

# DIVE INTO CODE

## 卒業課題

～再帰型ニューラルネットを用いた  
脳MRIデータからの年齢予測～

2020年 4月期

機械学習エンジニアコース

齋藤 拓実



# 自己紹介 / 入校動機

< 略歴 >

2012年 大学医学部卒業

脳神経外科に入局、7年間臨床に従事

→ 「**最善を尽くせていない**」の思いから退職

2020年 DIC入校

動機：「**医療の経験を、機械学習を  
手段として価値に還元する**」

価値 … 業務効率化、疾病予防

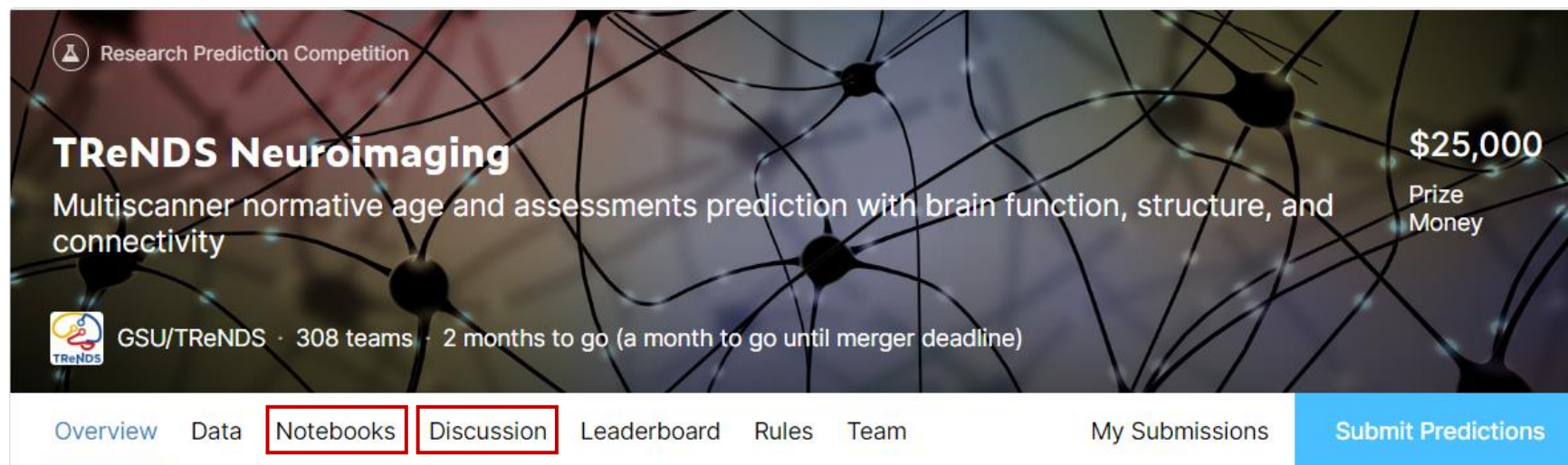
目標：

Engineer  $\leq$  DataScientist  
(手段)  $\leq$  (応用)



# 課題設定

## < Kaggle >



\* 2020/06/29 終了

## < Kaggleとは >

- ・ データ分析コンペティションのプラットフォーム
- ・ 社会問題解決のため各機関等が開催
- ・ notebooks / discussion上での情報共有も可能



# TReNDS Neuroimaging

## < 概要 >

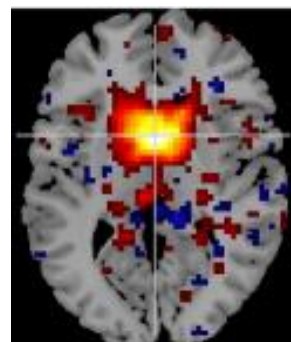
脳MRIデータ（数値・画像）から年齢と4変数を予測（回帰問題）

table)

Id	IC_01	IC_07
10001	0.006070	0.014466
10002	0.009087	0.009291
10003	0.008151	0.014684
10004	0.004675	0.000957
10005	-0.000398	0.006878

image)

&



```
Id, Predicted
10003_age, 0
10003_domain1_var1, 50.0
10003_domain1_var2, 50.0
10003_domain2_var1, 50.0
10003_domain2_var2, 50.0
etc.
```

背景：脳年齢と実年齢の乖離は、認知症等の神経疾患と関連がある

## < 選定理由 >

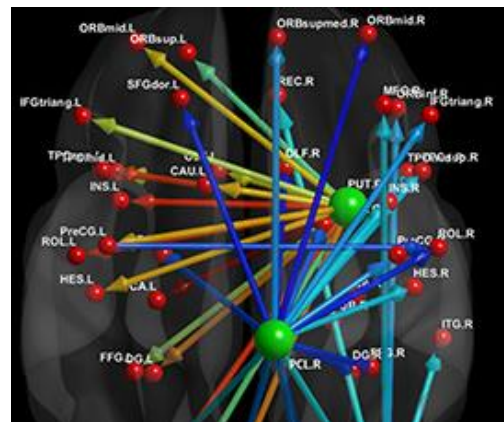
- ・プロダクトの作成 ≒ データ分析技術の習熟
- ・データ取得が容易、スコアで評価が可能
- ・自分の経歴に関連する領域

# Table Data

< 安静時の脳MRI >

相関係数

... 血流 (機能)

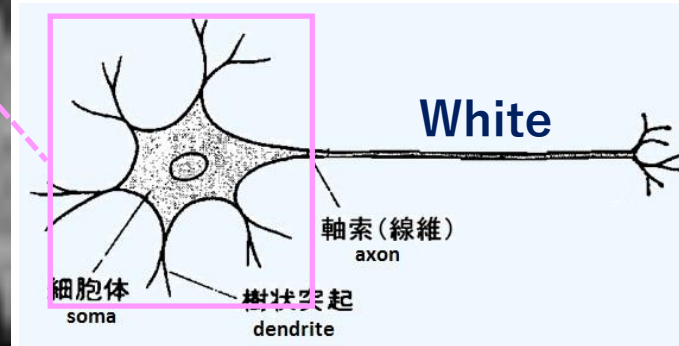


Id	SCN(53)_vs_SCN(69)	SCN(98)_vs_SCN(69)
10001	0.368580	0.166876
10002	0.151696	-0.024819
10003	0.343415	0.109974
10004	0.132793	0.258255
10005	0.291921	0.251254

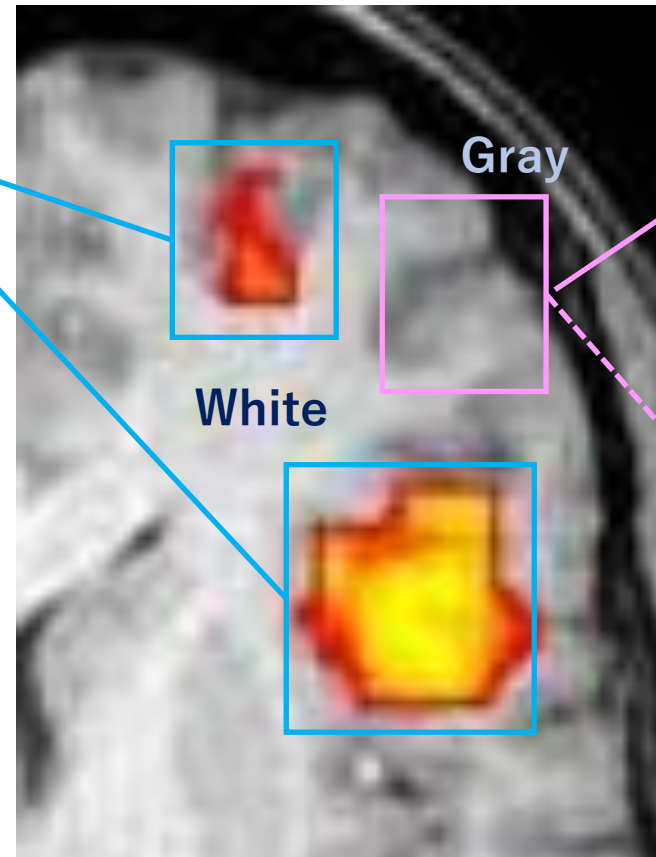
信号強度

... 厚み (形態)

Gray



White

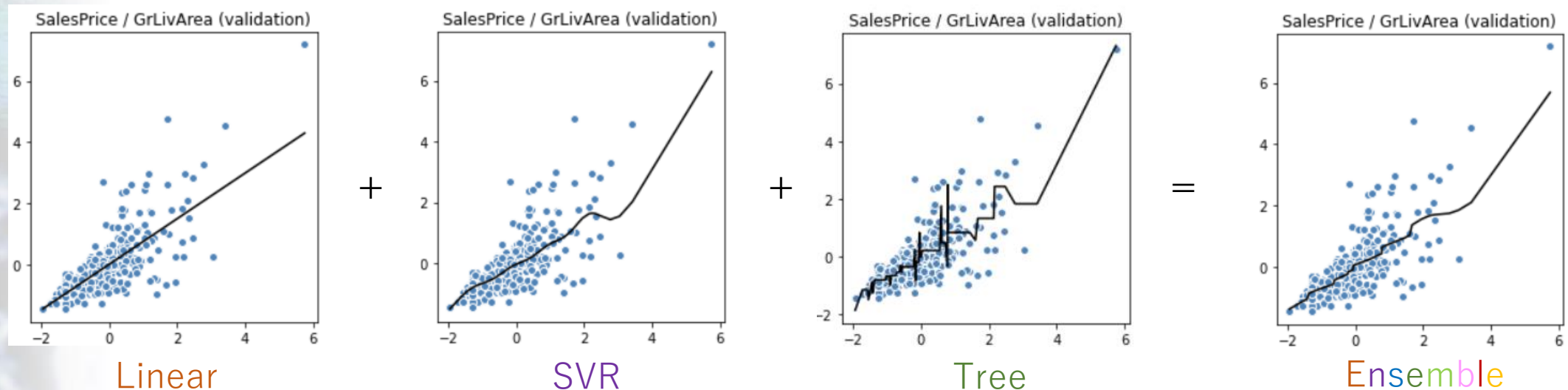


Id	IC_01	IC_07
10001	0.006070	0.014466
10002	0.009087	0.009291
10003	0.008151	0.014684
10004	0.004675	0.000957
10005	-0.000398	0.006878

# Approach ①

<調査した内容>

- ・ 線形回帰やSVRなど古典的手法が有効
  - ・ LGBやXGBなど決定木系は効果が少ない
  - ・ 多様なモデルのアンサンブル学習が、やはり有効
- … 手法に関するものが多く、特徴量作成については情報なし



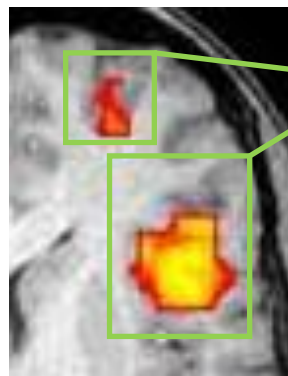


# Approach ②

<独自の取り組み>

## 1. 特徴量の作成

部位毎に、機能データを形態データで割る



相関係数 ÷ 信号強度

→ 効果は誤差程度...

age : 0.1567 → 0.1558 (上位 : 0.14台)

frontal & insular

SCN(53)_vs_SCN(69)	
0.368580	
IC_01	IC_07
0.006070	0.014466

frontal

insular

## 2. アンサンブル

Blending (Ridge, SVR, LGB, XGB, Neural) : 0.164

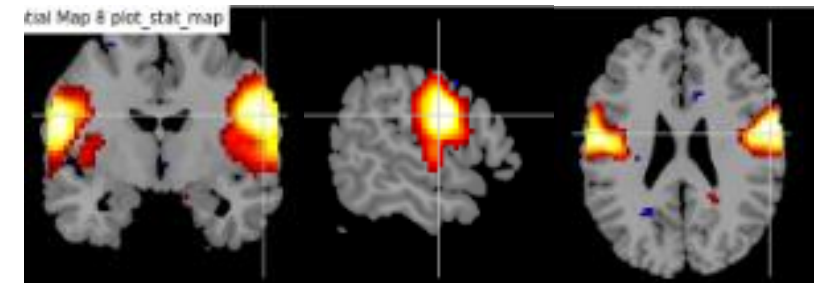
Stacking (SVR→LGB→KNN→Ridge) : 0.162 (best)

→ 画像は扱えず、812位で終了 (銅メダル : 0.159以下)

# Image Data / Approach ③

<Image data>

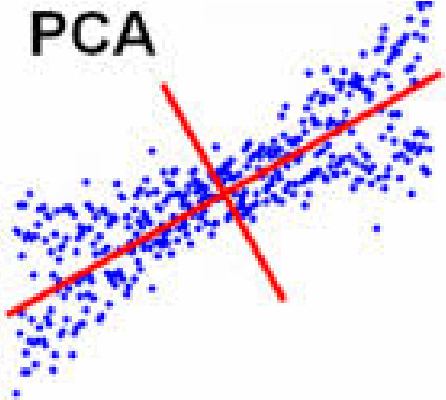
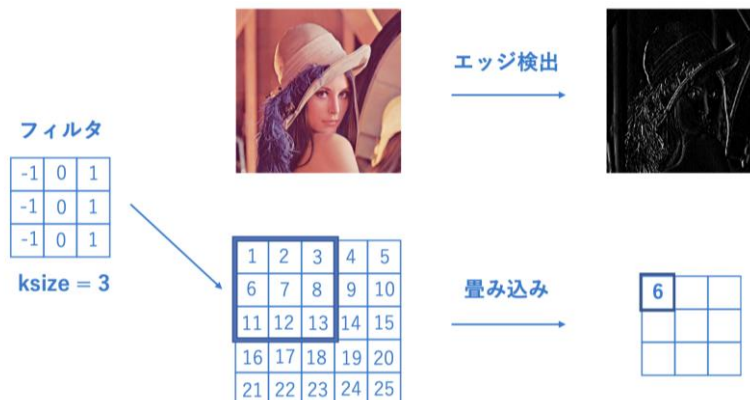
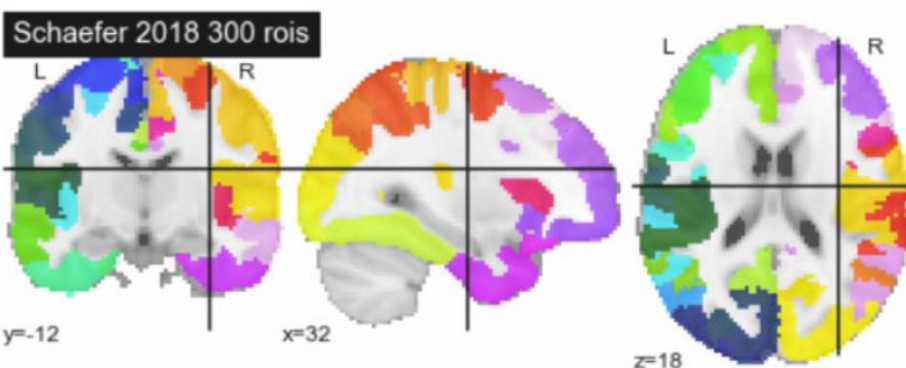
- $(53 \times 52 \times 63 \times 53)$  の4次元配列  
空間座標  $(x, y, z)$  部位
- 脳実質ではなく、血流信号のみ



$(x, y, z) \times 53$

<調査した内容>

- 3D convolution、AutoEncode → PCA など...
- Nilearn の機能マップから、該当部位の統計量抽出 など...





# Recurrent Neural Network ①

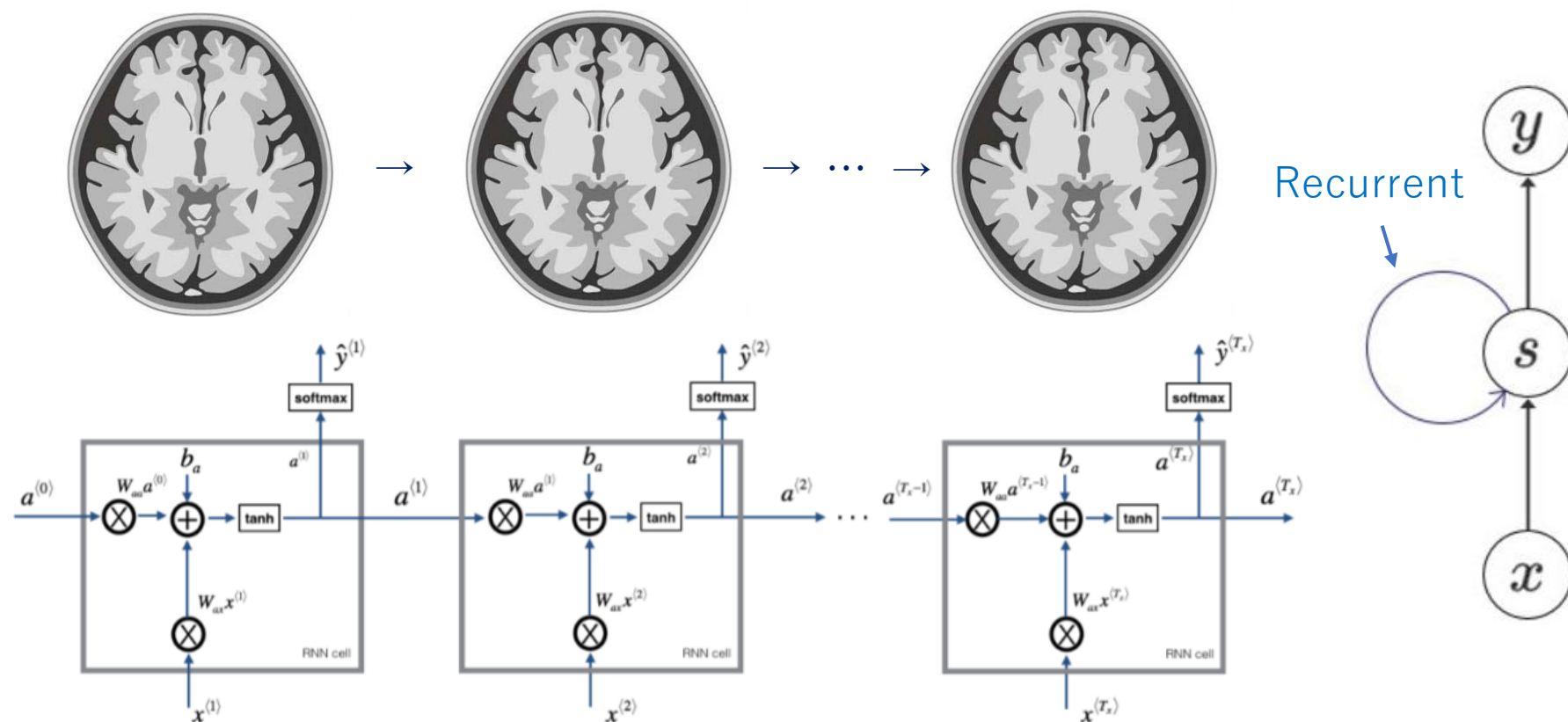
<独自の取り組み>

- ・年齢は時系列データでは？ → 再帰型ニューラルネットワーク

$t$  歳

$(t+1)$  歳

$\dots (t+n)$  歳



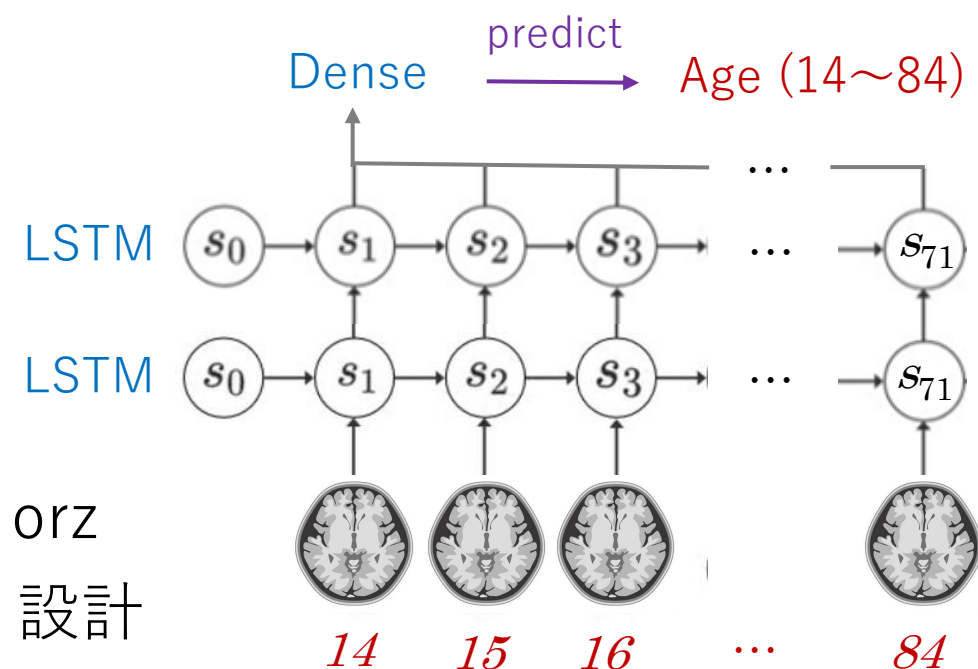
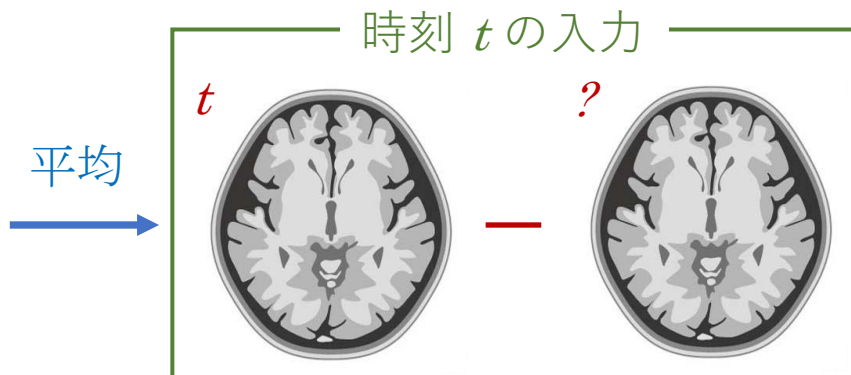
# Recurrent Neural Network ②

## <方法>

1. 年齢別にデータをまとめ、平均をとる（欠損は内分）
2. 経年変化を表す系列データを作成（14～84歳）
3. 上記から予測対象を引いたものを入力 → 年齢出力を学習  
（モデルはLSTMなどRNN 1～3層 + 全結合層）

$t$  歳のデータ

Id	age	IC_01	IC_07
16958	14.0	-0.003879	0.013171
14430	14.0	0.004617	0.014784
13914	14.0	0.004755	0.005741
11392	14.0	0.006085	0.010992



## <結果>

最高でもscore 0.182 程度 ... orz

原因：目的外使用、離散値、設計

# まとめ / 今後に向けて

## < 課題で得たもの >

- ・ DICで学んだ各手法の実践経験（*習うより慣れよ*）
- ・ 予測精度を高める技術（*score = 価値?*）

## < 反省点 >

- ・ コンペ自体は中途半端
- ・ 情報収集・発信が不十分

## < 今後に向けて >

- この経験を出発点とし、分析技術向上を図る
- 画像（特に神経）関連のコンペ/タスクに応用
- 画像に限らず、学んだ事を活かせる場を探していく

