

# Home Credit Default Risk

宮坂

# 結果

0.79119から始めて、  
0.79496。

その後は少しずつ  
下がっていききました...

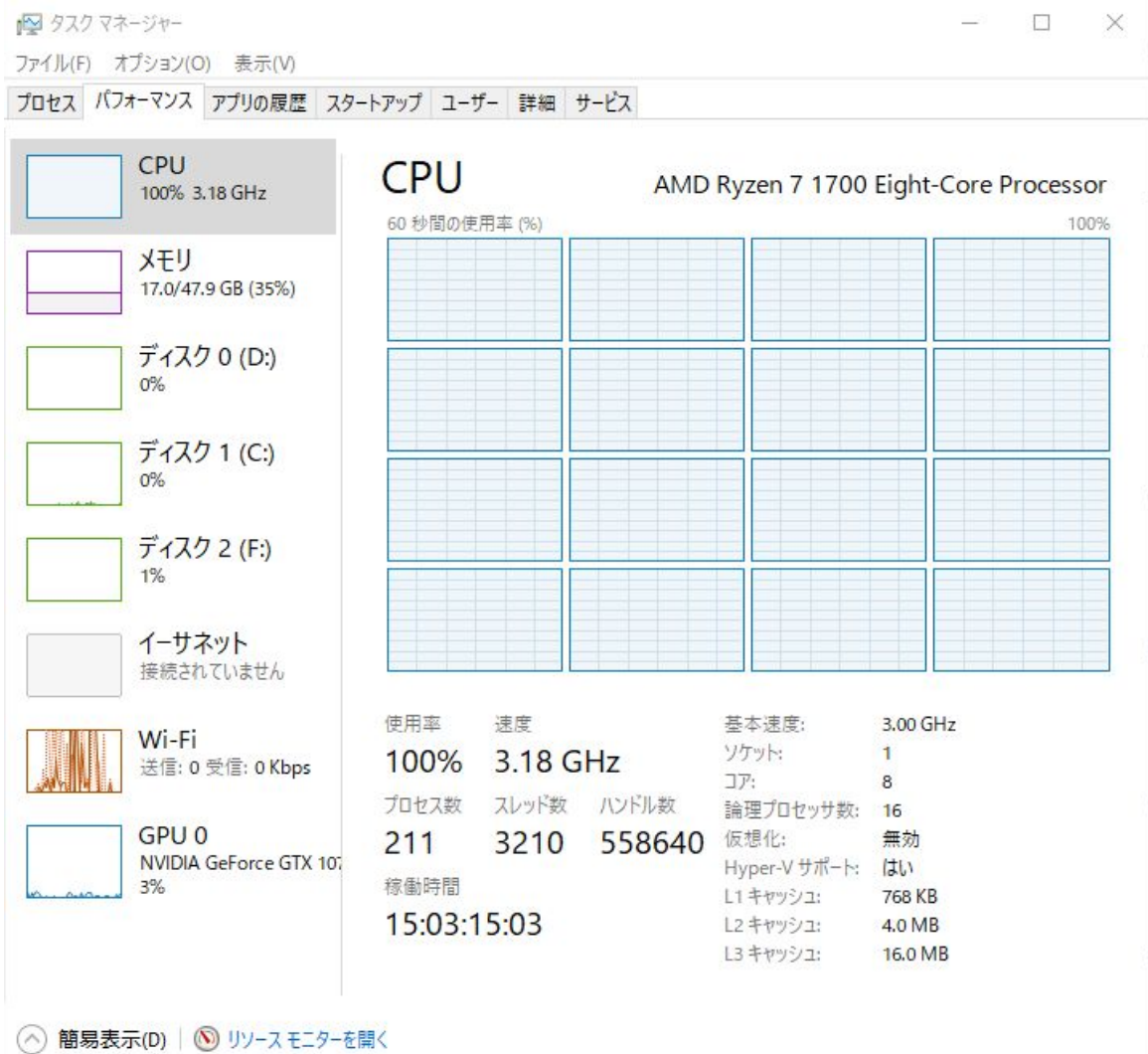
<a href="#">submission_0826_3.csv</a> 5 days ago by <a href="#">Lucky</a> <a href="#">add submission details</a>	0.78975	<input type="checkbox"/>
<a href="#">submission_0826_2.csv</a> 5 days ago by <a href="#">Lucky</a> <a href="#">add submission details</a>	0.79475	<input type="checkbox"/>
<a href="#">submission_0826.csv</a> 5 days ago by <a href="#">Lucky</a> <a href="#">add submission details</a>	0.79475	<input type="checkbox"/>
<a href="#">submission_0825_4_4.csv</a> 5 days ago by <a href="#">Lucky</a> <a href="#">add submission details</a>	0.79336	<input type="checkbox"/>
<a href="#">submission_0825_4.csv</a> 5 days ago by <a href="#">Lucky</a> <a href="#">add submission details</a>	0.79496	<input type="checkbox"/>
<a href="#">submission_0825_2.csv</a> 5 days ago by <a href="#">Lucky</a> <a href="#">add submission details</a>	0.79162	<input type="checkbox"/>
<a href="#">submission_0825.csv</a> 6 days ago by <a href="#">Lucky</a> <a href="#">add submission details</a>	0.79119	<input type="checkbox"/>

# 環境

ローカルPC

前回からSSD(m.2)追加

ファイル書きが3倍速くなった



# まずはデータを理解

application\_{train | test}.csv: 1行=1ローン

bureau.csv: クライアントの以前のクレジット

bureau\_balance.csv: クレジットの月間残高

POS\_CASH\_balance.csv: 以前持っていたPOSと現金ローン

credit\_card\_balance.csv: 以前のクレジットの月間残高

previous\_application.csv: 以前のアプリケーション

installments\_payments.csv: ホームクレジットの以前に支払われたクレジットの返済履歴。行なわれた支払いごとに1行プラス、不在支払いのために1行。1行は、1回の分割払いまたは1回の分割払いに相当し、前回のホームクレジットクレジットの1回の支払いに相当します。

HomeCredit\_columns\_description.csv

## 追加した特徴量(一部)

- ・EXT\_SOURCEのmean,\*,/,+,-,std,norm
- ・各データのSK\_ID\_CURRでgroupbyしたmin,max,size,mean,var,sum
- ・MONTHS\_BALANCEを1年単位で区切ってそれらのgroupbyした...
- ・一部の特徴をDAYS\_EMPLOYEDで/

# feature importance

列数は約12000

全列ではメモリが足りなかったなので先頭  
10万行で学習

feature importance が0<の列で全体を  
学習

fold数を5~30で試したが変わらず

