

# 予測と判断

(時系列データ分析、移動平均、  
RNNとLSTM の紹介)

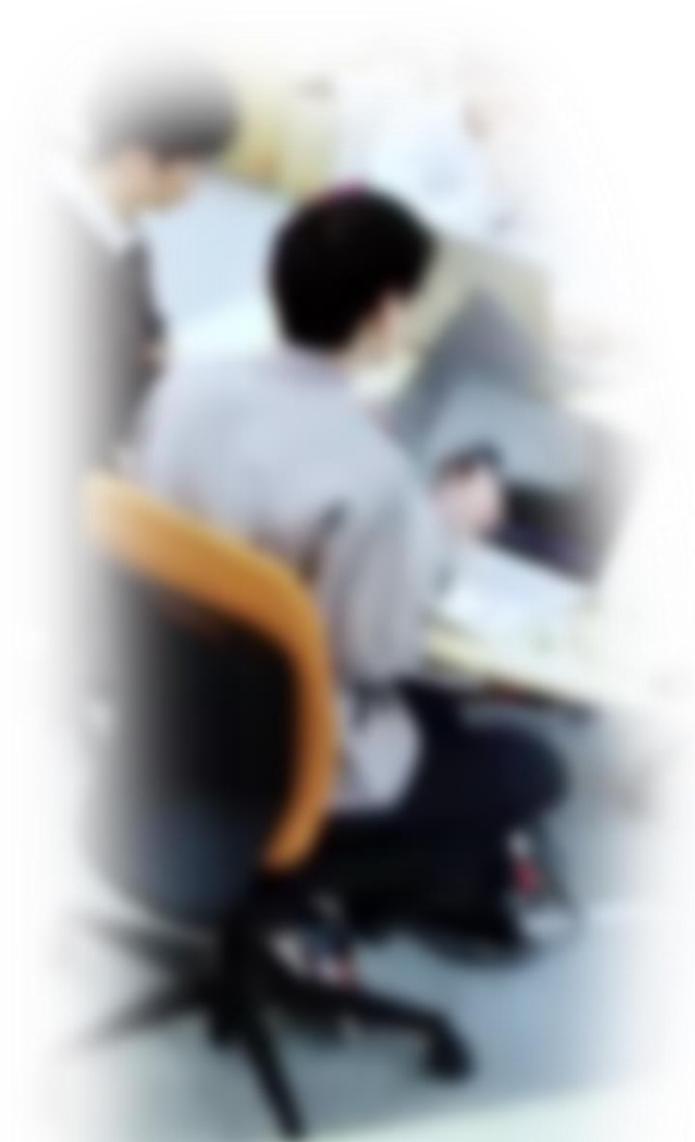
AI演習

<https://www.kkaneko.jp/ai/ae/index.html>

金子邦彦



- 
- ① **トレンドや周期性の分析など時系列データの基礎**
  - ② **時系列データを扱う技術の理解**
  - ③ **実用的なスキルの獲得**



# アウトライン

1. イントロダクション
2. 移動平均
3. 時系列データ分析の概要
4. 時系列データの分析（Prophetを使用）
5. リカレントニューラルネットワーク（RNN）の基本
6. LSTMの基本

# Google Colaboratory



Colaboratory へようこそ

Colaboratory は、完全にクラウドで実行される Jupyter ノートブック環境です。設定不要で、無料でご利用になれます。

```
x = [5, 4, 1, 3, 2]
for i in x:
    print(i * 120)
```

600  
480  
120  
360  
240

PRO ファイル 編集 表示 挿入 ランタイム

+ コード + テキスト | ドライブに...

```
[1] x = 100
```

```
if (x > 20):
    print("big")
else:
    print("small")
```

big

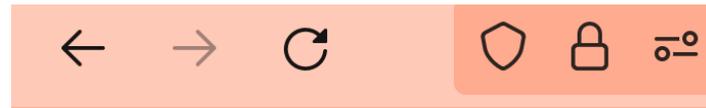
```
s = 0
for i in [1, 2, 3, 4, 5]:
    s = s + i
print(s)
```

15

URL: <https://colab.research.google.com/>

- オンラインで動く
- Python のノートブックの機能を持つ
- Python や種々の機能がインストール済み
- 本格的な利用には、Google アカウントが必要

# Google Colaboratory の全体画面



Colab の定期購入を最大限に活用する  
ファイル 編集 表示 挿入 ランタイム ツール ヘルプ

メニュー

+ コード + テキスト

コードセル, テキストセル  
の追加



メニュー

(目次, 検索と置換,  
変数, ファイル)

```
1. 変数
[2] x = 100
    y = 200

2. 式
▶ print(x + y)
  print(3 * x + y)
300
500

3. 条件分岐
[4] if (x > 50):
    print('big')
    else:
    print('small')
big
```

コードセル,  
テキストセルの  
並び

Web ブラウザの画面

# Google Colaboratory のノートブック



## コードセル, テキストセルの2種類

- **コードセル** : Python プログラム, コマンド, 実行結果
- **テキストセル** : 説明文, 図

2. 式

← テキストセル

```
[5] print(x + y)
     print(3 * x + y)
```

← コードセル

```
300
500
```

3. 条件分岐

← テキストセル

```
▶ if (x > 50):
    print('big')
else:
    print('small')
```

← コードセル

```
big
```

# 9-1. イントロダクション

# 人工知能

知的なITシステム

## 機械学習

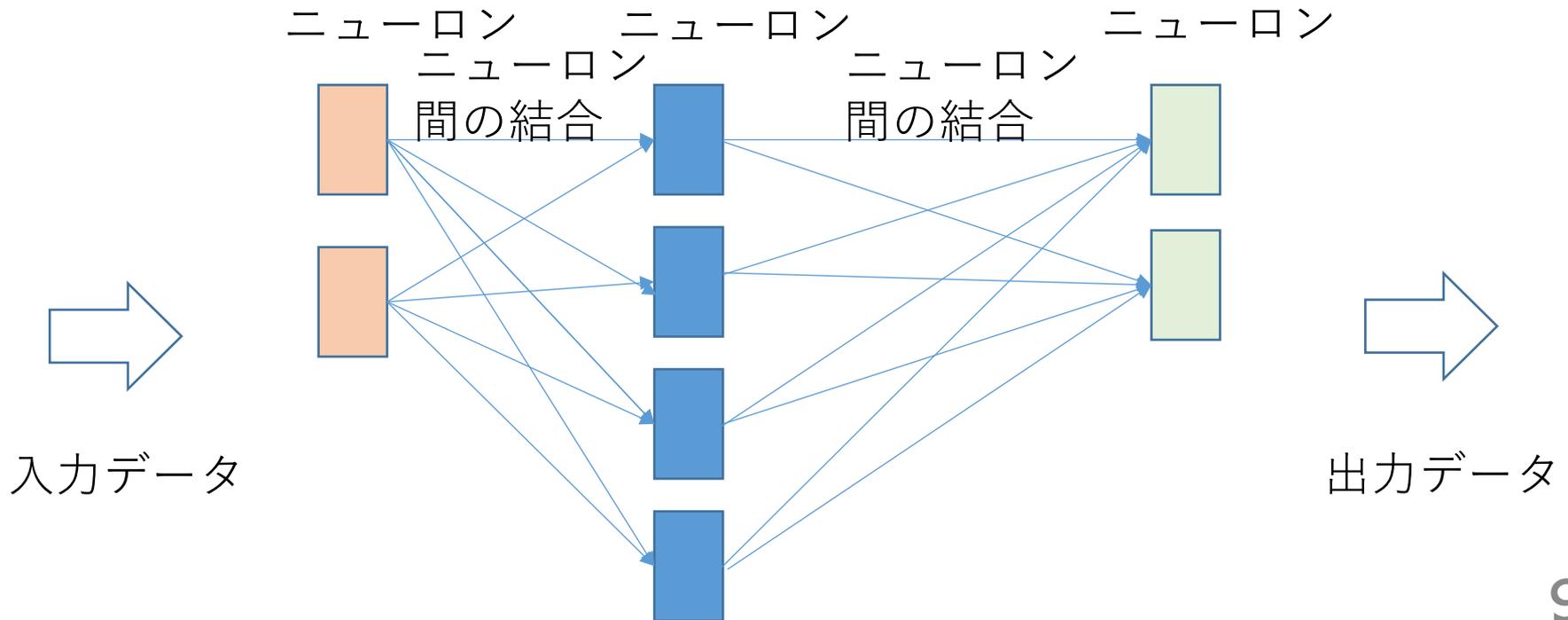
データから**学習**し、知的能力を向上

## ディープラーニング

データから**学習**し、複雑なタスクを実行。**多層のニューラルネットワーク**を使用

# ニューロンとニューラルネットワーク

- **ニューロン**は、**ニューラルネットワーク**の基本的な構成要素
- 一つ一つの**ニューロン**は、データの受け取り、処理、伝達を行う
- **ニューラルネットワーク**は、これらの**ニューロン**が多数組み合わさったもの

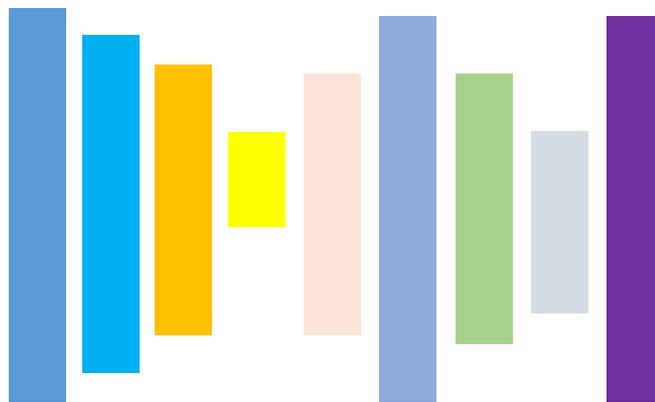


# ディープラーニング

ディープラーニングに「ディープ」とついているのは、多層のニューラルネットワークを使用するため



層の数が少ない

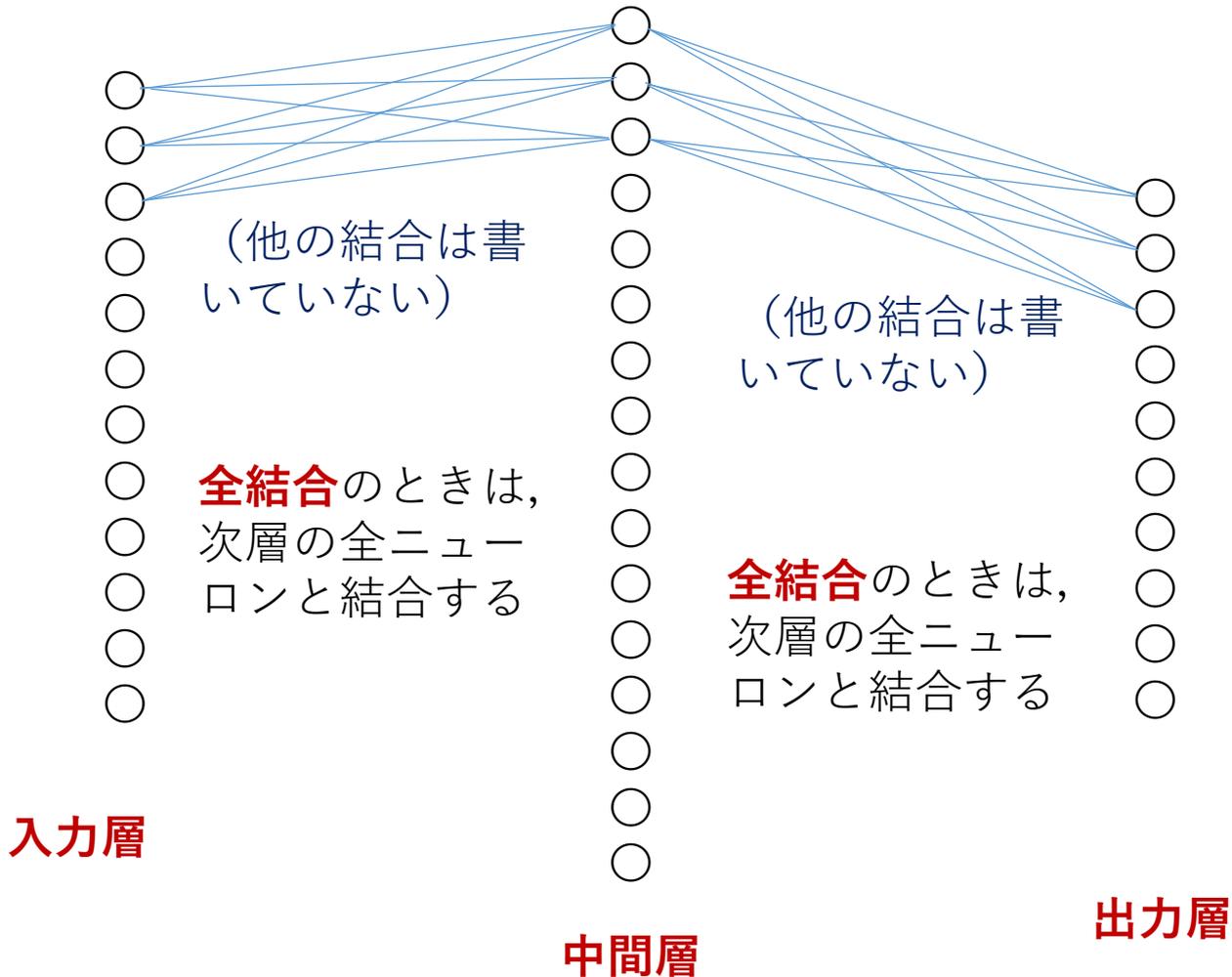


層の数が多し (ディープ)

# フォワードプロパゲーション

ニューロンは、前の層のニューロンからデータを受け取り、それに対して処理を行い、その結果（=活性化度）を次の層のニューロンに伝達する

○はニューロン，線は結合





## 9-2. 移動平均

# 平均



- **平均**は、データの**合計**を、**データの個数**で割った**もの**

10, 40, 30, 40 の**平均**:  $120 \div 4$  で **30**

- **データの代表値としての平均**：平均は、データの「中心」である。データの全体的な傾向を把握できる
- **誤差の軽減**：複数の計測値から平均を取ることで、個々の計測誤差を相殺し、より正確な推定値を得る
- **データの異常値や外れ値の識別**：平均値を知ることによって、異常値や外れ値があるかどうかを判断する手掛かりになる
- **移動平均の利用**：移動平均は、平均の一種。時系列データにおいて、特定の期間にわたる平均値を計算したもの

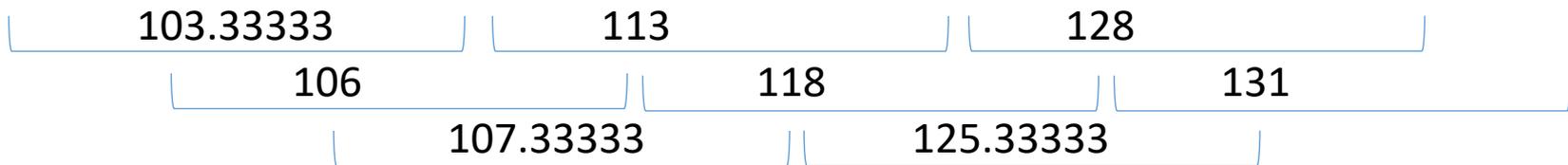
# 移動平均



**移動平均**は、時系列データにおいて、**特定の期間にわたる平均値**を計算したもの。短期的な変動が滑らかになり、トレンドや周期性を分析しやすくなる。

例

102, 105, 103, 110, 109, 120, 125, 131, 128, 134]



期間 3 の移動平均

# 平均と移動平均の要点



## 平均

- 平均はデータの合計をデータの個数で割ったもの  
例: 10, 40, 30, 40のデータの平均は30。

## 平均の有用性

- **データの代表値としての平均**：平均は、データの「中心」である。データの全体的な傾向を把握できる
- **誤差の軽減**：複数の計測値から平均を取ることで、個々の計測誤差を相殺し、**より正確な推定値を得る**
- **データの異常値や外れ値の識別**：平均値を知ることによって、**異常値や外れ値があるかどうか**を判断する手掛かりになる

## 移動平均

- **移動平均**は、時系列データにおいて、**特定の期間にわたる平均値**を計算したもの。
- 短期的な変動が滑らかになり、トレンドや周期性を分析しやすくなる。

## 9-3. 時系列データ分析の基礎

# 時系列データ

時系列データは、時間の経過に伴って順序付けられたデータ  
の並び

例

2023年12月3日の気温は15度

2023年12月4日の気温は13度

...

# 時系列データの例：太陽黒点観測データ

11314	1848	12	23	1848.977	353	23.8	1
11315	1848	12	24	1848.980	240	19.6	1
11316	1848	12	25	1848.982	275	21.0	1
11317	1848	12	26	1848.985	352	23.8	1
11318	1848	12	27	1848.988	268	20.8	1
11319	1848	12	28	1848.990	285	21.4	1
11320	1848	12	29	1848.993	343	23.5	1
11321	1848	12	30	1848.996	340	23.4	1
11322	1848	12	31	1848.999	238	19.6	1
11323	1849	1	1	1849.001	287	20.9	1
	year	month	day	dec_year	sn_value	sn_error	obs_num
72855	2017	6	21	2017.470	35	1.0	41
72856	2017	6	22	2017.473	24	0.8	39
72857	2017	6	23	2017.475	23	0.9	40
72858	2017	6	24	2017.478	26	2.3	15
72859	2017	6	25	2017.481	17	1.0	18
72860	2017	6	26	2017.484	21	1.1	25
72861	2017	6	27	2017.486	19	1.2	36
72862	2017	6	28	2017.489	17	1.1	22
72863	2017	6	29	2017.492	12	0.5	25
72864	2017	6	30	2017.495	11	0.5	30

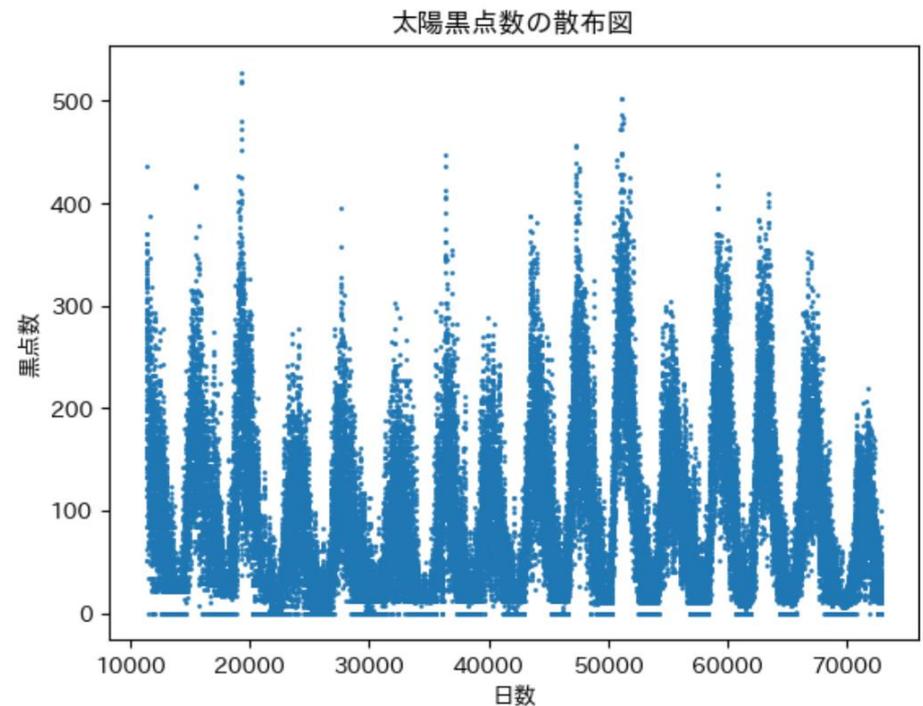


年月日の値 黒点数

黒点数の観測データ

1848年12月23日～2017年6月30日の  
毎日の黒点数データが公開されている

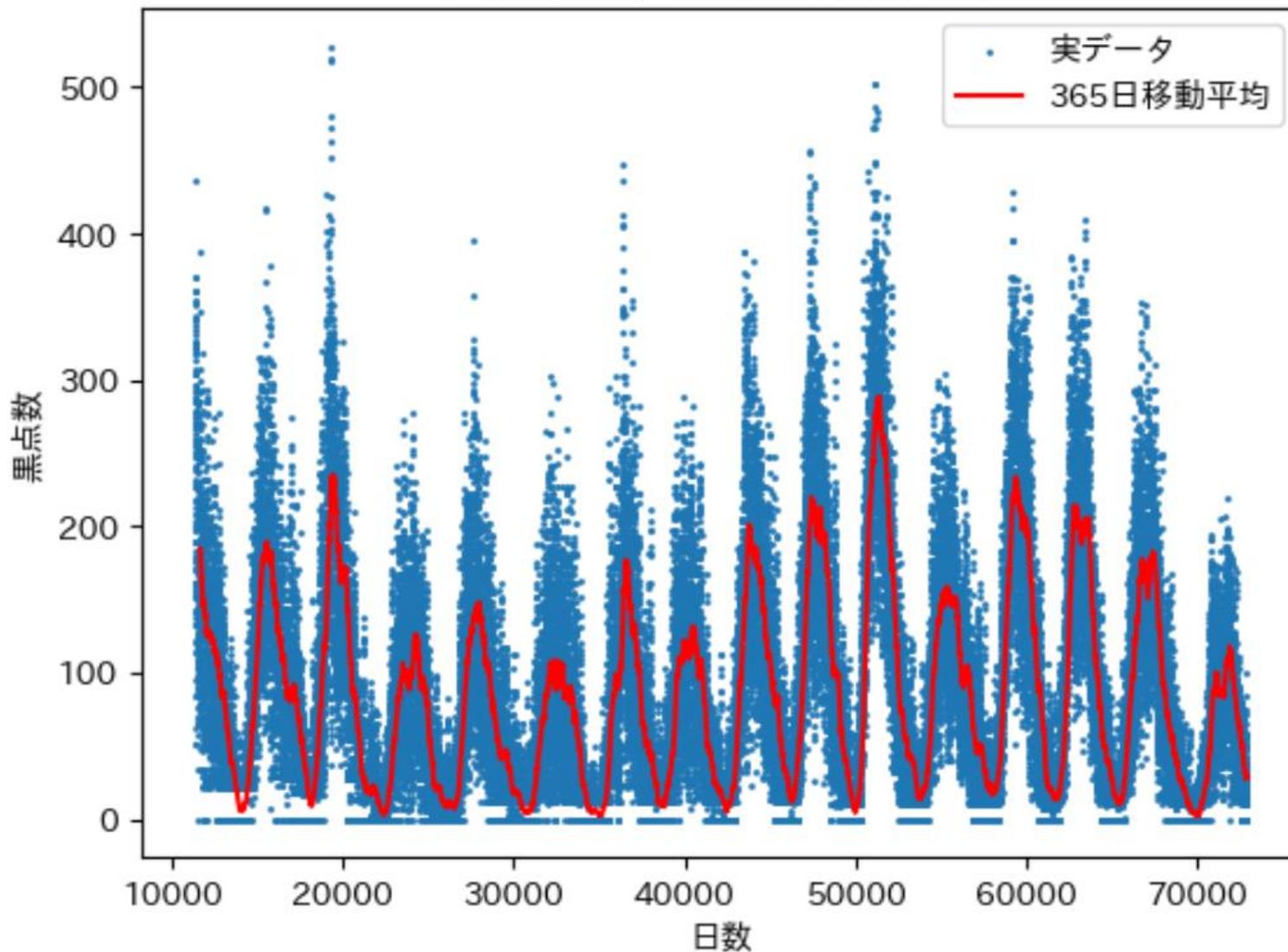
[https://data.heatonresearch.com/data/t81-558/SN\\_d\\_tot\\_V2.0.csv](https://data.heatonresearch.com/data/t81-558/SN_d_tot_V2.0.csv)



散布図でプロット

# 周期性の分析の例（太陽黒点データ）

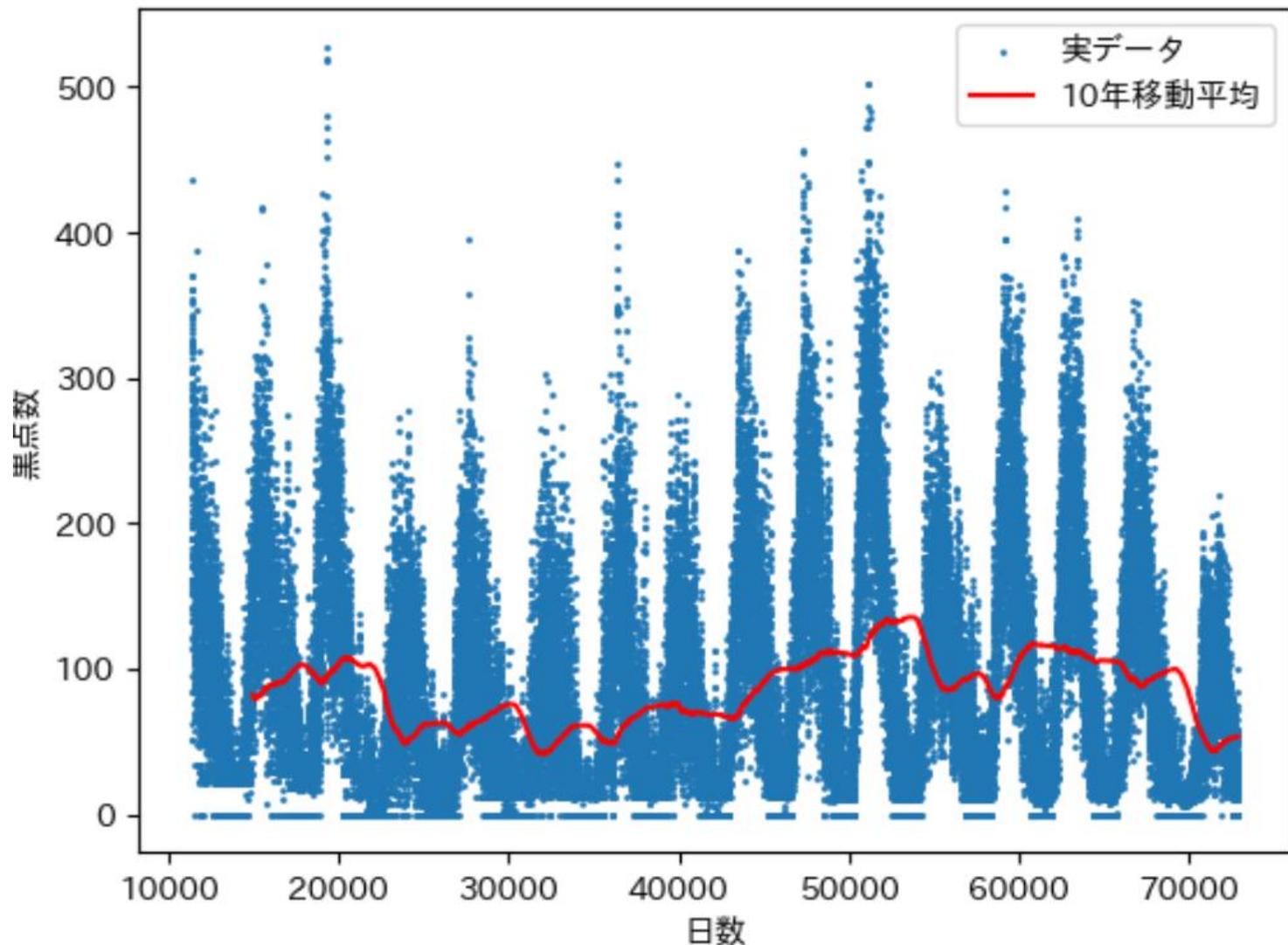
太陽黒点数の散布図と移動平均



赤線： 365日移動平均

# トレンドの識別の例（太陽黒点データ）

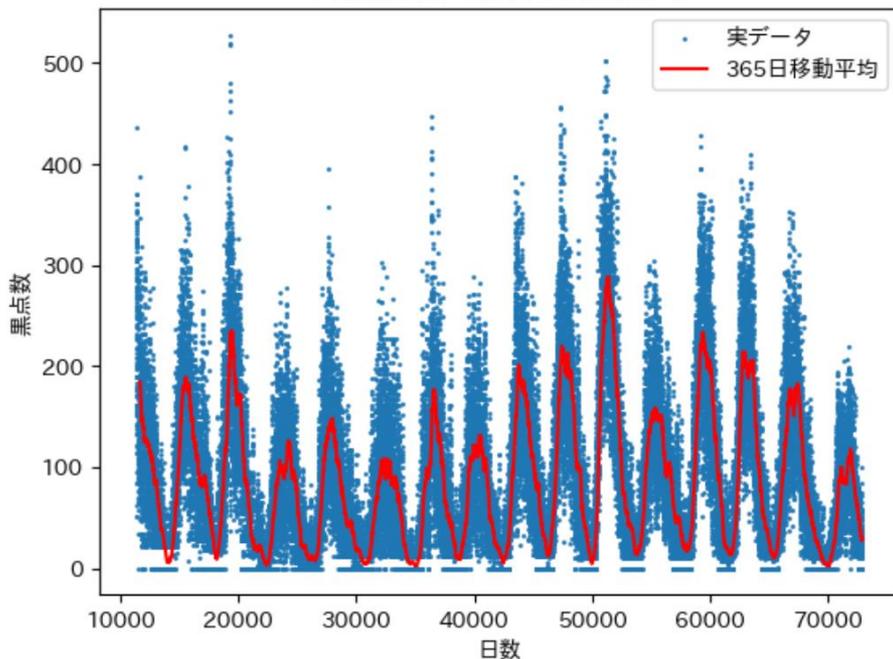
太陽黒点数の散布図と10年移動平均



赤線： 10年移動平均

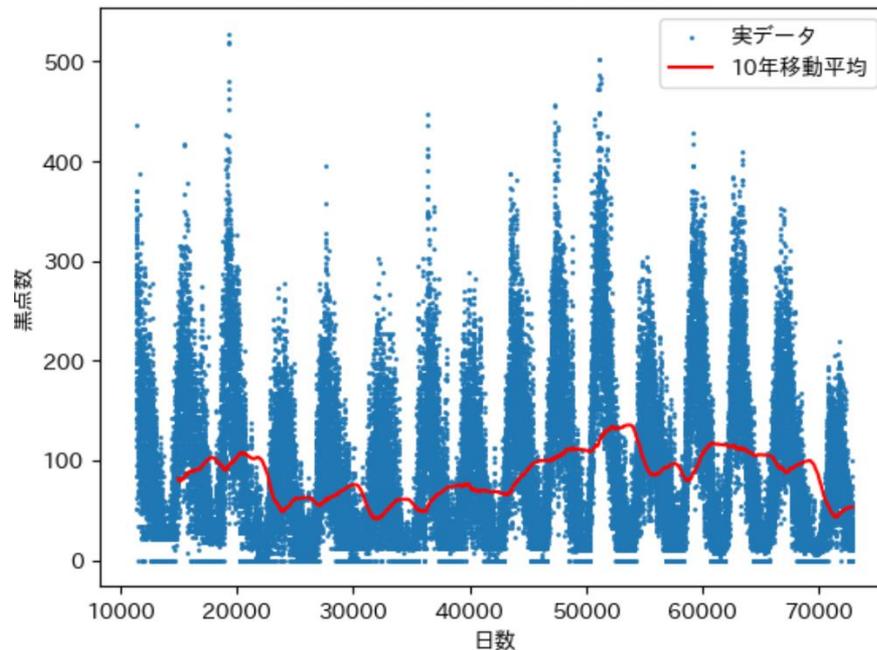
# 移動平均

太陽黒点数の散布図と移動平均



赤線： 365日移動平均

太陽黒点数の散布図と10年移動平均



赤線： 10年移動平均

- **移動平均**は、時系列データにおいて、**特定の期間にわたる平均値**を計算したものの。

例：ある日の移動平均値は、その日を含む過去365日や 3650日のデータポイントの平均値になる。

# 時系列データの特徴

- **周期性**：日、週、月、季節などの**定期的な間隔で繰り返されるパターン**  
例：自然現象、社会的活動
- **トレンド**：時間の経過による**増加、減少、一定レベルの維持などの方向性**
- **特定のイベントや時期**（例えば正月、学校の学期開始時期など）**との関連性**

# Python の Pandas データフレーム

表形式のデータ

	x	y
0	1	4
1	1	2
2	1	5
3	2	4
4	3	5
5	3	3

データ本体

# Pandas の基本的な使い方

- Google Colaboratoryにはpandasが**既にインストール**されている。
- **インポート**: pandasを使用する前に、「import pandas as pd」でインポート。

## 主要なメソッド

- **CSVファイルの読み込み**: read\_csvメソッド。  
例: `pd.read_csv('file.csv')`
- **Excelファイルの読み込み**: read\_excelメソッド。  
例: `pd.read_excel('file.xlsx')`
- **データの先頭部分、末尾部分を表示**: head(), tail() メソッド。  
例: `df.head()` `df.tail()`
- **1列の選択**: [ ]を使用して列を選択。  
例: `df['column_name']`
- **複数列の選択**: [[ ]]を使用して列を選択。  
例: `df[['a1', 'a2']]`



## 演習 1. 移動平均

### 【トピックス】

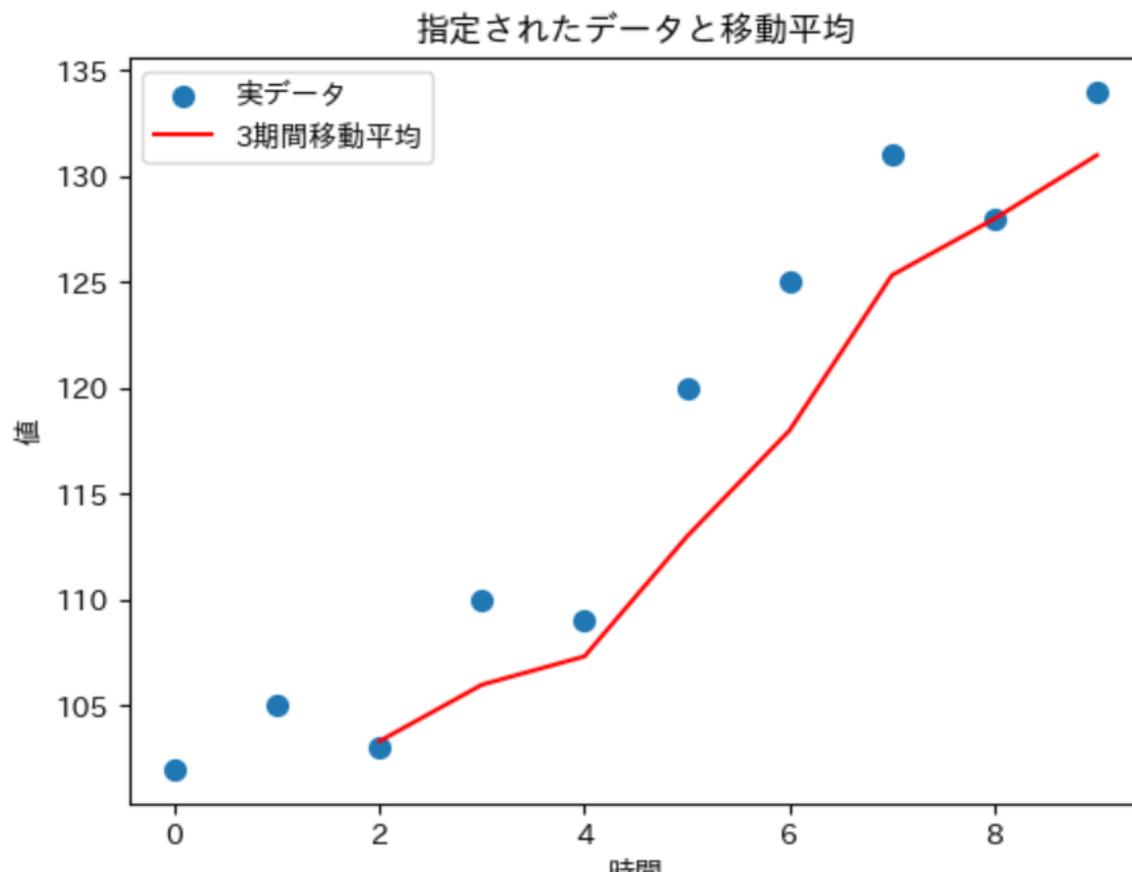
- Pandasデータフレーム
- 移動平均
- 散布図のプロット

## ① Google Colaboratory のページを開く

<https://colab.research.google.com/drive/1qXH5l0iEPUm-QRTuEBSqBLd3V9KvqltK?usp=sharing>

## ② 「9 - 2. 移動平均」を確認

説明とプログラムと実行結果を確認



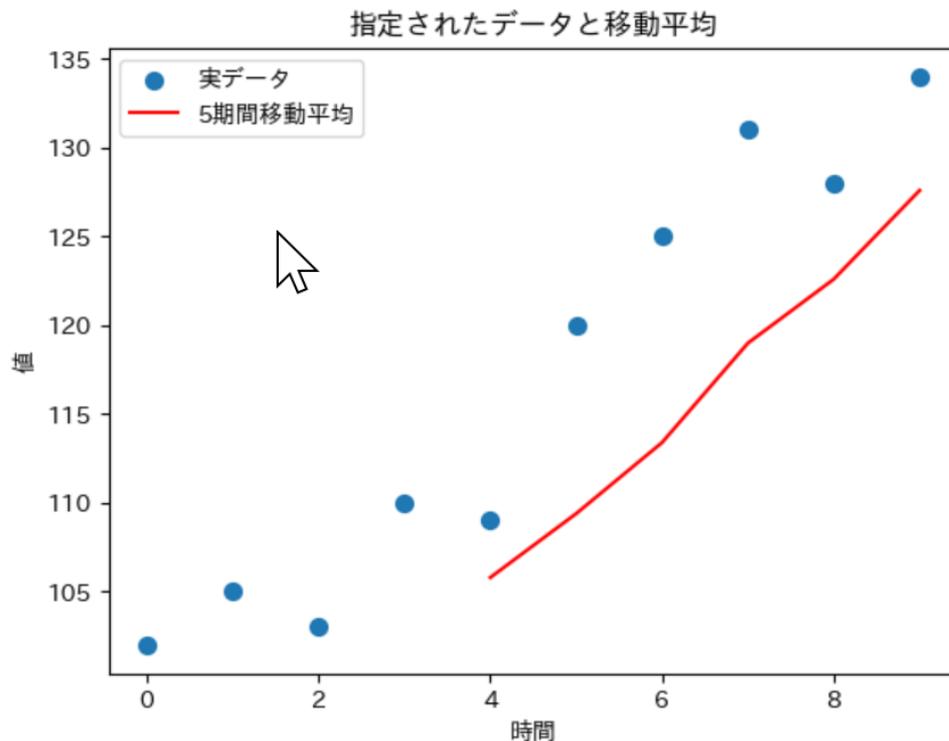
# 自習 1



目的：移動平均の理解を深める

## 演習 1 のページを使用

- コードセルの「rolling\_window = 3」を「rolling\_window = 5」など、より大きな値に書き換える（値は10以上にしないこと）
- 値を大きくすると、より滑らかになることを確認



## 演習 2. 時系列データ

### 【トピックス】

- Pandasデータフレーム
- 時系列データ
- 移動平均
- 散布図のプロット

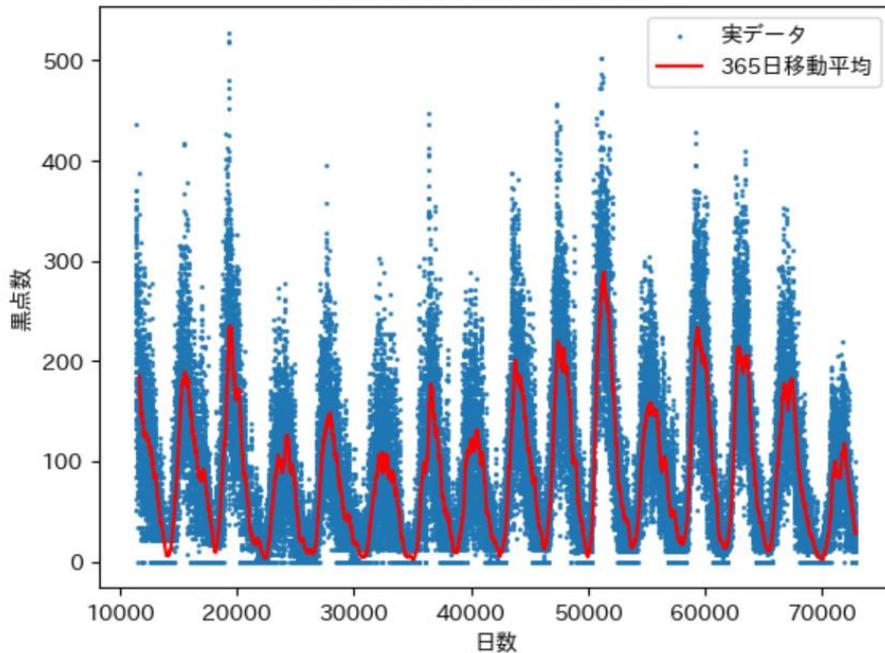
① Google Colaboratory のページを開く

<https://colab.research.google.com/drive/1qXH5l0iEPUm-QRTuEBSqBLd3V9KvqltK?usp=sharing>

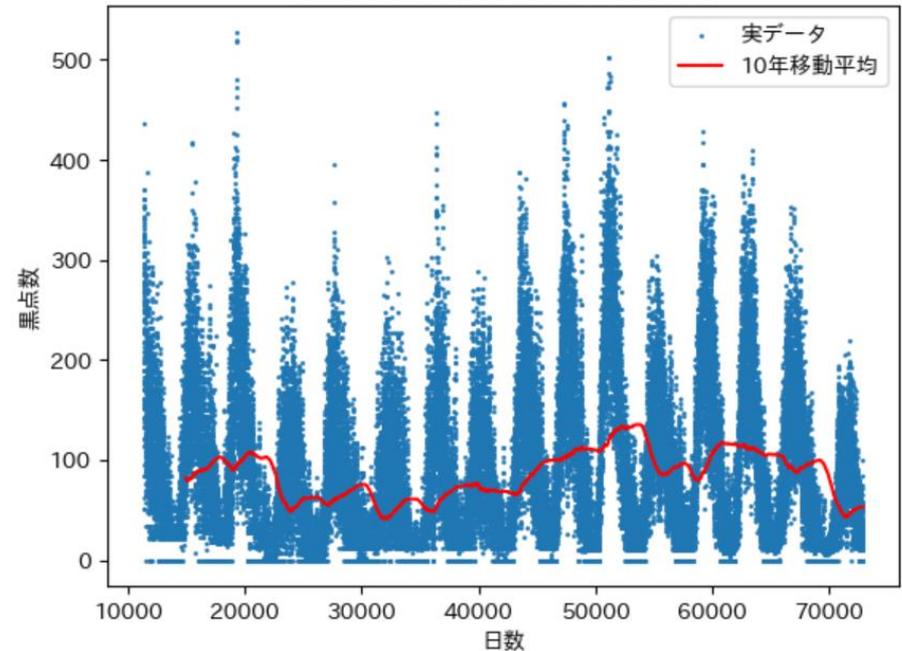
② 「9 - 3. 時系列データ分析の概要」を最後まで確認

説明とプログラムと実行結果を確認

太陽黒点数の散布図と移動平均



太陽黒点数の散布図と10年移動平均



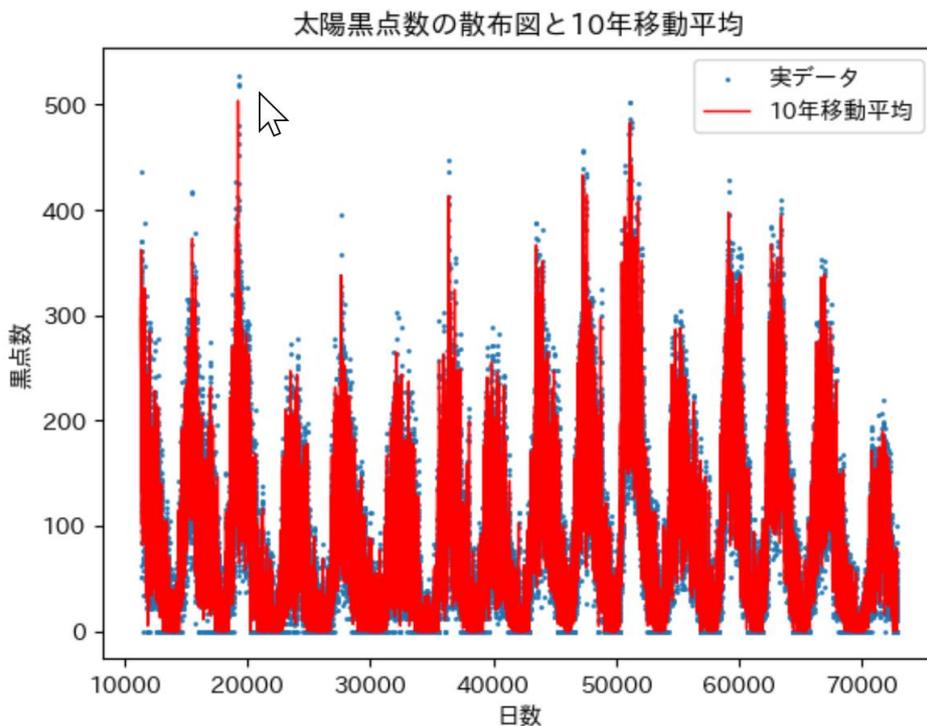
## 自習 2



目的：移動平均の理解を深める

### 演習 2 のページを使用

- コードセルの「**rolling\_window = 3650**」を「**rolling\_window = 5**」、「**rolling\_window = 50**」、「**rolling\_window = 500**」など、より大きな値に書き換える
- 値を大きくすると、より滑らかになることを確認



時系列データは、時間の経過に伴って順序付けられたデータ  
の並び

## 時系列データの特徴

- **周期性**：日、週、月、季節などの**定期的な間隔で繰り返されるパターン**  
例：自然現象、社会的活動
- **トレンド**：時間の経過による**増加、減少、一定レベルの維持などの方向性**
- **特定のイベントや時期**（例えば正月、学校の学期開始時期など）との**関連性**

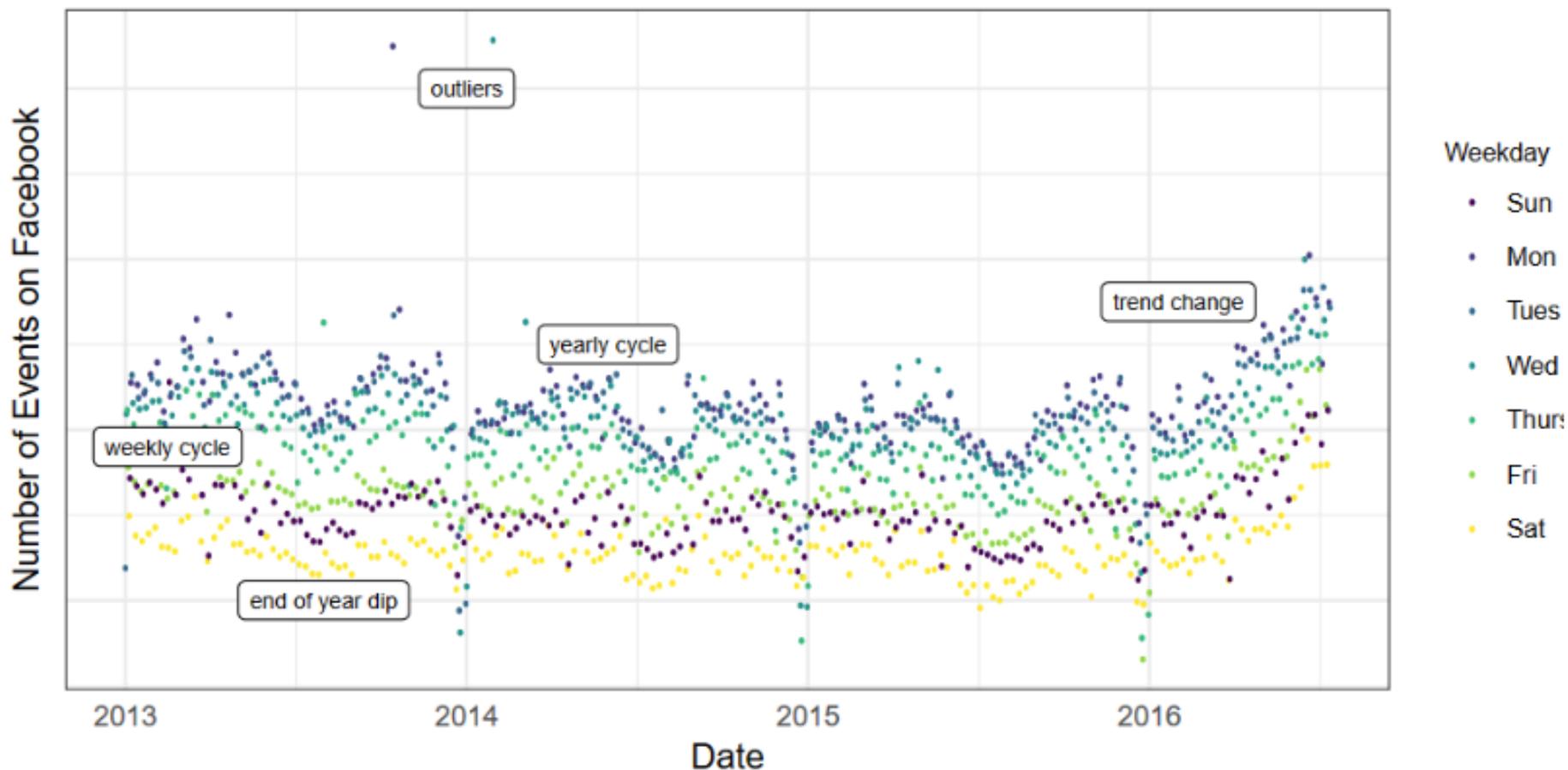
## 9-4. 時系列データの分析 (Prophetを使用)

# Prophet



- Prophet は Python のライブラリ
- **時系列データ**に対して、**周期性、トレンド、特定のイベントや時期との関連性を分析**する機能を持つ
- 統計手法を基礎とする。線形モデルと非線形モデルを組み合わせた回帰モデル。
- ディープラーニングではない

# Facebook イベント数の時系列データ

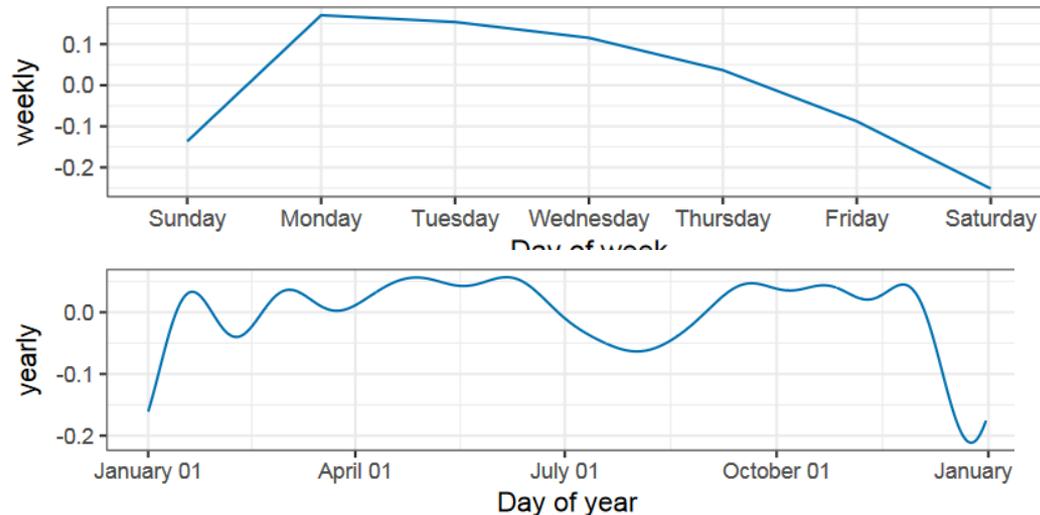


曜日で色を変えて**プロット** ⇒ 曜日単の周期性を読み取る

Taylor SJ, Letham B. 2017. Forecasting at scale. PeerJ Preprints 5:e3190v2  
<https://doi.org/10.7287/peerj.preprints.3190v2>

## 周期性の分析

元データをProphet で処理した結果

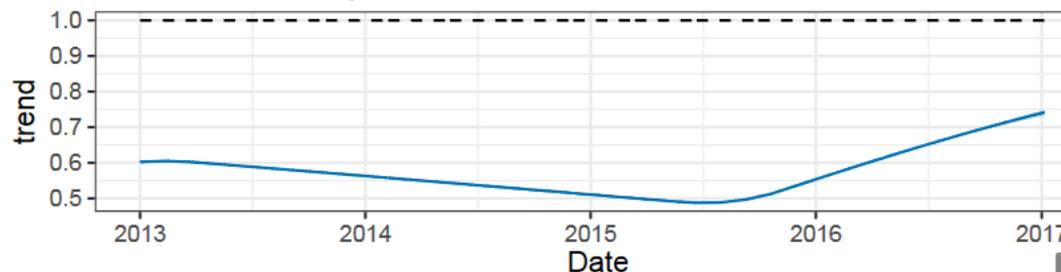


土日は少ない

年末年始は少ない

## トレンドの分析

元データをProphet で処理した結果



2015年からは増加



## 演習 3. トレンドの分析

### 【トピックス】

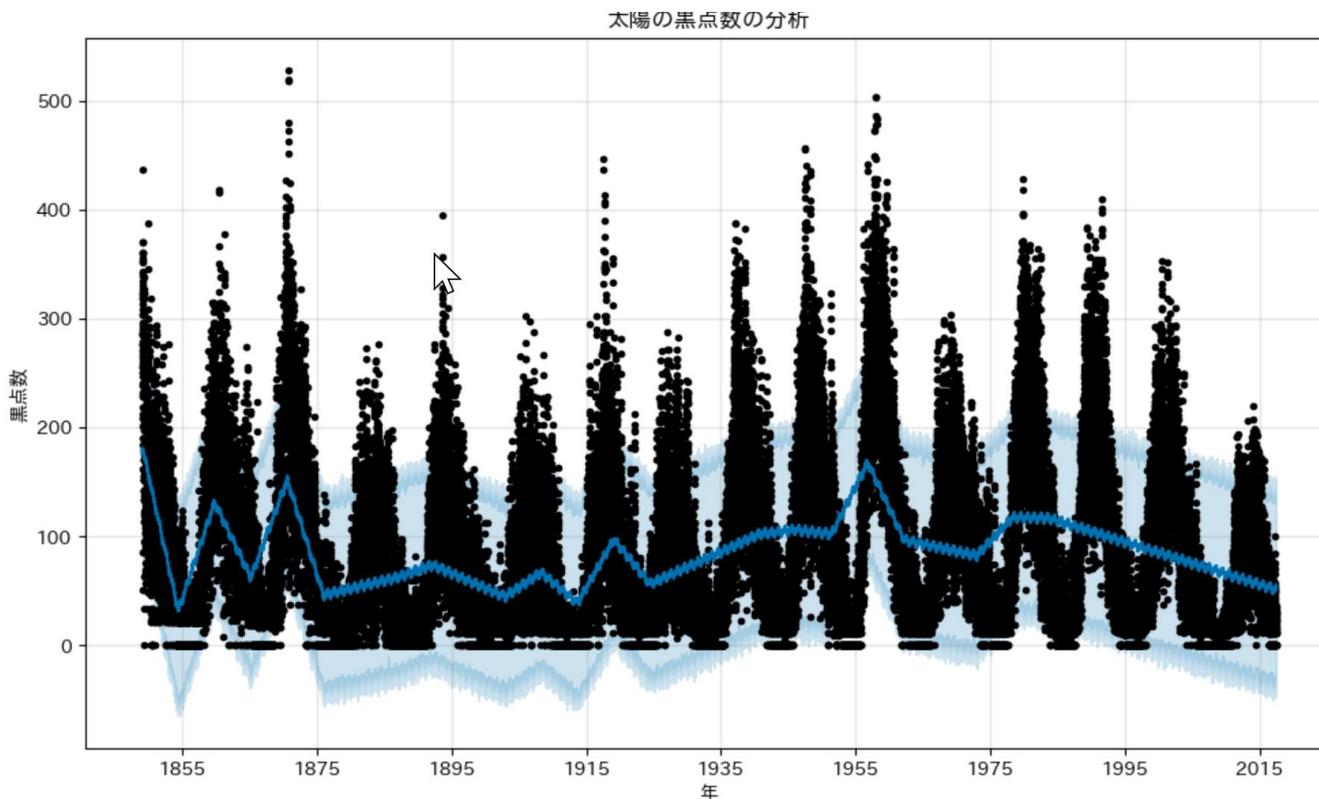
- 時系列データ
- Prophet

① Google Colaboratory のページを開く

<https://colab.research.google.com/drive/1qXH5l0iEPUm-QRTuEBSqBLd3V9KvqItK?usp=sharing>

② 「9 - 4. 時系列データの分析 (Phophet を使用)」を確認

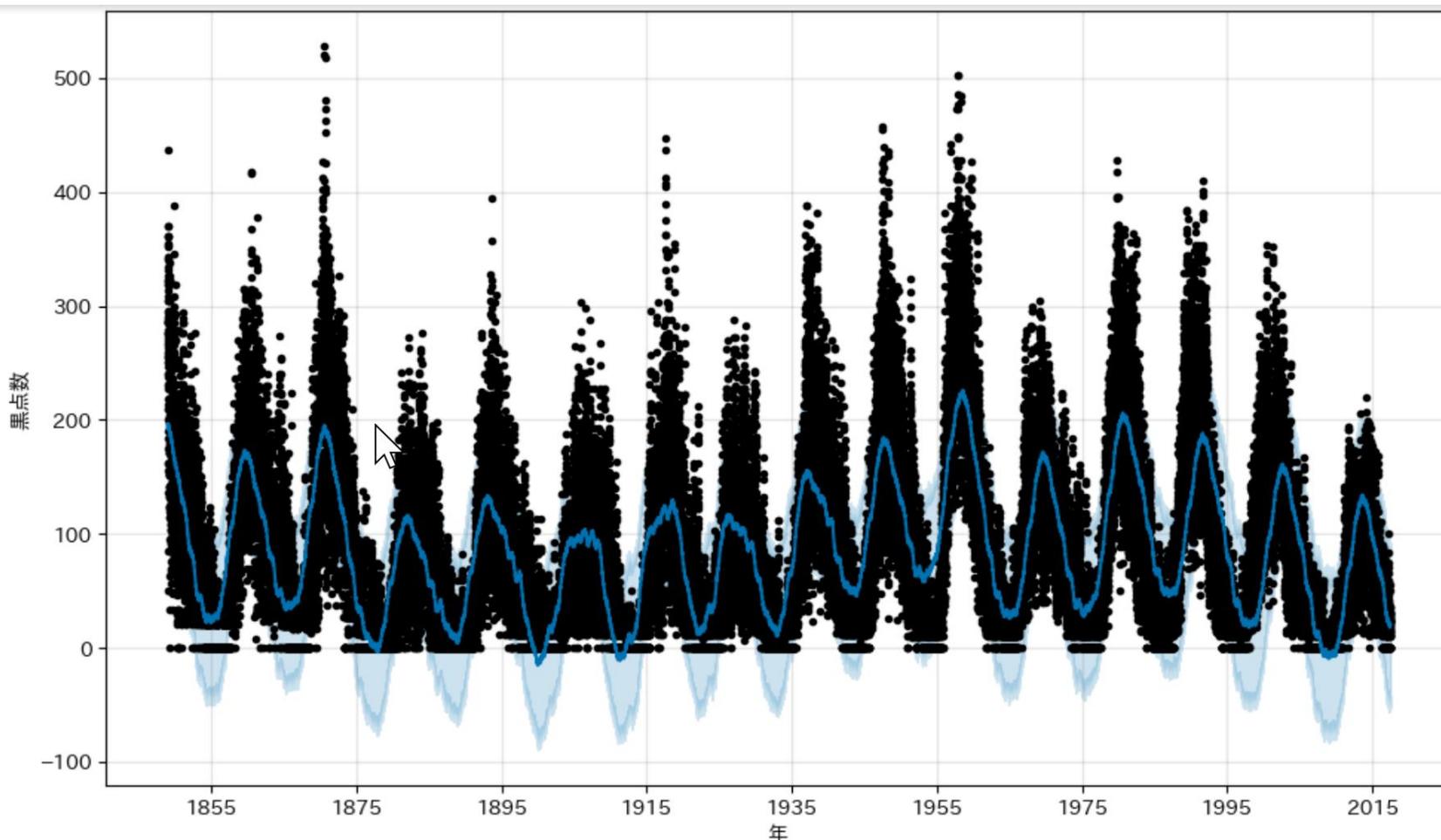
1 つめの説明とプログラムと実行結果を確認



トレンド分析  
ができる

③今度は、「(2) 11年周期の季節性を  
add\_seasonalityメソッドを使ってモデルに追加。  
を確認

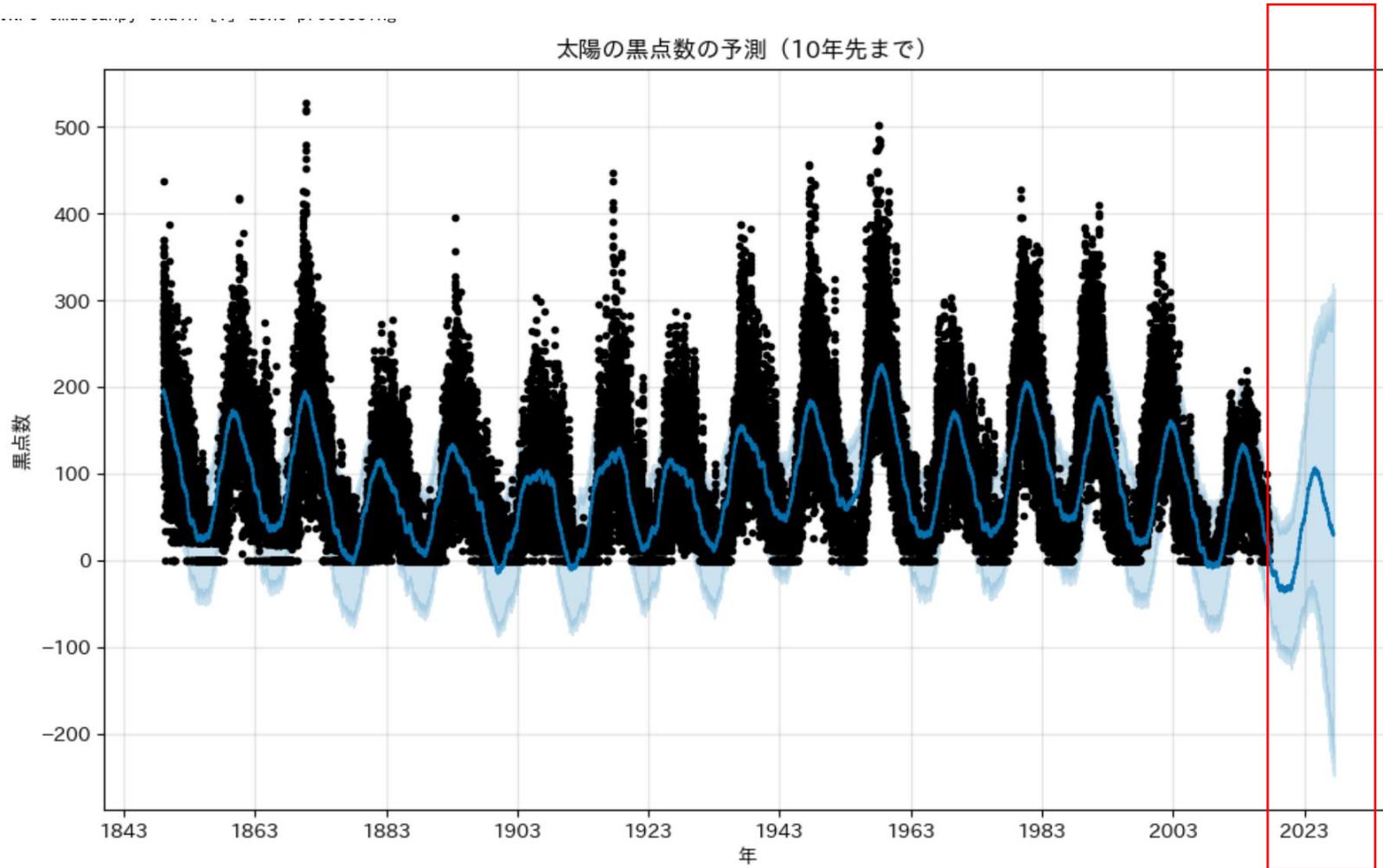
より精密な分析ができている



# ④今度は、「(3) Prophet を用いた将来予測」を確認



## より精密な分析ができています



# Prophet まとめ



- **時系列データ**に対して、**周期性、トレンド、特定のイベントや時期との関連性を分析する機能を持つ**
- 統計手法を基礎とする。線形モデルと非線形モデルを組み合わせた回帰モデル。
- ディープラーニングではない

## 9-4. リカレントニューラル ネットワークの基本

# ディープラーニングによる将来予測



ディープラーニングにより、**過去の時系列データを学習**することで、将来予測を行う

- **大量のデータが必要**である
- **予測の正確性は保証されない**

## 手法

- **リカレントニューラルネットワーク (RNN)** : 時系列データや、その他データの並びに**適したニューラルネットワーク**の一種
- **長短期記憶 (LSTM)**: RNNの一種. 長期的な依存関係を学習する能力
- **トランスフォーマー (Transformer)** : 大規模なシーケンスデータの処理

## 統計手法との比較

- ディープラーニングも統計手法（例えばProphet）も**過去の時系列データを用いる**
- 予測メカニズムは**異なる**
- ディープラーニングはより複雑なパターンをとらえ、**より高精度の予測を行う可能性がある**

# リカレントニューラルネットワーク (RNN)

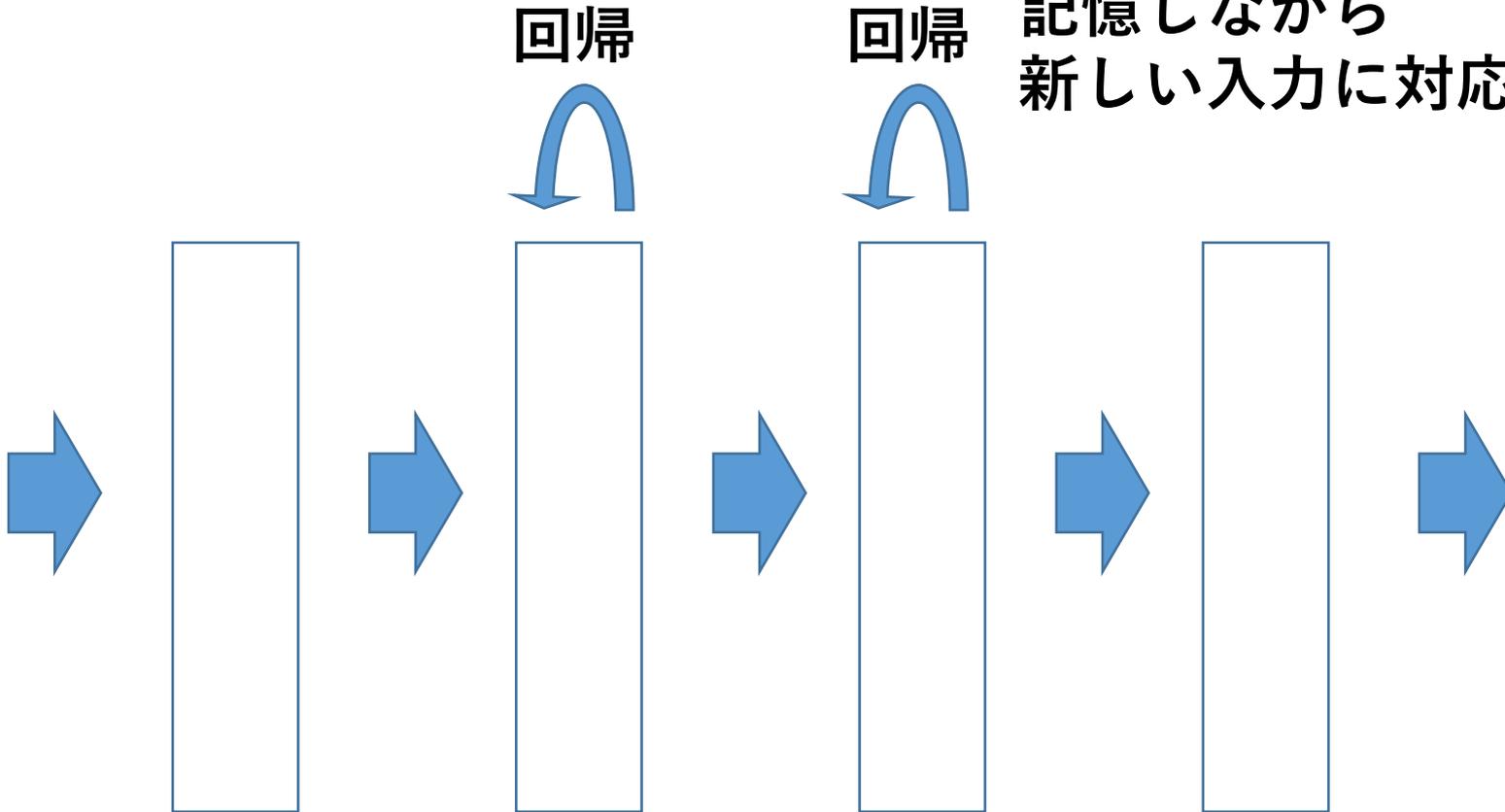


- **リカレントニューラルネットワーク (RNN) は、時系列データや、その他データの並びに適したニューラルネットワークの一種**
- **回帰のしくみが導入されている**
- **前回の実行時での結果を記憶しながら新しい入力に対応**

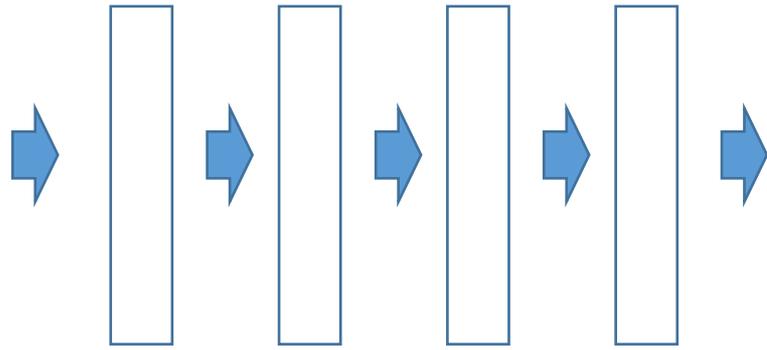
# リカレントニューラルネットワーク

## 回帰のしくみ

前回の実行時での結果を  
記憶しながら  
新しい入力に対応

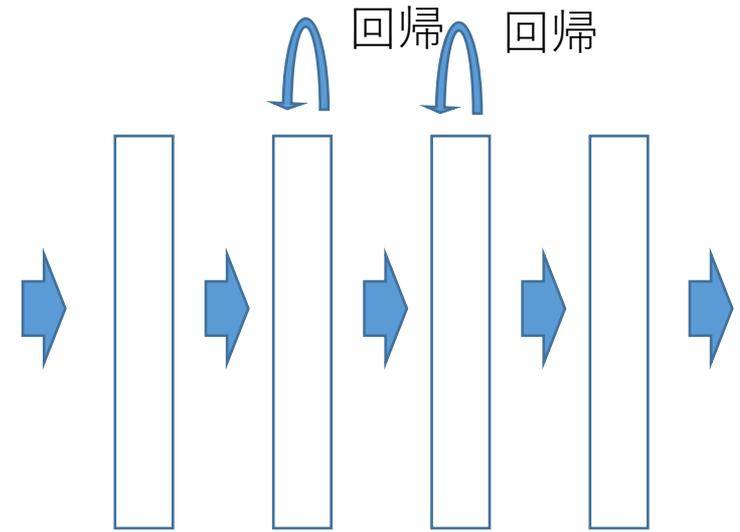


# フィードフォワードとリカレントニューラルネットワーク



## フィードフォワード

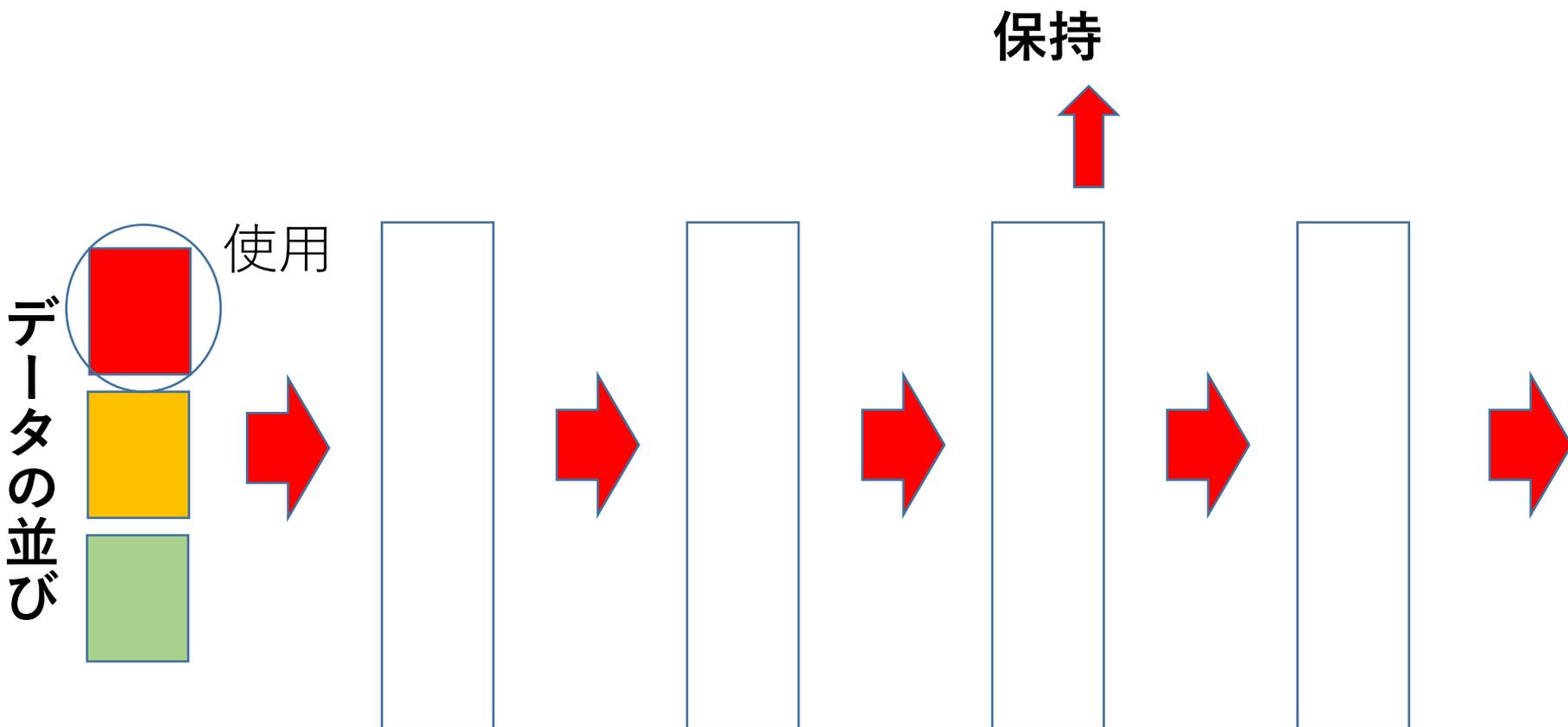
ある層の出力を，次の層が受け取る



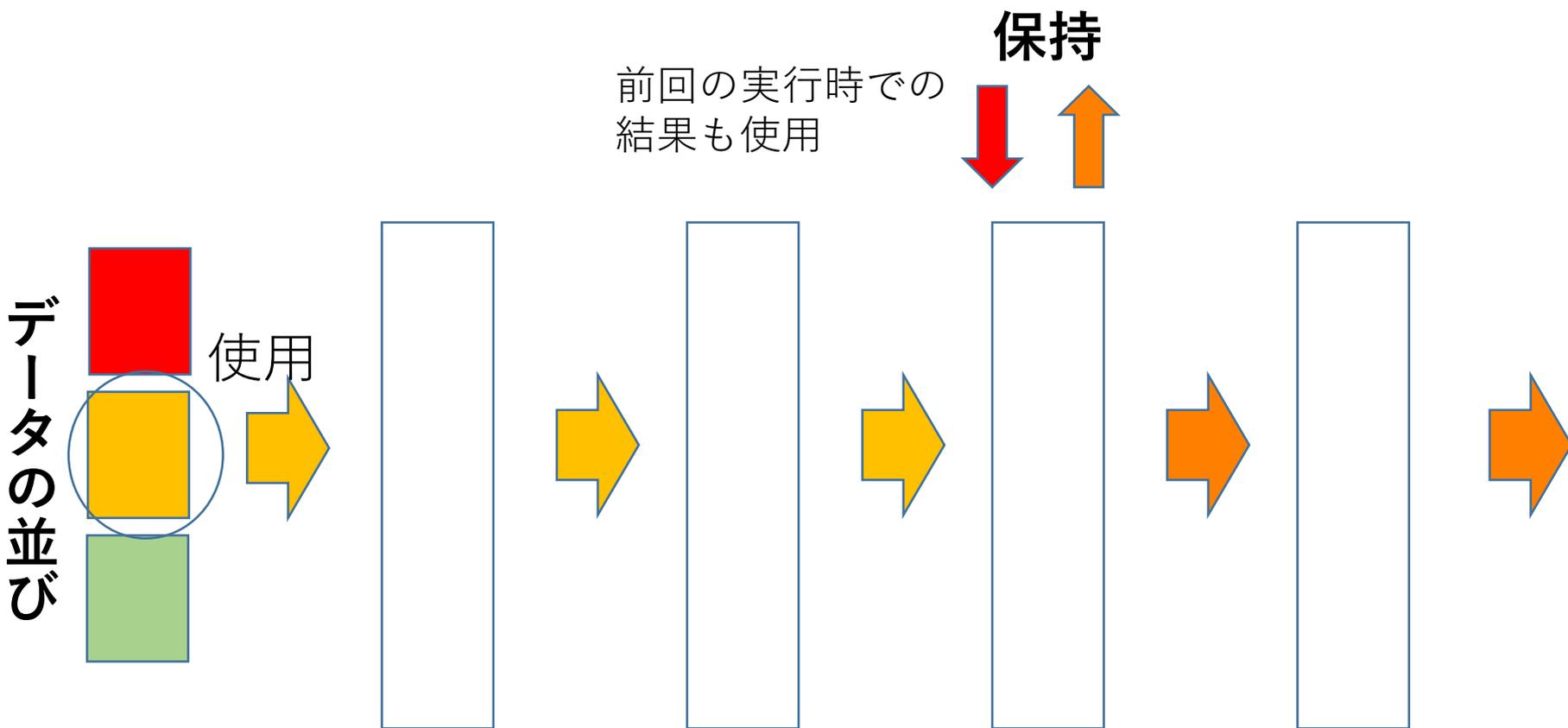
## リカレントニューラルネットワーク

**回帰**により，前回の実行時での結果の一部が，次の実行に反映される。

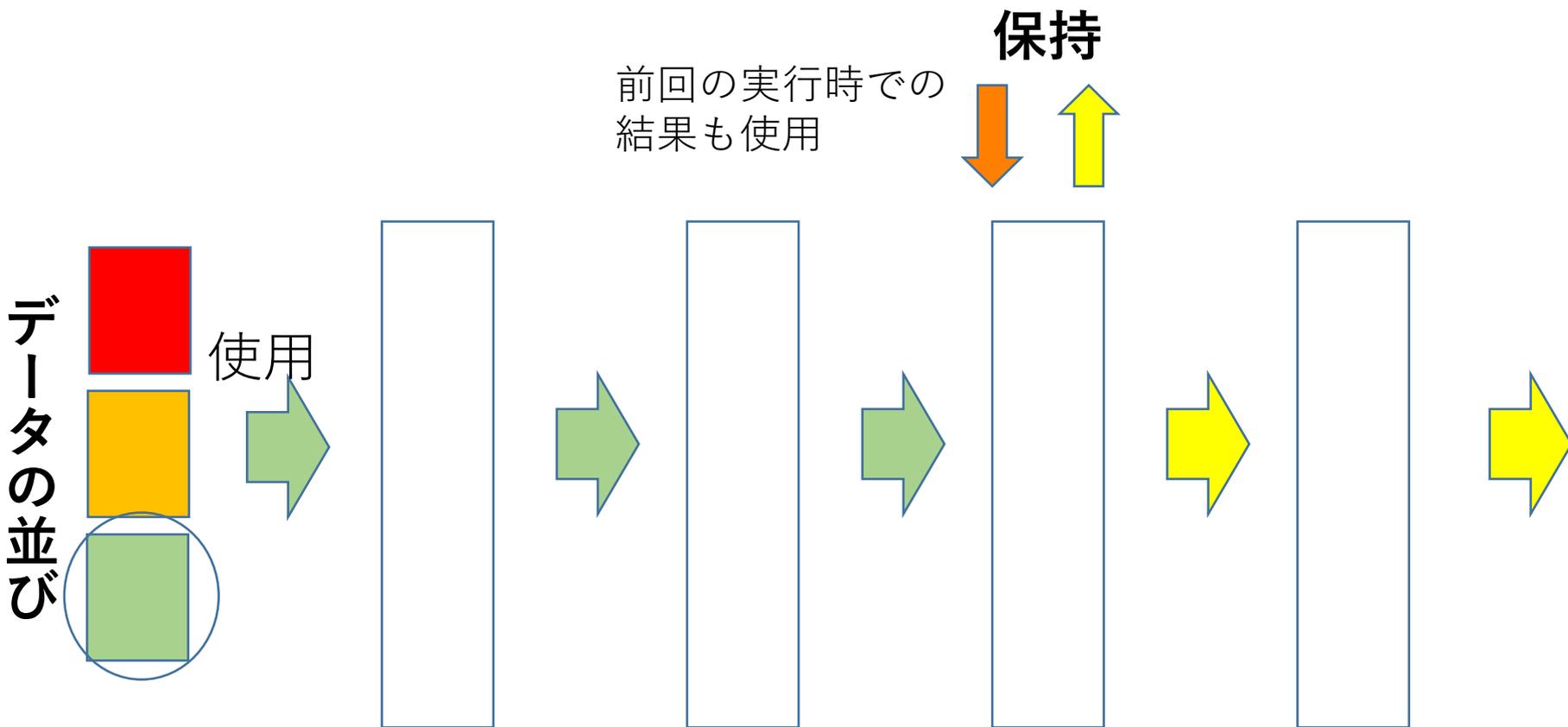
# リカレントニューラルネットワークの動作イメージ ①



# リカレントニューラルネットワークの動作イメージ ②



# リカレントニューラルネットワークの動作イメージ ③



# リカレントニューラルネットワークの応用

- 時系列データの将来予測
- 手書き文字認識
- 音声認識

# リカレントニューラルネットワークのニュース

## 数学に関する教科書を人工知能が学習。数学論文の偽物等を生成する能力を獲得

For  $\bigoplus_{n=1, \dots, m}$  where  $\mathcal{L}_{m,*} = 0$ , hence we can find a closed subset  $\mathcal{H}$  in  $\mathcal{H}$  and any sets  $\mathcal{F}$  on  $X$ ,  $U$  is a closed immersion of  $S$ , then  $U \rightarrow T$  is a separated algebraic space.

*Proof.* Proof of (1). It also start we get

$$S = \text{Spec}(R) = U \times_X U \times_X U$$

and the comparico in the fibre product covering we have to prove the lemma generated by  $\coprod Z \times_U U \rightarrow V$ . Consider the maps  $M$  along the set of points  $\text{Sch}_{\text{fppf}}$  and  $U \rightarrow U$  is the fibre category of  $S$  in  $U$  in Section, ?? and the fact that any  $U$  affine, see Morphisms, Lemma ?? . Hence we obtain a scheme  $S$  and any open subset  $W \subset U$  in  $\text{Sh}(G)$  such that  $\text{Spec}(R') \rightarrow S$  is smooth or an

$$U = \bigcup U_i \times_{S_i} U_i$$

which has a nonzero morphism we may assume that  $f_i$  is of finite presentation over  $S$ . We claim that  $\mathcal{O}_{X,x}$  is a scheme where  $x, x', x'' \in S'$  such that  $\mathcal{O}_{X,x'} \rightarrow \mathcal{O}_{X',x'}$  is separated. By Algebra, Lemma ?? we can define a map of complexes  $\text{GL}_{S'}(x'/S'')$  and we win.  $\square$

To prove study we see that  $\mathcal{F}|_U$  is a covering of  $\mathcal{X}'$ , and  $\mathcal{T}_i$  is an object of  $\mathcal{F}_{X/S}$  for  $i > 0$  and  $\mathcal{F}_p$  exists and let  $\mathcal{F}_i$  be a presheaf of  $\mathcal{O}_X$ -modules on  $\mathcal{C}$  as a  $\mathcal{F}$ -module. In particular  $\mathcal{F} = U/\mathcal{F}$  we have to show that

$$\widehat{M}^* = \mathcal{I}^* \otimes_{\text{Spec}(k)} \mathcal{O}_{S,s} - i_X^{-1} \mathcal{F}$$

is a unique morphism of algebraic stacks. Note that

$$\text{Arrows} = (\text{Sch}/S)_{\text{fppf}}^{\text{opp}}, (\text{Sch}/S)_{\text{fppf}}$$

and

$$V = \Gamma(S, \mathcal{O}) \mapsto (U, \text{Spec}(A))$$

is an open subset of  $X$ . Thus  $U$  is affine. This is a continuous map of  $X$  is the inverse, the groupoid scheme  $S$ .

*Proof.* See discussion of sheaves of sets.  $\square$

The result for prove any open covering follows from the less of Example ?? . It may replace  $S$  by  $X_{\text{spaces}, \text{étale}}$  which gives an open subspace of  $X$  and  $T$  equal to  $S_{\text{Zar}}$ , see Descent, Lemma ?? . Namely, by Lemma ?? we see that  $R$  is geometrically regular over  $S$ .

```
/*
 * Increment the size file of the new incorrect UI_FILTER group information
 * of the size generatively.
 */
static int indicate_policy(void)
{
    int error;
    if (fd == MARN_EPT) {
        /*
         * The kernel blank will coeld it to userspace.
         */
        if (ss->segment < mem_total)
            unblock_graph_and_set_blocked();
        else
            ret = 1;
        goto bail;
    }
    segaddr = in_SB(in. addr);
    selector = seg / 16;
    setup_works = true;
    for (i = 0; i < blocks; i++) {
        seq = buf[i++];
        bpf = bd->bd.next + i * search;
        if (fd) {
            current = blocked;
        }
    }
    rw->name = "Getjbbregs";
    bprm_self_clearl(&iv->version);
    regs->new = blocks[(BPF_STATS << info->historidac) | PFMR_CLOBATHING_SECONDS << 12];
    return segtable;
}
```

人工知能が生成した数学レポート

人工知能が生成したC言語プログラム

Andrej Karpathy のブログ記事

The Unreasonable Effectiveness of Recurrent Neural Networks. 2015.

<http://karpathy.github.io/2015/05/21/rnn-effectiveness/>

# リカレントニューラルネットワークのまとめ

リカレントニューラルネットワーク（RNN）は、**時系列データ**や、**その他データの並びに適したニューラルネットワーク**の一種

- **回帰**のしくみが導入されている
- **統計手法との比較**：ディープラーニングはより複雑なパターンをとらえ、**より高精度の予測を行う可能性**がある
- **応用例**：  
時系列データの将来予測、手書き文字認識、音声認識

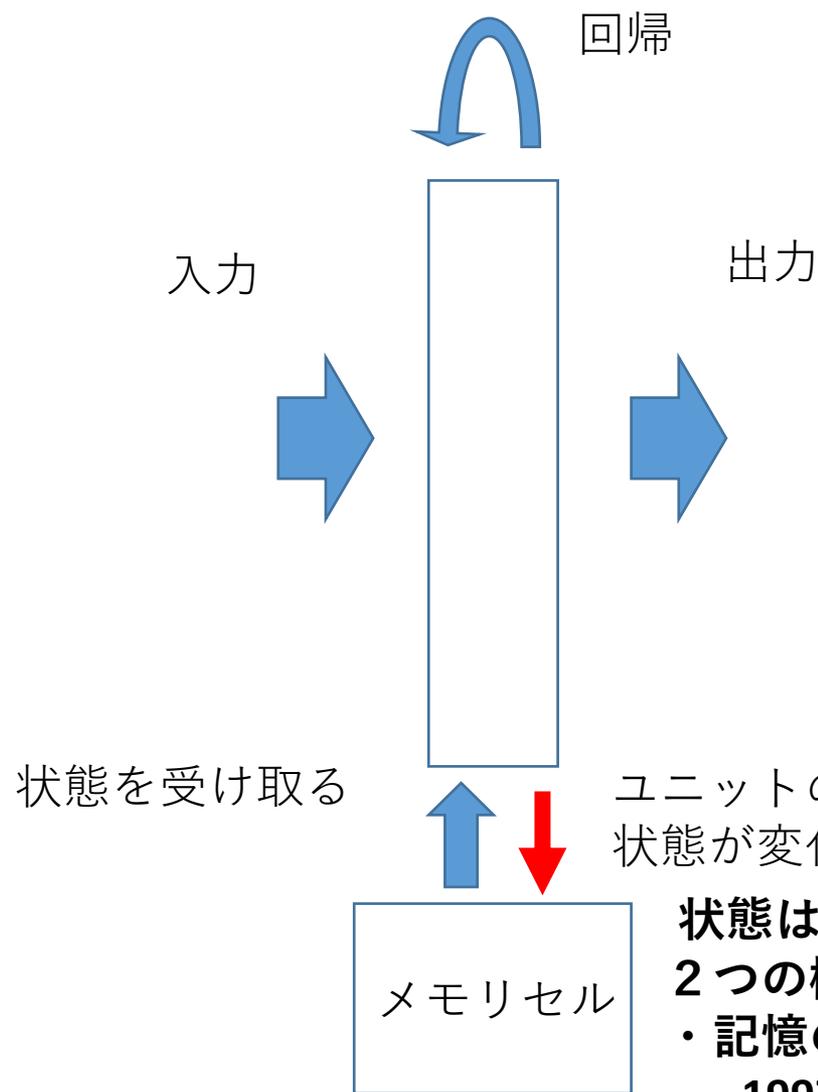


## 9-5. LSTM の基本

# LSTM の特質

- リカレントニューラルネットワーク（RNN）の一種。1997年
- 「**長期的な依存関係の学習が困難**」という、RNNの弱点を改良
- LSTM は、長期の記憶の保持能力を持ち、長期的な依存関係の学習を可能とする。複雑なデータを扱えるように。

# LSTM の仕組み

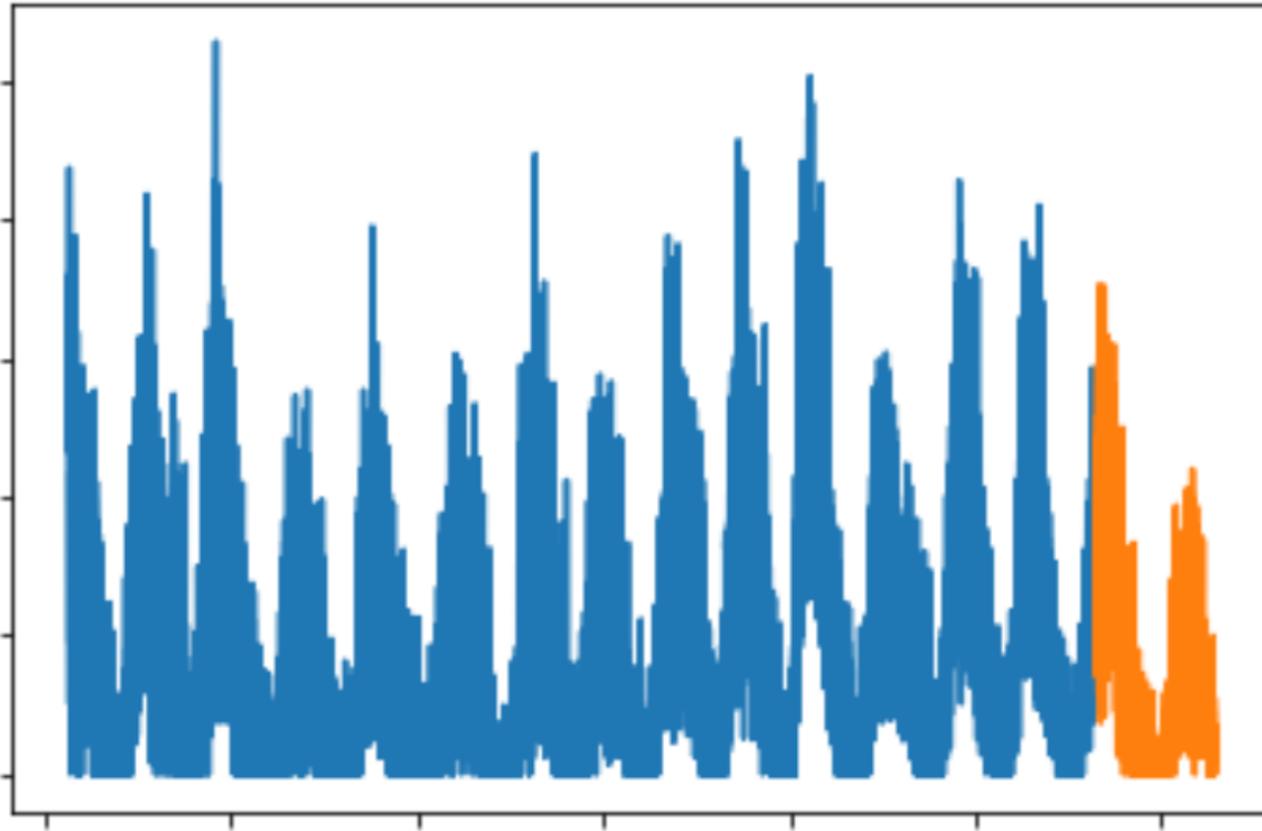


**LSTM のメモリセルは、長期の記憶の保持を可能とする**

状態は「メモリセル」に記憶されている  
2つの機能

- ・ 記憶の持続 (constant error carousel)  
1997年発表
- ・ 記憶の忘却 (forget gate)  
1999年発表

# LSTM を用いた予測の例

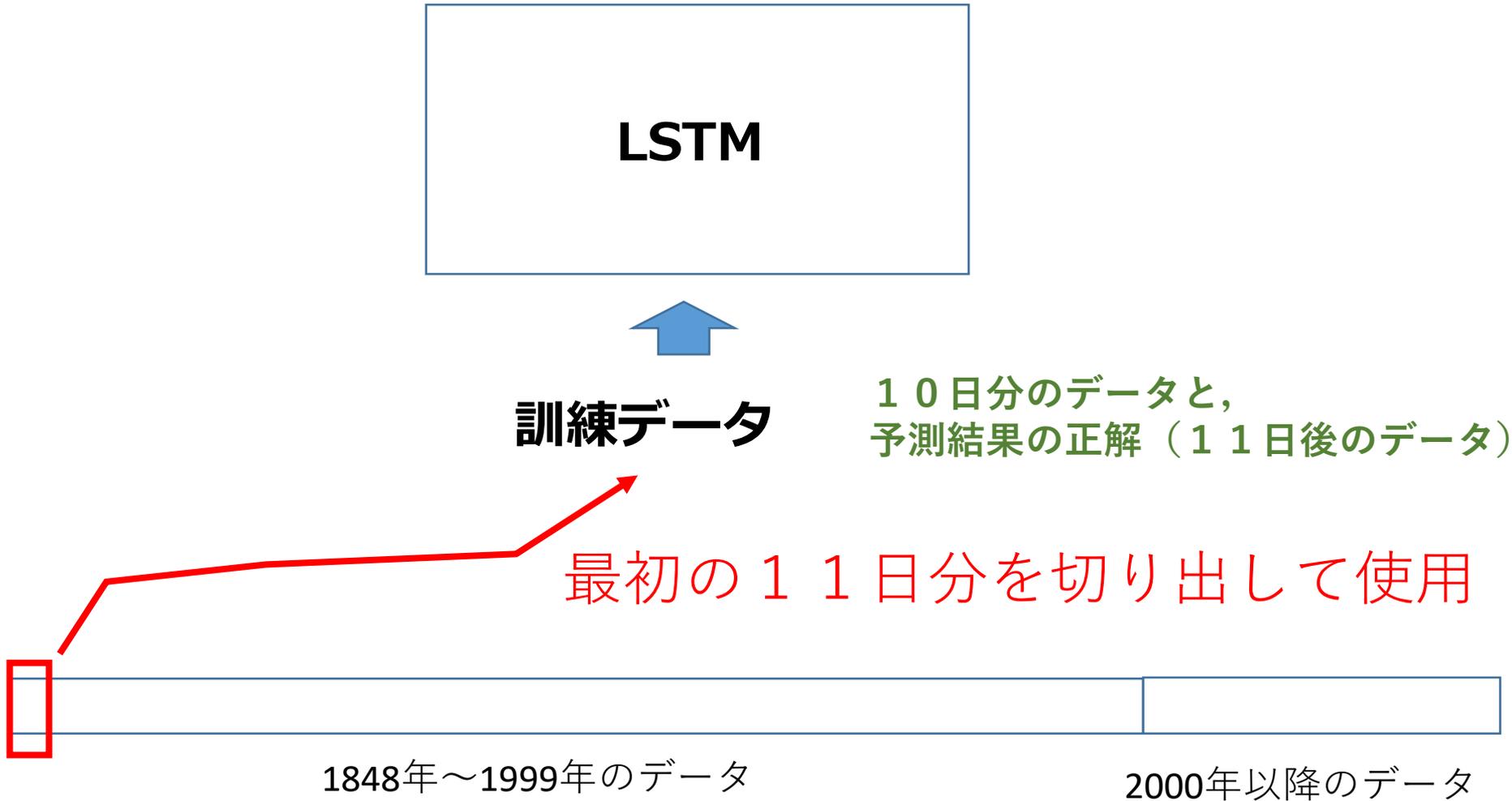


太陽の黒点数の変化

予測

**1848年～1999年のデータ**を用いて、**2000年以降を予測**  
(ディープニューラルネットワークによる予測)

①



LSTM での学習の繰り返し

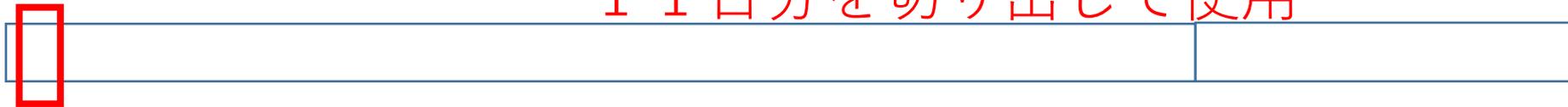
②



訓練データ

10日分のデータと、  
予測結果の正解（11日後のデータ）

少しずつ、  
11日分を切り出して使用



1848年～1999年のデータ

2000年以降のデータ

LSTM での学習の繰り返し

③



訓練データ

10日分のデータと、  
予測結果の正解（11日後のデータ）

最後まで使い切って  
学習を終了

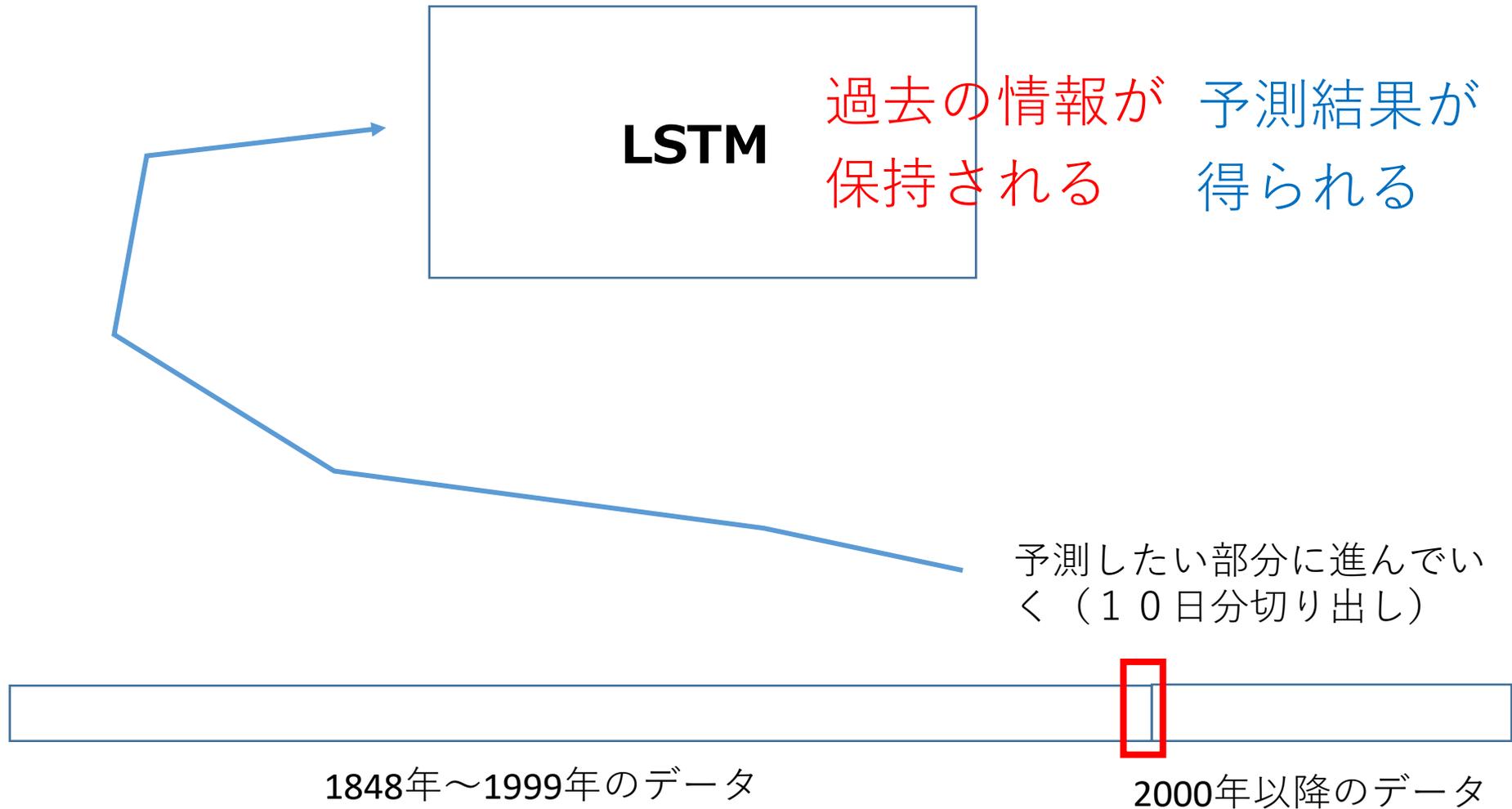


1848年～1999年のデータ

2000年以降のデータ

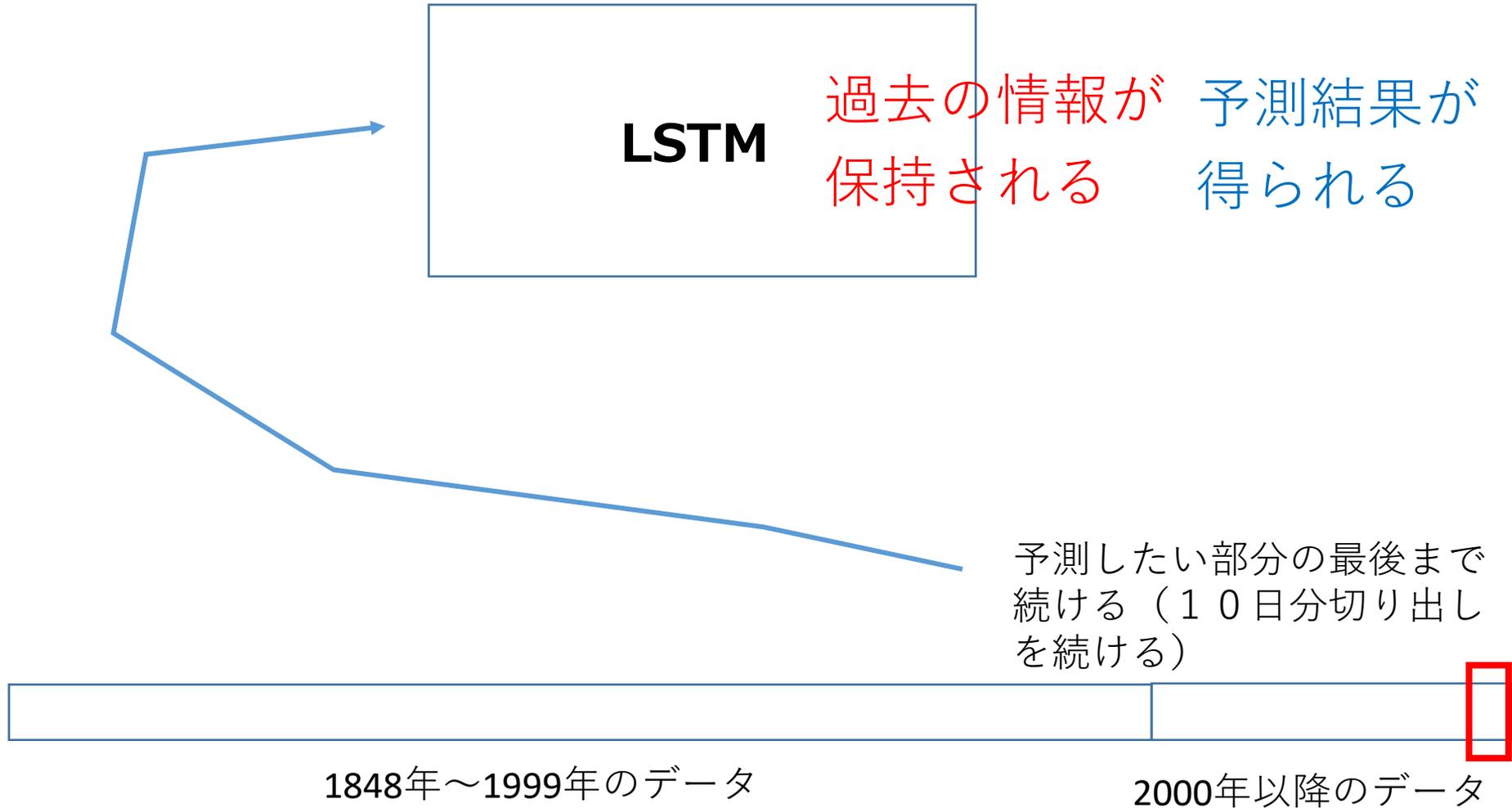
LSTM での学習の繰り返し

④



## LSTMによる予測

⑤



## LSTMによる予測



## 演習 4 . LSTM による将来予測

### 【トピックス】

- 時系列データ
- ディープラーニング
- 将来予測
- 太陽黒点データ
- LSTM

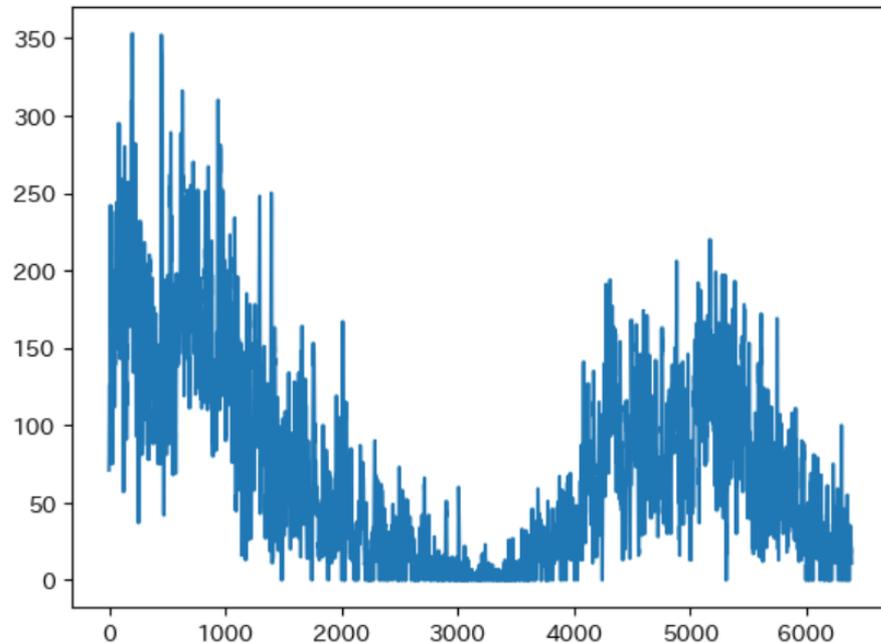
## ① Google Colaboratory のページを開く

<https://colab.research.google.com/drive/1qXH5lOiEPUm-QRTuEBSqBLd3V9KvqltK?usp=sharing>

## ② 「9 - 5. LSTM の基本」を確認

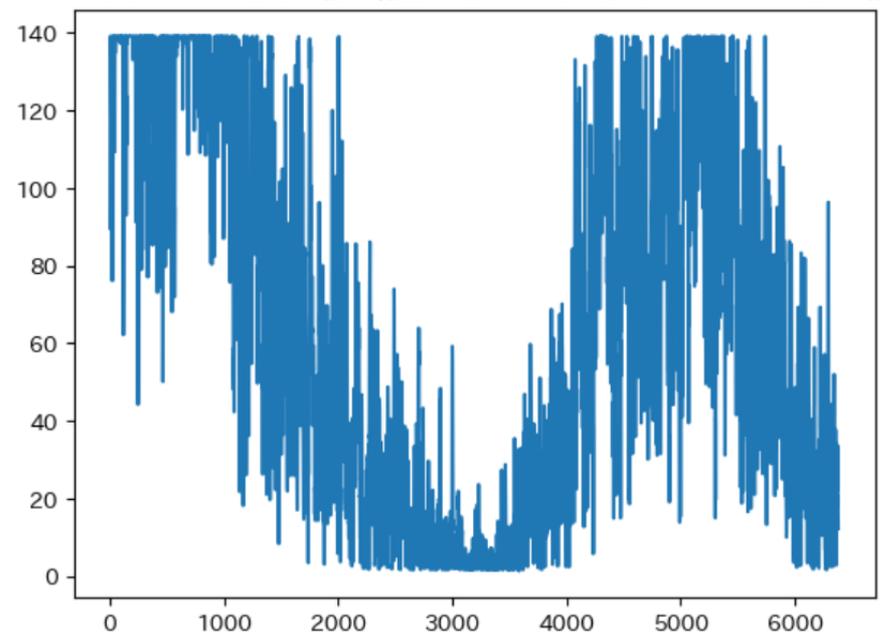
説明とプログラムと実行結果を確認

2000年以降の黒点数（実測値）



実際のデータ

1848年から1999年までの黒点数をもとに AI が予測した 2000年以降の黒点数



過去データからの予測

# 全体まとめ

## 時系列データ分析の基礎

- **時系列データは、時間の経過に伴って順序付けられたデータの並び**
- 特徴は、**周期性、トレンド、特定のイベントや時期との関連性**

## 平均と移動平均

- 平均は、**データの「中心」である**。データの全体的な傾向を把握できる
- **移動平均**は、時系列データにおいて、**特定の期間にわたる平均値**を計算したものの。短期的な変動が滑らかになり、トレンドや周期性を分析しやすくなる。

## Prophetライブラリ

- **時系列データに対して、周期性、トレンド、特定のイベントや時期との関連性を分析する機能を持つ**
- 統計手法を基礎とする。線形モデルと非線形モデルを組み合わせた回帰モデル。

## ディープラーニングによる将来予測

- **リカレントニューラルネットワーク (RNN) : 時系列データや、その他データの並びに適したニューラルネットワークの一種**
- **長短期記憶 (LSTM): RNNの一種。長期的な依存関係を学習する能力**
- **統計手法との比較 : ディープラーニングはより複雑なパターンをとらえ、より高精度の予測を行う可能性がある**

## 自習① 移動平均の理解

**目的:** 移動平均の基本概念を理解

この資料やインターネットで「**移動平均**」とは何かをより深く調べ、その**定義と計算方法を理解する**。次に、自分の選んだ**時系列データに移動平均をあてはめるとどのようになりそうか、考察する**

ヒント：

**移動平均**は、時系列データにおいて、**特定の期間にわたる平均値**を計算したもの。短期的な変動が滑らかになり、トレンドや周期性を分析しやすくなる

例：株価の移動平均からトレンドが見やすくなる

## 自習② RNNとLSTM の基本

**目的:**リカレントニューラルネットワーク（RNN）と LSTM の基本的な概念と違いを理解

この資料やインターネットで、**RNNとLSTMについて調べ**、それぞれの基本的な概念、機能、および違いを理解する。

ヒント：LSTMはRNNの一種で、RNNの弱点について改良が行われている

## 自習③ 時系列データ分析

**目的:**時系列データの特徴を理解

この資料やインターネットで**時系列データの特徴（周期性、トレンド、特定のイベントや時期との関連性など）**について調べ、それらをどのように分析するかをリサーチする

ヒント：

周期性（日々の変動など）

トレンド（長期的な増加や減少や一定量の維持）

特定のイベントや時期との関連性（正月は多いなど）