



AI WorkShop

金融データを扱った機械学習の演習

ファイナンシャル機械学習 第9章 交差検証法によるハイパーパラメータの調整

2022/03/25

土田晃司

金融データを扱った機械学習の演習

ファイナンシャル機械学習 第9章 評価指標とハイパーパラメータの調整

目次

1. K ~交差検証法
2. グリットサーチ交差検証法
3. ランダムサーチ交差検証法
4. ベイズ最適化サーチ交差検証法
5. 対数一様分布
6. 評価指標とハイパーパラメータの調整 メタラベリング
7. 評価指標とハイパーパラメータの調整 それ以外

1. 交差検証法

k-交差検証法：データセットを分割して学習を繰り返す

金融の取引では、連続した時系列を扱うので、通常の交差検証法では、オーバーフィッティングしやすい。

7章で提案された、テストデータの前後と学習データの手前を間引く、パージ・エンバゴしたk-分割交差検証法を用いる

メタラベリングの場合(0,1の予測)は、f1でモデルを評価する

それ以外の場合は、負の対数損失を評価する

2. グリットサーチ交差検証法

グリットサーチ：与えられたハイパーパラメータの値に対して、すべてのパターンを検証して結果が良かったパラメータを抽出する。パラメータの範囲を決める



3. ランダムサーチ交差検証法

ランダムサーチ：与えられたハイパーパラメータの値に対して、ランダムのパターンを検証して結果が良かったパラメータを抽出する。想定の時間内に必ず処理が終わる。パラメータの組み合わせを検討する必要がある



4. ベイズ最適化サーチ交差検証法

ベイズ最適化サーチ：与えられたハイパーパラメータの値に対して、どのパラメータが精度が出やすいかをベイズ推定を元に自動で判断してパターンを検証して結果が良かったパラメータを抽出する。

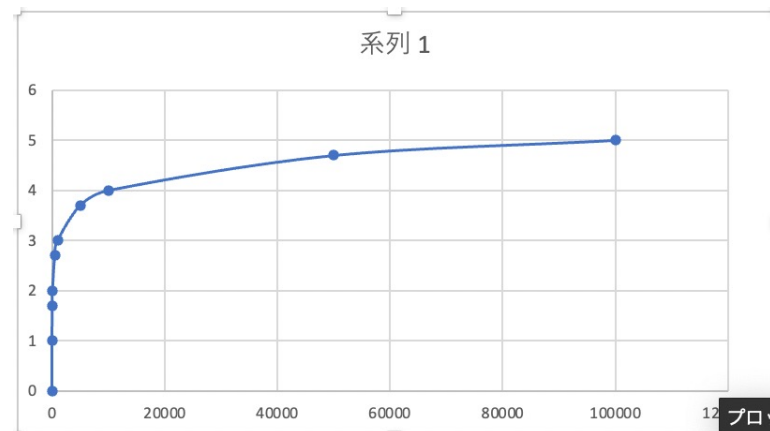


5. 対数一様分布

パラメータの実現可能性領域を探索するための方法

多くのハイパーパラメータは、0 から100の範囲で乱数で生成して利用するが多いが、0.001から1の間の値と、1から100までの値の増加量を同じにすると、計算量だけが増え非効率になる。

対数を使って値を生成することで、パラメータの値が小さい場合に増加量を小さくして、値が大きくなるにつれて増加量を大きくしていくことで効率的にハイパーパラメータのチューニングができる。(一様分布の確立密度変数を対数で)



6. 評価指標とハイパーパラメータの調整 メタラベリング

混同行列：二値分類（正事例と負事例の予測）

- ▶ Precision(適合率)
- ▶ Recall (再現率)
- ▶ Accuracy(正解率)
- ▶ F値(適合率と再現率の調和平均)

精度と再現率両方が良い結果のパラメータを見つけることで、取引の結果が安定する

偽陽性(FN)	真陰性(TN)
真陽性(TP)	偽陽性(FP)

7. 評価指標とハイパーパラメータの調整 それ以外

Accuracy(正解率)

確率の高い誤った予測と確率の低い誤った予測を同等に扱う
低い確率の大当たりと高い確率の損失を同等に扱う

=> 売買損益に着目していない => 金融取引では損益が重要

No_log_loss(クロスエントロピー誤差)

正解から離れるほど、誤差が急激に大きくなる特徴がある
PLに関するパラメータに着目した評価ができる

▶ <https://cognicull.com/ja/evb1pj5y>

クロスエントロピー誤差の値を評価することが金融取引では重要

演習

- ▶ 人工データの実験を行う

- ▶ Colaboratory

<https://colab.research.google.com/notebooks/intro.ipynb?hl=ja#>

- ▶ Github

<https://github.com/ktsuchida11/AIWorkshop202203>

- ▶ 参考になるコード

<https://github.com/boyboi86/AFML>