

アイスクリームの年間支出を予測する

神戸電子専門学校

情報処理学科

WebエンジニアIIコース

大山竜生

課題概要

金沢市の月平均気温とアイスクリーム年間支出金額から
アイスクリームの年間支出金額を予測した。

実行環境

- Google Colaboratory
- Python3.6
- Pandas
- Numpy
- Matplot
- statsmodels

データセット

earnings :

- アイスクリームの年間支出金額

temperature :

- dateの月平均気温

金沢市

2014/1/1~2019/12/1

までのデータ

```
[9] row_data.head()
```



	date	earnings	temperature
0	2014/1/1	396.0	3.9
1	2014/2/1	309.0	3.5
2	2014/3/1	447.0	7.4
3	2014/4/1	520.0	12.4
4	2014/5/1	803.0	18.0

データの自然処理(1)

1. データセットの型を整える
2. データを4つの成分に分ける
 - 周期成分(今回のデータだと12ヶ月周期)
 - トレンド(全体の傾向)
 - 外因性(自己相関では説明できない動き。今回は月ごとの平均気温のみを採用)
 - ホワイトノイズ

データの自然処理(2)

3. データからトレンド成分を抜く
4. トレーニングデータとテストデータに分割

モデル推定

statsmodelsのSARIMAXを使ってモデルを推定する。

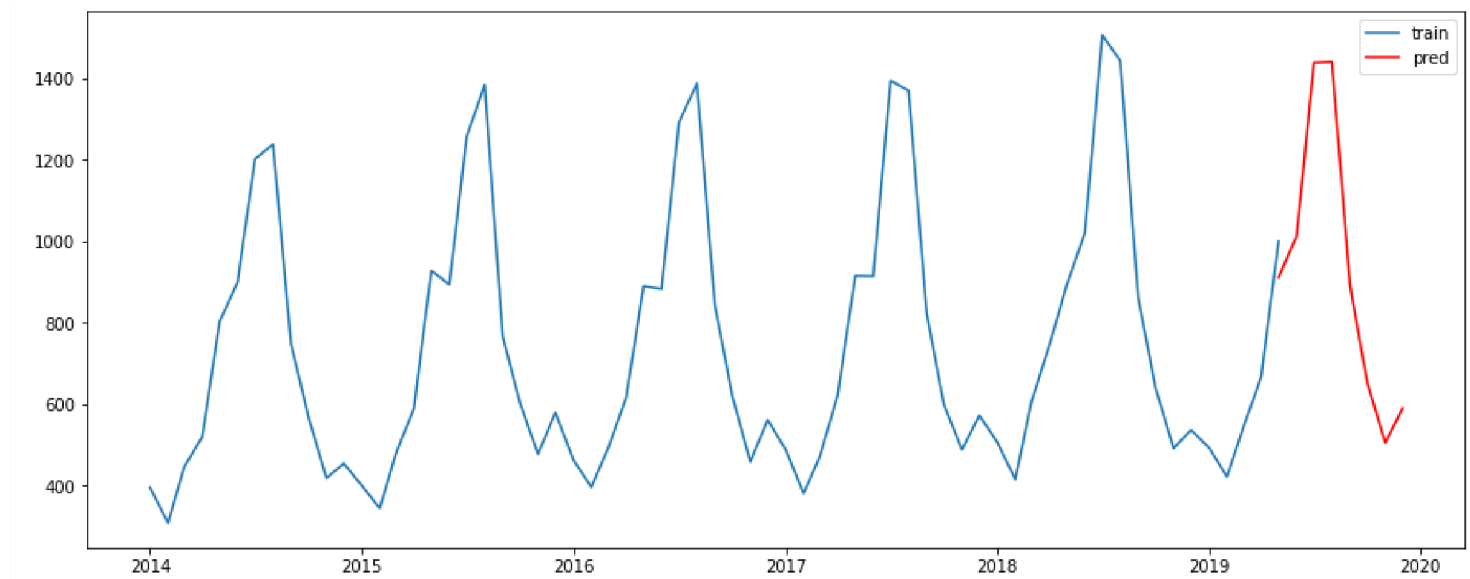
その時総当たりでAIC(統計モデルの良さを評価する指標)が最小となる次数を探す

モデルの評価

Ljungbox検定を用いて評価する。

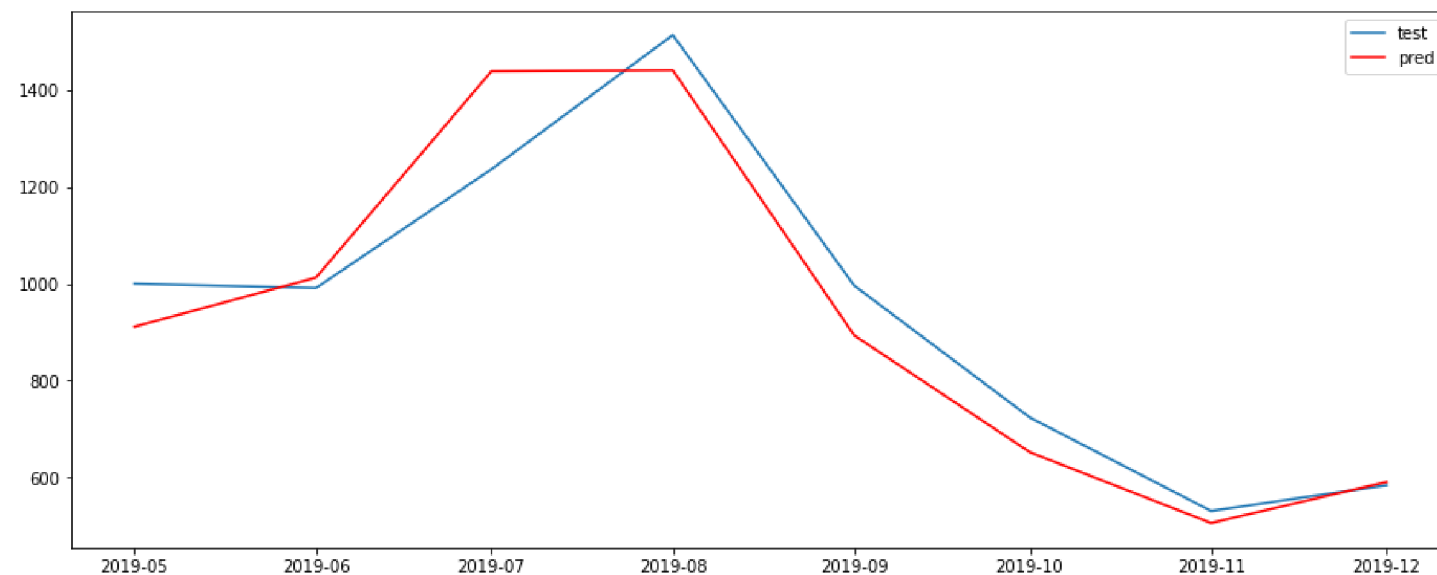
ホワイトノイズに有意な自己相関(時系列上にある異なる点同士の相関のこと)が一つもない状態になっているか検証できる。

作成したモデル で年間支出金額 を予測(1)



作成したモデル で年間支出金額 を予測(2)

正解データと合わせて図示



結論

それっぽい予測を行うことはできた。

時系列分析のことについて少し調べながら予測したが
理解できている気がしない。