

競馬予測界の  
ディープインパクトを  
作るために

ウマアナリティクス #1

henry

# 自己紹介

- ・ 大元 司 (Tsukasa ŌMOTO, henry)
- ・ [@henry0312](#)
- ・ <https://github.com/henry0312>
- ・ 学生の頃の専門はトピックモデル
- ・ 仕事はレコメンダーとその基盤の開発

# 競馬歴

- ・ ちょうど5ヶ月ぐらい
  - ・ 競馬の予測をガチでやってみた - stockedge.jpの技術メ  
モ
- ・ 読んだ本
  - ・ 勝ち馬がわかる競馬の教科書
  - ・ 馬券裁判 -競馬で1億5000万円儲けた予想法の真実-
- ・ 競馬場に一度も行ったことが無いので年内に一度は・・・

# 今日の内容

- ・ ディープラーニングを使って最高に勝てる競馬予測モデルを作るために試行錯誤した事例を紹介します
- ・ **今日話すモデルを作っても十中八九勝てません**
  - ・ 買い目やレースを絞れが可能性はあると思うが・・・
- ・ ただし、今後何をすれば良いのか？というのがわかれば**勝てる**モデルに近づくことができると信じています

# 目次

- ・ 競馬予測とは
- ・ ディープラーニングについて
- ・ データ
- ・ 3着以内予測モデル
- ・ 走破タイム予測モデル
- ・ 今後の展望

# 目次

- ・ 競馬予測とは
- ・ ディープラーニングについて
- ・ データ
- ・ 3着以内予測モデル
- ・ 走破タイム予測モデル
- ・ 今後の展望

# 競馬予測

- ・ 着順を正確に当てることは目的ではない
- ・ 馬券に絡んで来そうな馬を予想する
- ・ 期待値で回収率100%を超える馬の組み合わせ（馬券の購入方法）を探す
- ・ 競馬場に行くこと無く、データから推定する
- ・ さらに金曜日に全ての作業を終えたい・・・

# 目次

- ・ 競馬予測とは
- ・ ディープラーニングについて
- ・ データ
- ・ 3着以内予測モデル
- ・ 走破タイム予測モデル
- ・ まとめ

# ディープラーニング

- ・ ディープラーニング（深層学習）は、多層のニューラルネットワークを用いた機械学習の方法論
- ・ 近年その有効性が広く認知され、多くの研究者や技術者の関心を爆発的に集めている

# フレームワーク

- ・ 研究はもちろんのこと、深層学習フレームワークも戦国時代にある
  - ・ Theano, Chainer, Tensorflowなど
- ・ Kerasがおすすめ！
  - ・ TheanoとTensorflowをバックエンドに利用可能
  - ・ <http://keras.io/ja/>

# 目次

- ・ 競馬予測とは
- ・ ディープラーニングについて
- ・ データ
- ・ 3着以内予測モデル
- ・ 走破タイム予測モデル
- ・ まとめ

# データ

- ・ JRDB
  - ・ 固定長のテキスト
  - ・ 月額 1,980円～
- ・ JRA-VANにも同様のデータがあるのだろうか？
  - ・ plain textデータが見当たらない・・・

# 特徴選択の大前提

- ・ 基本的には事実のみを利用する
  - ・ 人の感覚による指数（調教指数や人気指数など）は利用しない
- ・ レース前日までに利用できる情報を利用する
  - ・ 当日の情報（オッズ、体重など）は利用しない
- ・ 血統情報は利用しない
  - ・ 血統の善し悪しは成績に表れると仮定
    - ・ 新馬戦どうするんだ・・・（課題）

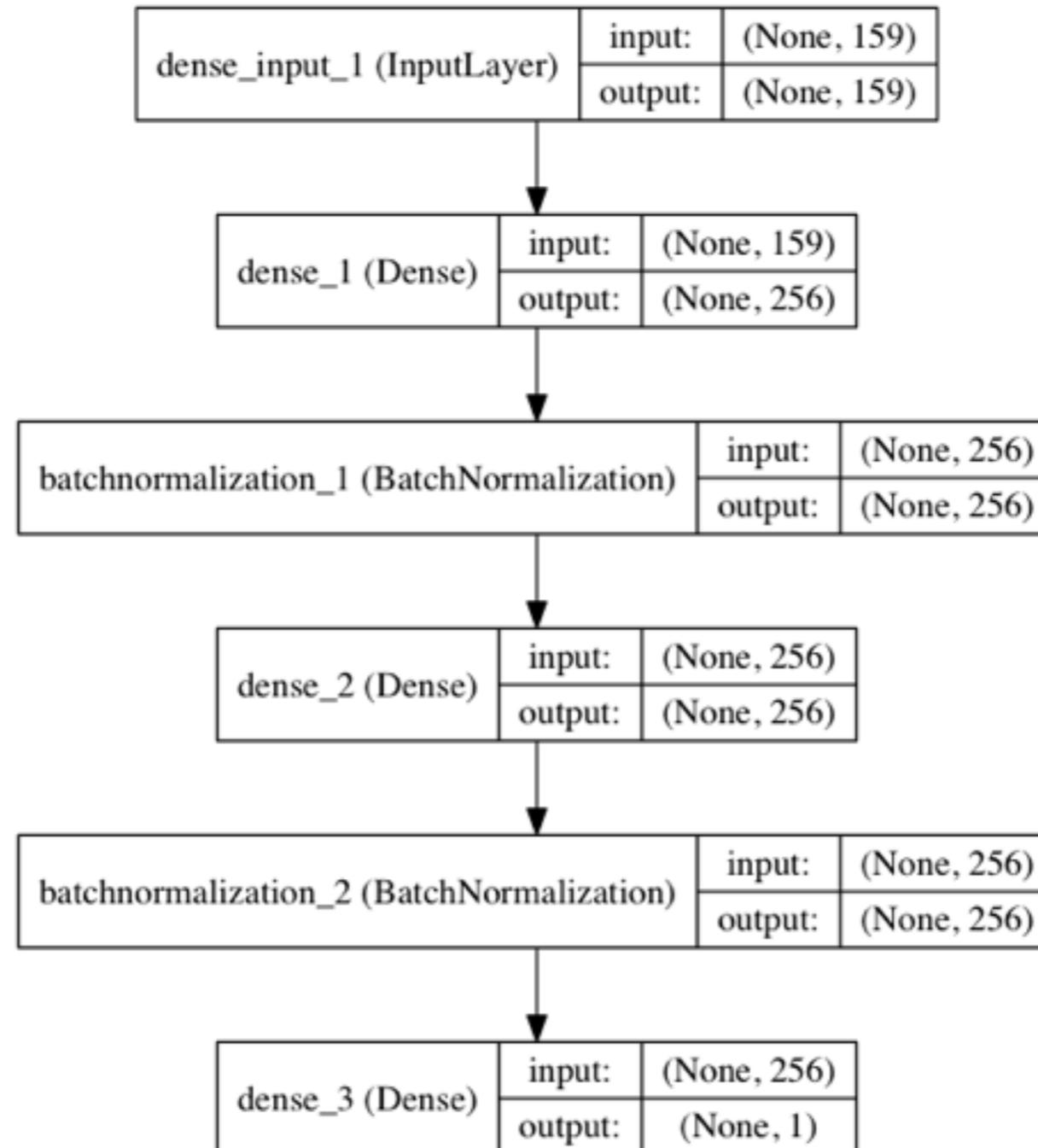
# 目次

- ・ 競馬予測とは
- ・ ディープラーニングについて
- ・ データ
- ・ 3着以内予測モデル
- ・ 走破タイム予測モデル
- ・ まとめ

# 3着以内予測モデル

- ・ 出走馬が3着以内に入る確率を予測する
- ・ 競馬場 x 芝・ダート x 距離の組み合わせでモデルを組む
  - ・ 1,600m以下、1,600m超 2,200m未満、2,200m以上
- ・ 正例（3着以内に着た馬）：負例 = 1：1

# ネットワーク図



# モデル精度

## 回収率

ワイドBOX	78.486783
三連単BOX	84.189233
三連複BOX	76.453724
単勝	88.223659
枠連BOX	37.837842
複勝	83.731238
馬単BOX	87.469308
馬連BOX	83.147319

dtype: float64

## 標準偏差

ワイドBOX	9.125506
三連単BOX	46.147552
三連複BOX	32.951770
単勝	11.499427
枠連BOX	3.711086
複勝	3.481503
馬単BOX	23.282112
馬連BOX	18.653187

dtype: float64

## 的中率

ワイドBOX	57.726272
三連単BOX	45.033313
三連複BOX	45.020106
単勝	78.643832
枠連BOX	72.813497
複勝	72.115332
馬単BOX	60.531429
馬連BOX	60.491465

dtype: float64

## 標準偏差

ワイドBOX	5.518842
三連単BOX	6.387588
三連複BOX	6.303718
単勝	7.772632
枠連BOX	7.702738
複勝	8.615014
馬単BOX	8.086571
馬連BOX	8.075304

dtype: float64

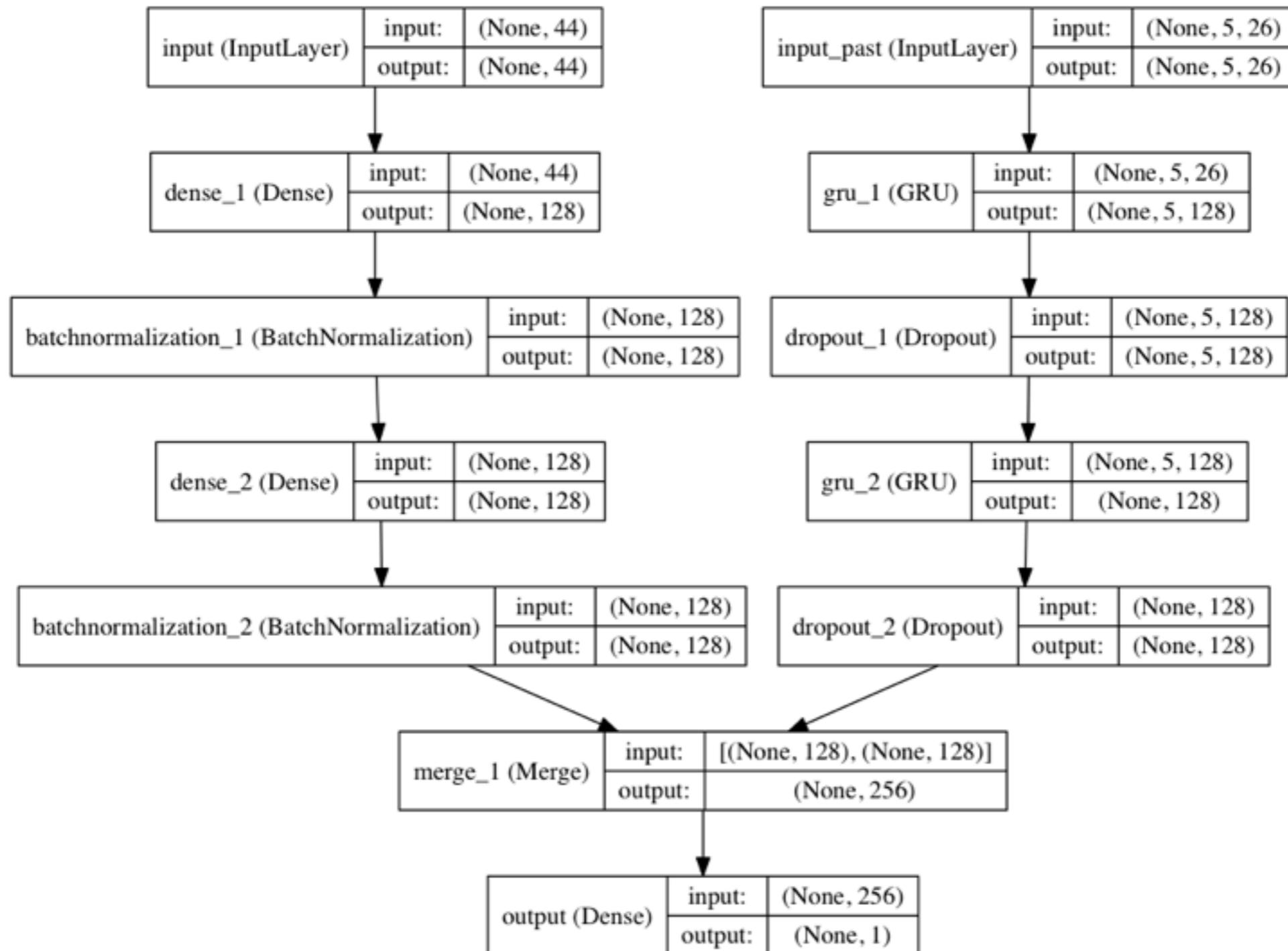
## Performance

[ [ 8.96325857e-01 3.36629750e-01 ]
[ 6.34557985e-01 7.12994499e-01 ]
[ 7.41784392e-01 4.56268326e-01 ]
[ 2.24000000e+03 5.81600000e+02 ] ]
[ [ 0.02059411 0.0211319 ]
[ 0.04366687 0.07982712 ]
[ 0.02693963 0.03081221 ]
[ 20.18910597 3.0724583 ] ]

## Confusion Matrix

[ [ 1422.2 817.8 ]
[ 167. 414.6 ] ]
[ [ 109.3387397 91.17324169 ]
[ 46.94677838 46.01999565 ] ]

# ネットワーク図 (RNN)



# モデル精度 (RNN)

## 回収率

ワイドBOX	84.548811
三連単BOX	94.529267
三連複BOX	93.415239
単勝	82.832431
枠連BOX	39.712275
複勝	84.768054
馬単BOX	80.423860
馬連BOX	80.351189

dtype: float64

## 標準偏差

ワイドBOX	10.431649
三連単BOX	40.413319
三連複BOX	37.318956
単勝	11.696210
枠連BOX	5.228255
複勝	5.723460
馬単BOX	21.220352
馬連BOX	19.813006

dtype: float64

## 的中率

ワイドBOX	57.726272
三連単BOX	45.033313
三連複BOX	45.020106
単勝	78.643832
枠連BOX	72.813497
複勝	72.115332
馬単BOX	60.531429
馬連BOX	60.491465

dtype: float64

## 標準偏差

ワイドBOX	5.518842
三連単BOX	6.387588
三連複BOX	6.303718
単勝	7.772632
枠連BOX	7.702738
複勝	8.615014
馬単BOX	8.086571
馬連BOX	8.075304

dtype: float64

## Performance

```
[[ 9.06453930e-01  4.00489146e-01]
 [ 7.20744328e-01  7.11150216e-01]
 [ 8.02145294e-01  5.10294641e-01]
 [ 1.12000000e+03  2.90800000e+02]]
[[ 0.01380861  0.02718656]
 [ 0.04132838  0.06323865]
 [ 0.02190195  0.02415163]
 [ 13.51295675  2.74954542]]
```

## Confusion Matrix

```
[[ 807.5  312.5]
 [  84.1  206.7]]
[[ 52.49047533  45.42081021]
 [ 19.07590103  17.51599269]]
```

# 考察

- ・ 的中率は良いのに回収率が上がってこない
- ・ 買い目が多すぎる

東京競馬場 10 芝 2400 東京優駿

3着以内

[ 1 3 4 5 7 8 9 10 12 14 15 16 ]

[ 0.60095537 0.99876684 0.97611648 0.68347794 0.98636204 0.9980619

0.9801656 0.93103594 0.98858529 0.99711454 0.91811293 0.75677949 ]

3着以内 (RNN)

[ 1 3 7 8 9 10 12 14 ]

走破タイム

[10 13 14 12 5]

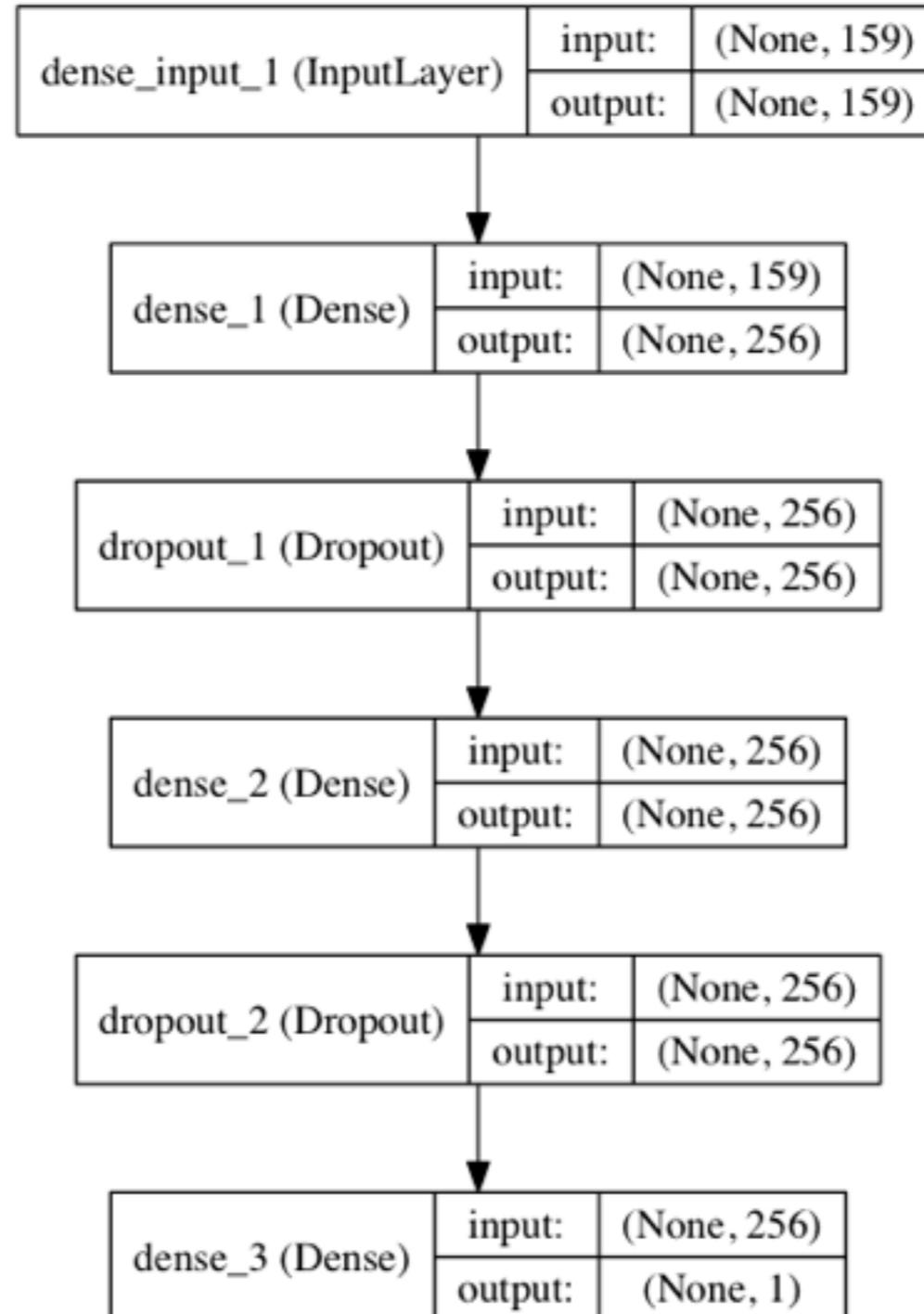
# 目次

- ・ 競馬予測とは
- ・ ディープラーニングについて
- ・ データ
- ・ 3着以内予測モデル
- ・ 走破タイム予測モデル
- ・ まとめ

# 走破タイム予測モデル

- ・ 出走馬のレース走破タイムを予測する
- ・ 競馬場 x 芝・ダート x 距離の組み合わせでモデルを組む
- ・ 全データを利用する（利用できる）

# ネットワーク図



# モデル精度 (Top-1)

回収率	
ワイドBOX	0.000000
三連単BOX	0.000000
三連複BOX	0.000000
単勝	96.650919
枠連BOX	0.000000
複勝	74.378328
馬単BOX	0.000000
馬連BOX	0.000000

dtype: float64

標準偏差	
ワイドBOX	0.000000
三連単BOX	0.000000
三連複BOX	0.000000
単勝	73.231450
枠連BOX	0.000000
複勝	27.965845
馬単BOX	0.000000
馬連BOX	0.000000

dtype: float64

的中率	
ワイドBOX	0.000000
三連単BOX	0.000000
三連複BOX	0.000000
単勝	11.411074
枠連BOX	0.000000
複勝	9.926037
馬単BOX	0.000000
馬連BOX	0.000000

dtype: float64

標準偏差	
ワイドBOX	0.000000
三連単BOX	0.000000
三連複BOX	0.000000
単勝	1.333824
枠連BOX	0.000000
複勝	0.847060
馬単BOX	0.000000
馬連BOX	0.000000

dtype: float64

# モデル精度 (Top-2)

```
回収率
ワイドBOX      0.000000
三連単BOX      0.000000
三連複BOX      0.000000
単勝           92.546744
枠連BOX        37.553038
複勝           81.161230
馬単BOX        30.547869
馬連BOX        30.115361
dtype: float64
```

```
標準偏差
ワイドBOX      0.000000
三連単BOX      0.000000
三連複BOX      0.000000
単勝           36.627278
枠連BOX        18.738282
複勝           13.429145
馬単BOX        9.943144
馬連BOX        15.828053
dtype: float64
```

```
的中率
ワイドBOX      0.000000
三連単BOX      0.000000
三連複BOX      0.000000
単勝           22.667167
枠連BOX        7.892763
複勝           19.217707
馬単BOX        2.997203
馬連BOX        2.997203
dtype: float64
```

```
標準偏差
ワイドBOX      0.000000
三連単BOX      0.000000
三連複BOX      0.000000
単勝           1.888502
枠連BOX        1.984853
複勝           0.895552
馬単BOX        1.030230
馬連BOX        1.030230
dtype: float64
```

# モデル精度 (Top-3)

回収率  
ワイドBOX 237.712911  
三連単BOX 28.224868  
三連複BOX 22.793651  
単勝 85.982252  
枠連BOX 40.095633  
複勝 77.351539  
馬単BOX 197.371495  
馬連BOX 202.818064  
dtype: float64

標準偏差  
ワイドBOX 155.760314  
三連単BOX 48.615895  
三連複BOX 36.206676  
単勝 39.079830  
枠連BOX 12.969737  
複勝 16.616993  
馬単BOX 263.699637  
馬連BOX 254.973322  
dtype: float64

的中率  
ワイドBOX 5.907276  
三連単BOX 0.631191  
三連複BOX 0.634921  
単勝 32.962130  
枠連BOX 16.265467  
複勝 27.448960  
馬単BOX 7.736670  
馬連BOX 7.736670  
dtype: float64

標準偏差  
ワイドBOX 1.631101  
三連単BOX 0.352884  
三連複BOX 0.354931  
単勝 2.325535  
枠連BOX 2.146979  
複勝 2.781015  
馬単BOX 1.562513  
馬連BOX 1.562513  
dtype: float64

# モデル精度 (Top-4)

## 回収率

ワイドBOX	112.859674
三連単BOX	42.086156
三連複BOX	51.077647
単勝	81.894888
枠連BOX	38.111088
複勝	78.993907
馬単BOX	141.345228
馬連BOX	143.154460

dtype: float64

## 標準偏差

ワイドBOX	30.603909
三連単BOX	30.922401
三連複BOX	37.532119
単勝	29.946957
枠連BOX	13.787862
複勝	11.592624
馬単BOX	130.213306
馬連BOX	124.250231

dtype: float64

## 的中率

ワイドBOX	11.292359
三連単BOX	2.833533
三連複BOX	2.849702
単勝	40.888639
枠連BOX	26.070491
複勝	35.370509
馬単BOX	12.473695
馬連BOX	12.473695

dtype: float64

## 標準偏差

ワイドBOX	1.596563
三連単BOX	1.322045
三連複BOX	1.331466
単勝	3.455137
枠連BOX	1.918288
複勝	2.574581
馬単BOX	2.775989
馬連BOX	2.775989

dtype: float64

# モデル精度 (Top-5)

回収率

ワイドBOX	66.022022
三連単BOX	39.138626
三連複BOX	43.261705
単勝	71.974203
枠連BOX	36.970113
複勝	77.384952
馬単BOX	94.753506
馬連BOX	96.687714

dtype: float64

標準偏差

ワイドBOX	13.347196
三連単BOX	30.955663
三連複BOX	28.648507
単勝	22.879004
枠連BOX	10.336719
複勝	7.674877
馬単BOX	75.841384
馬連BOX	71.586333

dtype: float64

的中率

ワイドBOX	16.619697
三連単BOX	5.355795
三連複BOX	5.384425
単勝	46.279215
枠連BOX	37.125359
複勝	42.391479
馬単BOX	17.833161
馬連BOX	17.833161

dtype: float64

標準偏差

ワイドBOX	2.343000
三連単BOX	1.882840
三連複BOX	1.906668
単勝	2.616403
枠連BOX	1.257308
複勝	2.936184
馬単BOX	2.366673
馬連BOX	2.366673

dtype: float64

# 考察

- ・ 的中率が良くない (悪い)
- ・ JRA-VAN
  - ・ データマイニング予測の仕組み
  - ・ 予測の検証結果
  - ・ 回収率と的中率を比較すると結果は妥当か
- ・ レースを絞れば勝てるのか？

# まとめ

- ・ 3着以内予測モデルは買いすぎ
- ・ 走破タイム予測モデルは的中率がいまいち
  - ・ モデルが得意なレースを探せば可能性はある？
- ・ いずれのモデルも出走馬間の優劣を直接モデル化していない