデータに対する全体平均loss,　分類の平均loss, 回帰の平均lossをライブプロットするようにしてください