



SIVA(シヴァ)AI競馬開発エンジニアに
よる

「競馬で始める機械学習」 ハンズオン

株式会社GAUSS
Yuta Miyawaki

アジェンダ

- 自己紹介
- 機械学習について
- 競馬予測の手順
- 機械学習アルゴリズムの紹介
- 環境構築
- 作業フォルダの構成
- 予測AIのチューニング

自己紹介

<?xml version="1.0"?>

<speaker>

<name>**Yuta Miyawaki**</name>

<age>**23**</age>

<lang>**python, Japanese**</lang>

<drink>**Monster Energy**</drink>

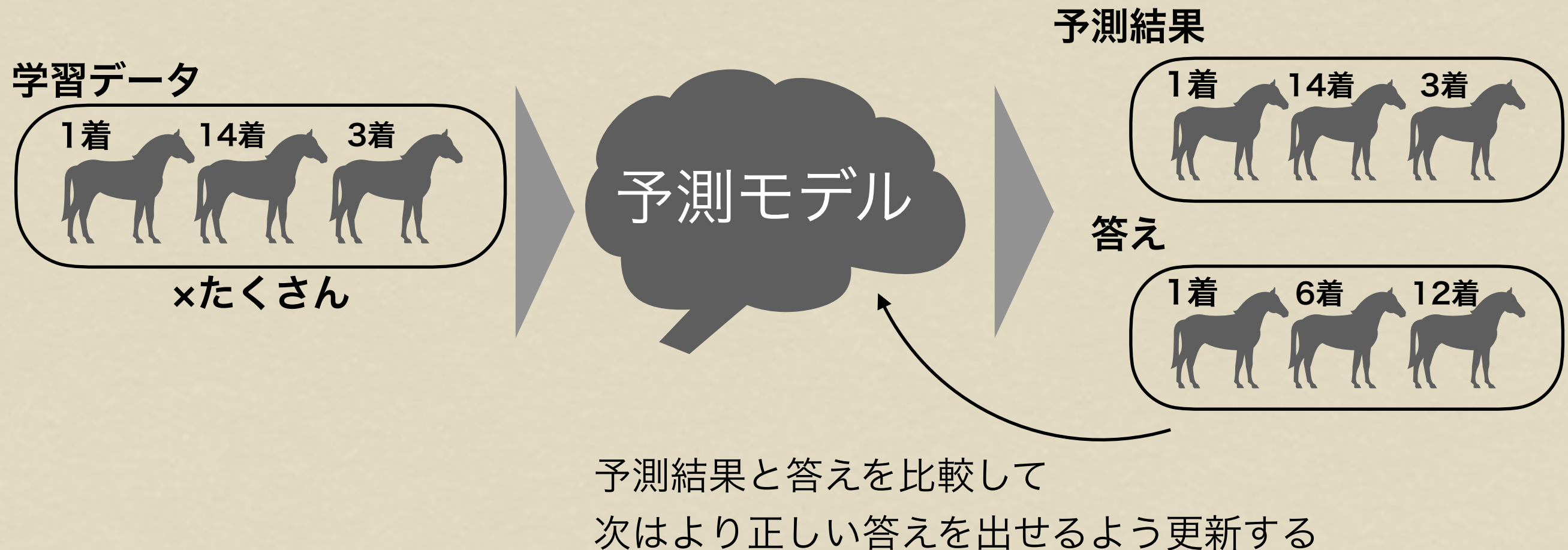
<twitter>**@nami73nbj**</twitter>

</speaker>

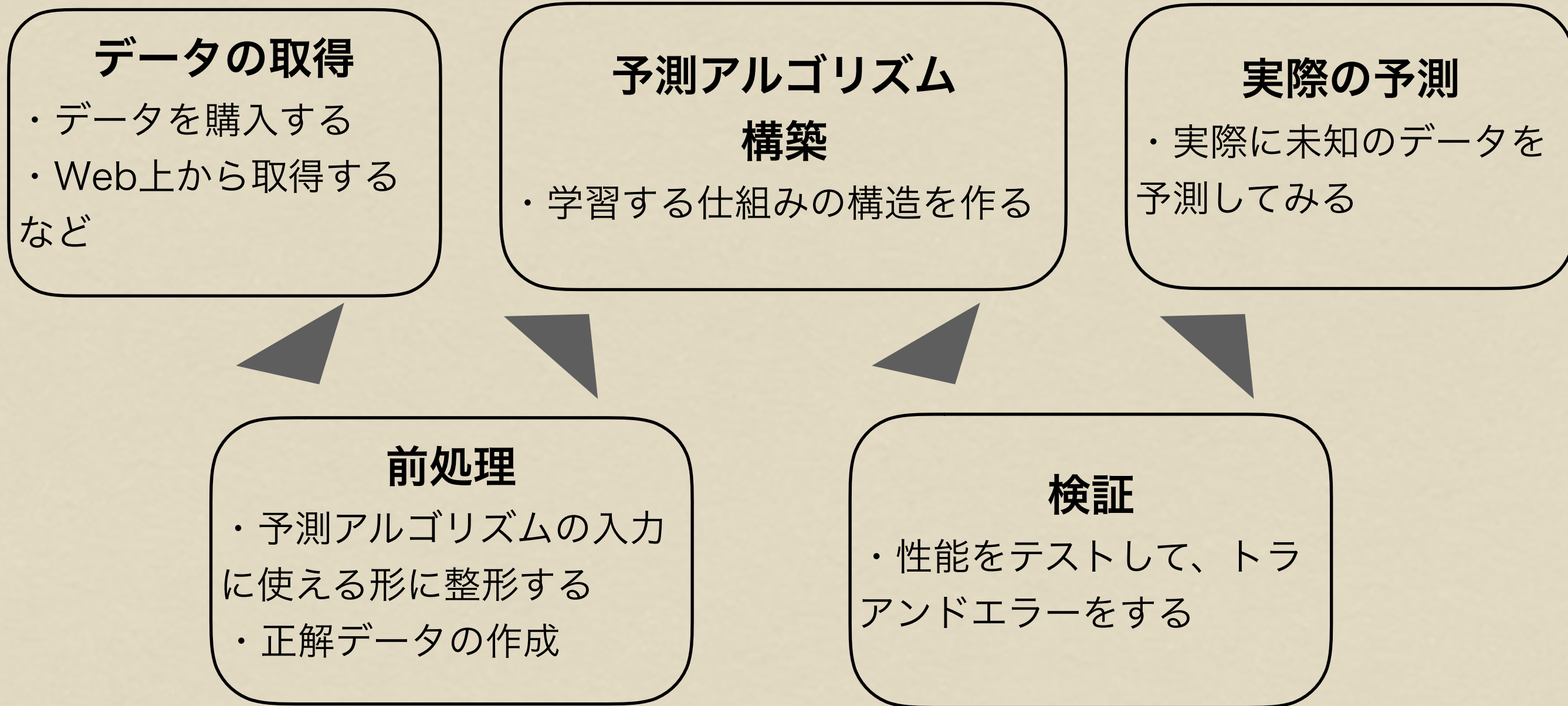


機械学習について

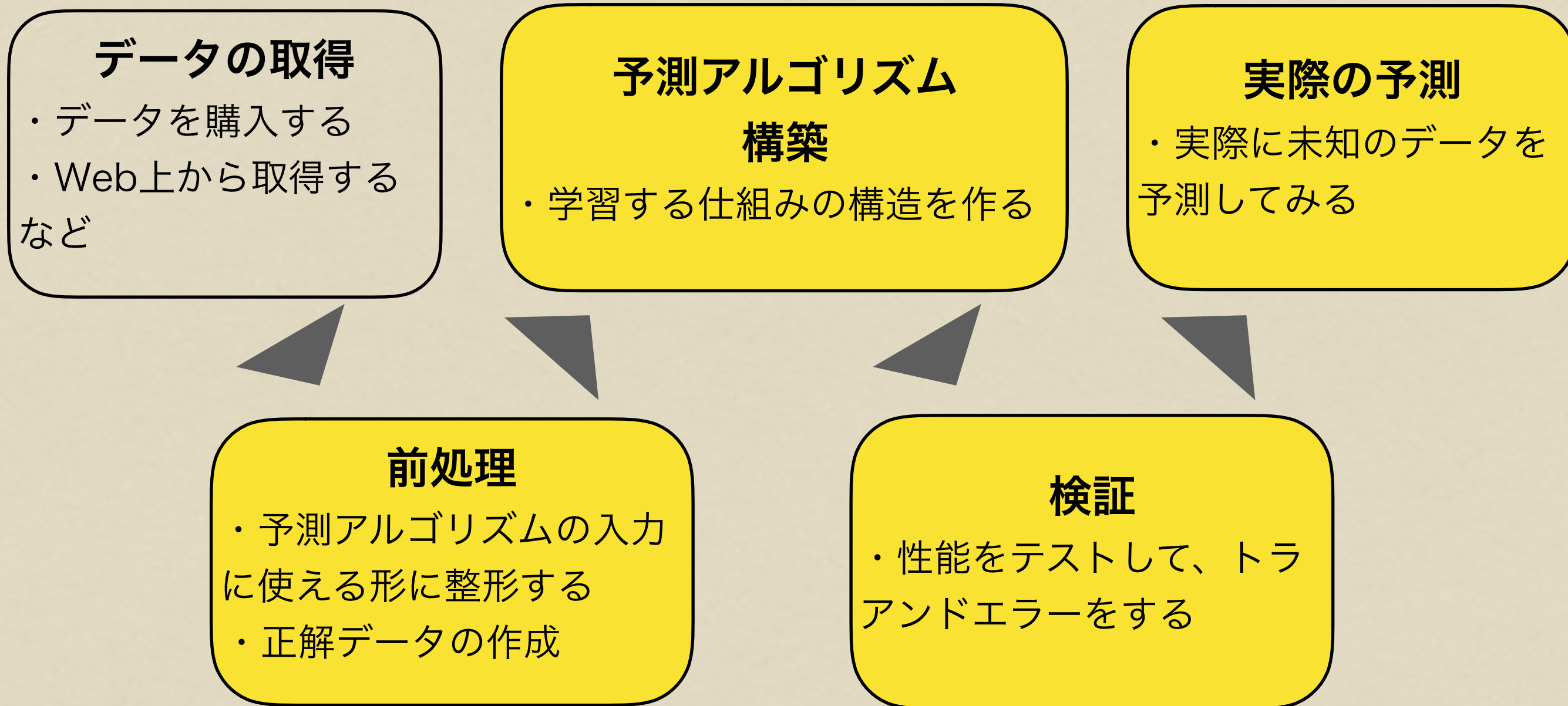
学習データから特徴・パターンを見つけ出し、
未知のデータに対する答えを見つけ出すタスク・アルゴリズム



競馬予測の手順

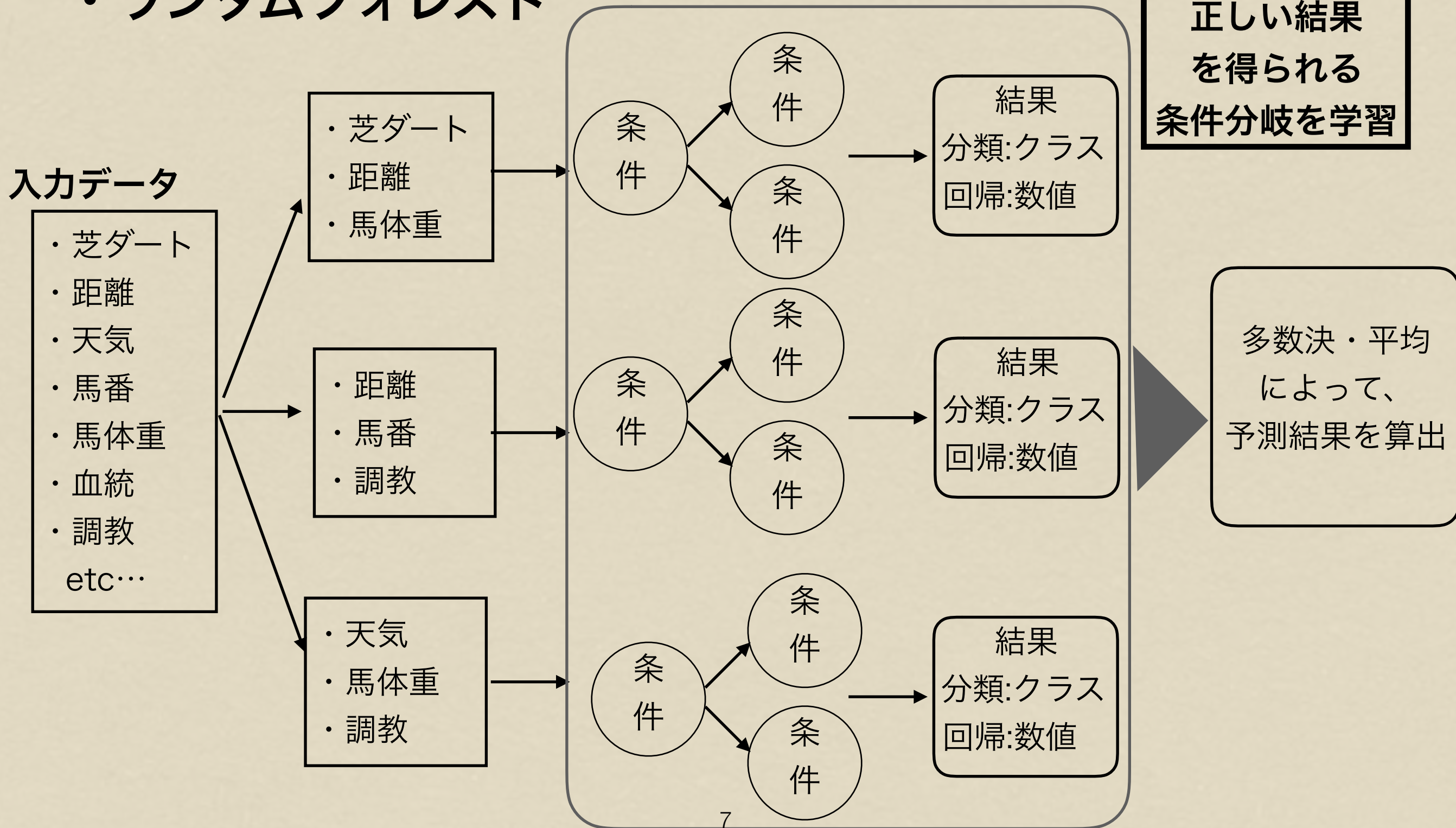


競馬予測の手順



機械学習アルゴリズムの紹介

・ランダムフォレスト

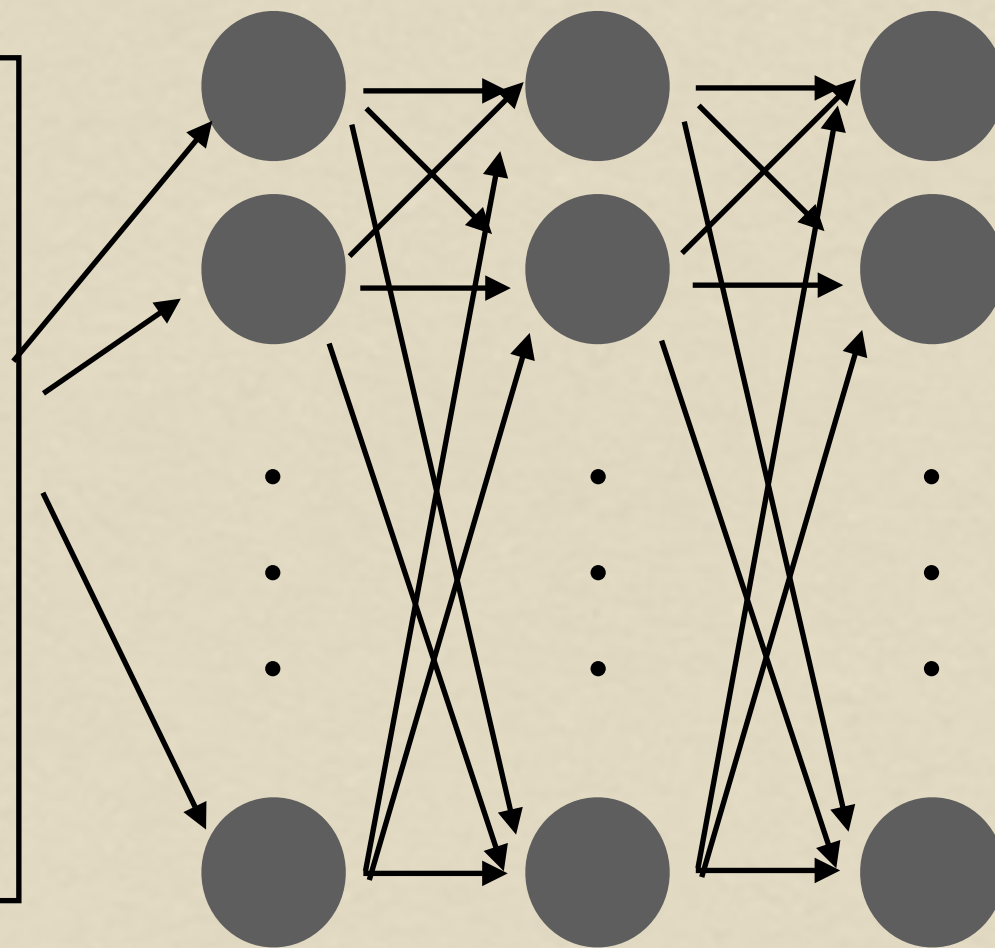


機械学習アルゴリズムの紹介

・ニューラルネット

競馬データ

- ・ 芝ダート
- ・ 距離
- ・ 天気
- ・ 馬番
- ・ 馬体重
- ・ 血統
- ・ 調教
- etc...



入力データから
正しい結果を得られる
パラメータ(矢印)を学習

予測結果

作業フォルダの構成

siva_hands_on_files

- |ー SIVA_hands_on.ipynb メイン作業ノート

- |ー util.py

- |ー preprocessing.py 前処理カンニング用

- |ー SIVA_hands_on_slide このスライド

- L data

- |ー data_train.csv 学習データ

- |ー data_test.csv 検証データ

- |ー data_today_fukushima_11r.csv 福島11Rの検証データ

- |ー data_today_tyukyo_11r.csv 中京11Rの検証データ

- |ー data_today_hakodate_11r.csv 函館11Rの検証データ

- |ー horse_name_fukushima_11r.csv 福島11Rの馬名データ

- |ー horse_name_tyukyo_11r.csv 中京11Rの馬名データ

- L horse_name_hakodate_11r.csv 函館11Rの馬名データ

環境構築

- リポジトリからGit clone
‘git clone https://github.com/nami73b/siva_hands_on’
- Docker Imageのビルド
‘docker build ./ -t siva_handson’
- Dockerコンテナの起動
‘docker run -it -p 8888:8888 siva_handson’

予測モデルのチューニング

前処理編

[例1]

目的変数の分類の閾値を変更する(デフォルト=3)

[例2]

空値を補完する数値を変更する(デフォルト=0)

予測モデルのチューニング

特徴量編

[例1]

騎手コードの有無で精度を比較する

[例2]

走破タイムを「分.秒.コンマ秒」→「秒」などに変換する

(注意)

実装した場合「drop_columns_list」から走破タイム_n走前を取り除く必要がある

予測モデルのチューニング

特徴量編

[例3]

開催日付ファクタを月と日付に分割する

開催日付:1123 -> 開催月:11,開催日:12

[例4]

前走からの経過日数を計算する

datetimeライブラリを使用する

予測モデルのチューニング

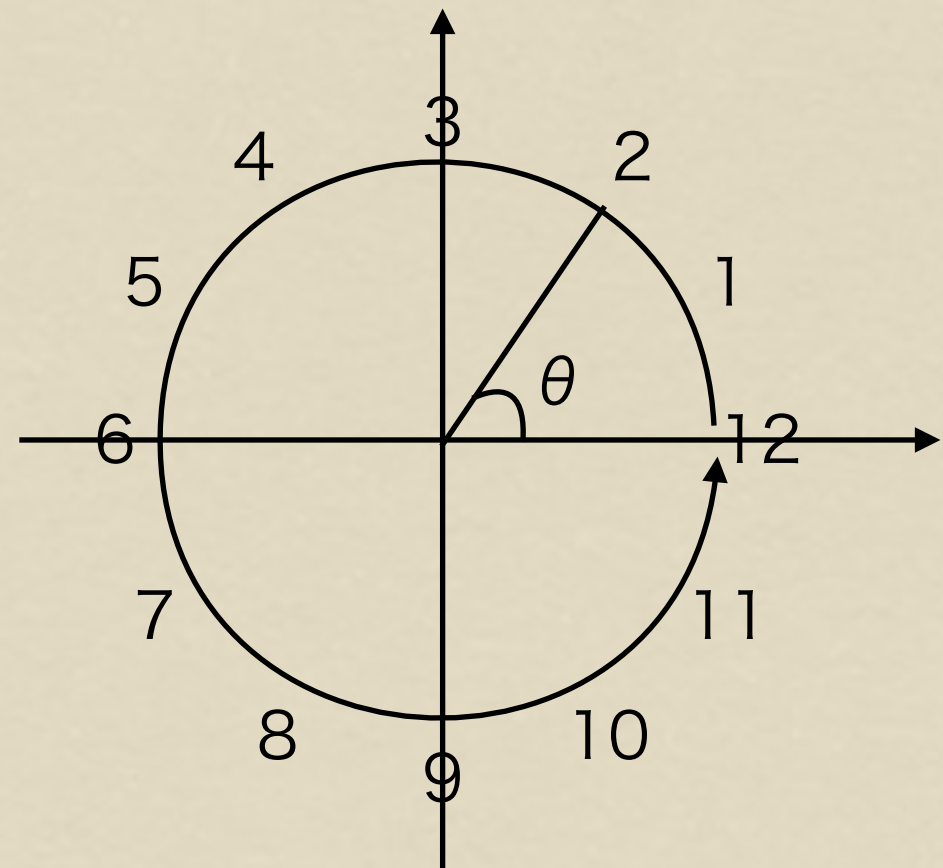
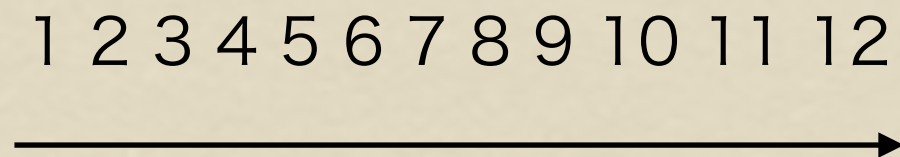
特徴量編

[例5]

開催月、開催日を三角関数で表現する

$$\text{df}[\text{'開催月_sin'}] = \sin(\theta)$$

$$\text{df}[\text{'開催月_cos'}] = \cos(\theta)$$



予測モデルのチューニング

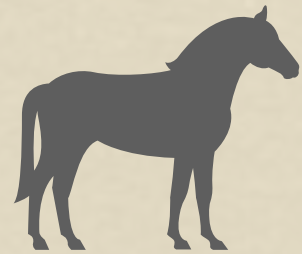
アルゴリズム編

- ・ ランダムフォレスト

fit関数のハイパーパラメータを調整する

- n_estimators : 木の本数
- max_depth : 木の深さの最大
- max_features : 1つの木がランダムで選択するファクタ数
など

予測モデルのチューニング



自由時間です！

各自チューニングした予測モデルで
本日の以下3レースを予測し、
各レース3頭を推奨し、ホワイトボードに記入してください

・ 函館11R
15:25
TVh杯

・ 中京11R
15:35
白川郷S

・ 福島11R
15:45
テレビユー福島賞