



# 機械学習を活用した 台風被害の予測研究

大同火災海上保険株式会社  
赤嶺 康友  
玉城 博史

※本稿の内容は、発表者の所属法人の見解を表すものではありません。

# AGENDA

1. 研究の動機
2. データサイエンスについて
3. 機械学習の概要と台風分析について
4. ディープラーニングの概要と台風分析について
5. まとめ

# AGENDA

## 1. 研究の動機

- (1) 沖縄県の台風リスクについて
- (2) 台風被害予測モデルの概要
- (3) 今回の研究内容について

# 1. 研究の動機

## (1) 沖縄県の台風リスクについて

・沖縄県は日本に接近する台風の約半数を占めている台風常襲地域である。

### ■年間台風接近数

年	2011	2012	2013	2014	2015	2016	2017	2018	平均
沖縄地方	7	11	9	10	6	7	7	13	8.8
本土*	5	6	6	5	6	9	5	10	6.5

\*北海道、本州、四国、九州

出典：気象庁HP



出典：当社HP（2003年台風14号（マエミー）襲来後の宮古島の様子）

# 1. 研究の動機

## (1) 沖縄県の台風リスクについて

- ・過去には大規模な被害をもたらした台風も多く発生している。
- ・特に2018年の台風24号では約118億の保険金支払が発生しており、沖縄県において過去最大の支払額を記録した。

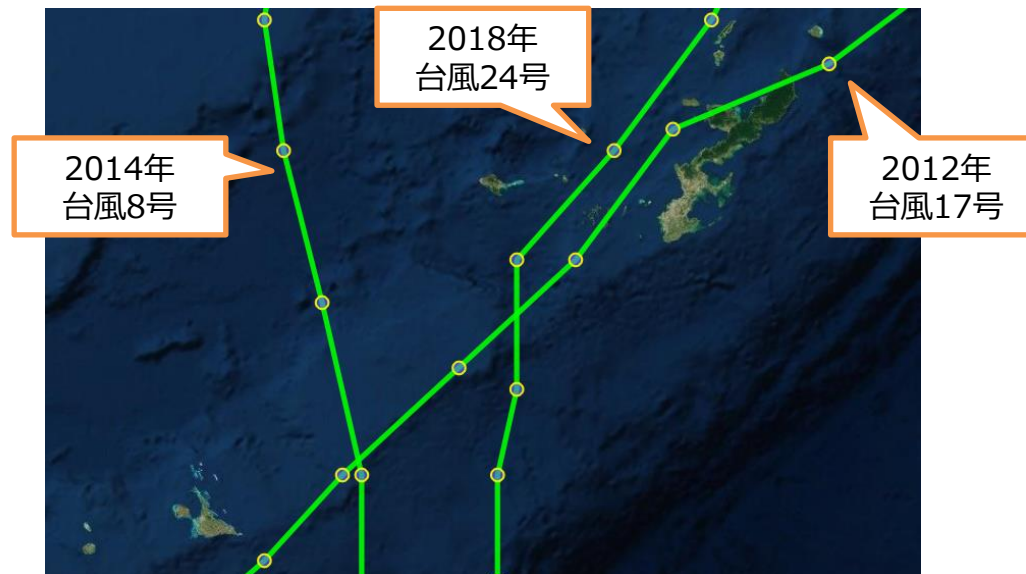
	火災保険		自動車保険		その他		合計	
	件数	支払保険金	件数	支払保険金	件数	支払保険金	件数	支払保険金
2018年 台風24号	11,543件	990,117万円	3,288件	83,955万円	1,892件	109,544万円	16,723件	118億36,16万円
2012年 台風17号	9,966件	853,004万円	5,684件	137,466万円	318件	25,594万円	15,968件	101億6,064万円
2003年 台風14号	1,295件	319,569万円	358件	7,825万円	100件	30,476万円	1,753件	35億7,870万円
2002年 台風16号	2,972件	265,633万円	841件	13,531万円	471件	30,994万円	4,284件	31億 158万円
2014年 台風8号	5,164件	260,953万円	1,166件	26,391万円	188件	7,268万円	6,518件	29億4,612万円
2011年 台風 2号	4,220件	234,640万円	2,516件	43,950万円	222件	10,412万円	6,958件	28億9,002万円
1999年 台風18号	2,008件	189,503万円	617件	14,300万円	446件	30,395万円	3,071件	23億4,198万円
2011年 台風 9号	3,492件	203,258万円	1,190件	18,500万円	133件	6,675万円	4,815件	22億8,433万円

出典：損保協会HPより筆者編集

# 1. 研究の動機

## (2) 台風被害予測モデルの概要

- ・大規模な台風が襲来すると、全ての損害額が確定するまでに長期間を要することから、台風襲来前後で損害額をできるだけ正確に予測できることが望ましい。  
⇒例えば9月末に台風が襲来した場合、中間決算では9月末時点での保険金を見積もる必要があるが、事故受付がほとんどない状況の中で予測しなければならない（IBNR 備金の見積もり）
- ・未発生 of 台風リスクを評価するにあたっては、様々な台風の経路や規模に応じた損害額をシミュレートできるモデルの構築が重要である。



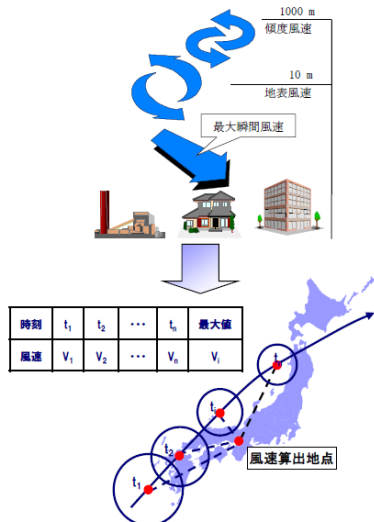
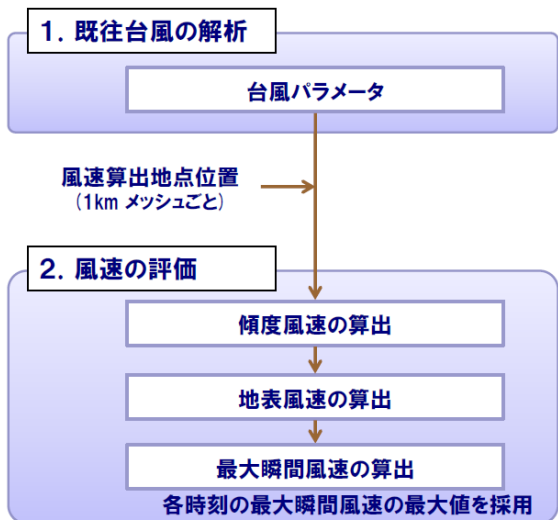
# 1. 研究の動機

## (2) 台風被害予測モデルの概要

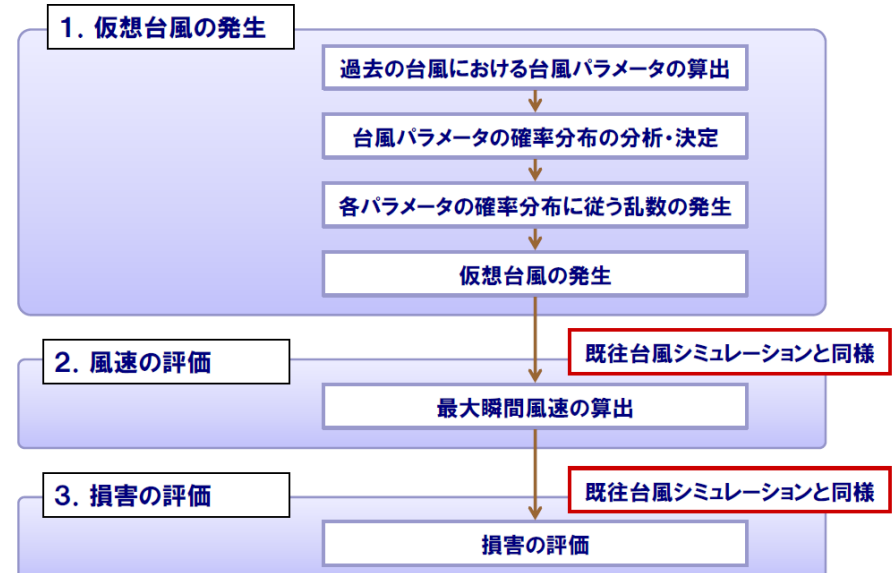
- ・一般に自然災害は過去データが少なく、統計的手法では未発生災害の予測が難しいと言われている（特に未発生の自然災害の予測）。
- ・台風リスクを評価するモデルとしては、中心気圧や最大瞬間風速、建物脆弱性などの工学的知見から台風損害を予測する「工学的事故発生モデル」が活用されている。

### 既往台風シミュレーション 2. 風速の評価

- 解析した台風パラメータを用いて、各風速算出地点位置における最大瞬間風速の評価を行う。



### 台風モンテカルロシミュレーションの概要



出典：損害保険料率算出機構HP

# 1. 研究の動機

## (3) 今回の研究内容について

- 今回の研究では、過去データを機械学習やディープラーニングによって学習させる「統計的アプローチ」からの検討を行うこととした。
- 機械学習やディープラーニングを活用することで、工学的手法や従来の統計的手法とは異なる**特徴量**を学習して予測することが期待できる。
- なお、台風に関しては進路予測や建物被害棟数予測などでディープラーニングを活用した研究が行われているが、今回は自社データから保険金を予測する点で他の研究と異なっている。
- また、当社データが沖縄県に限られていることから、本研究では沖縄県に襲来した台風情報を用いて学習・予測を行った。





# AGENDA

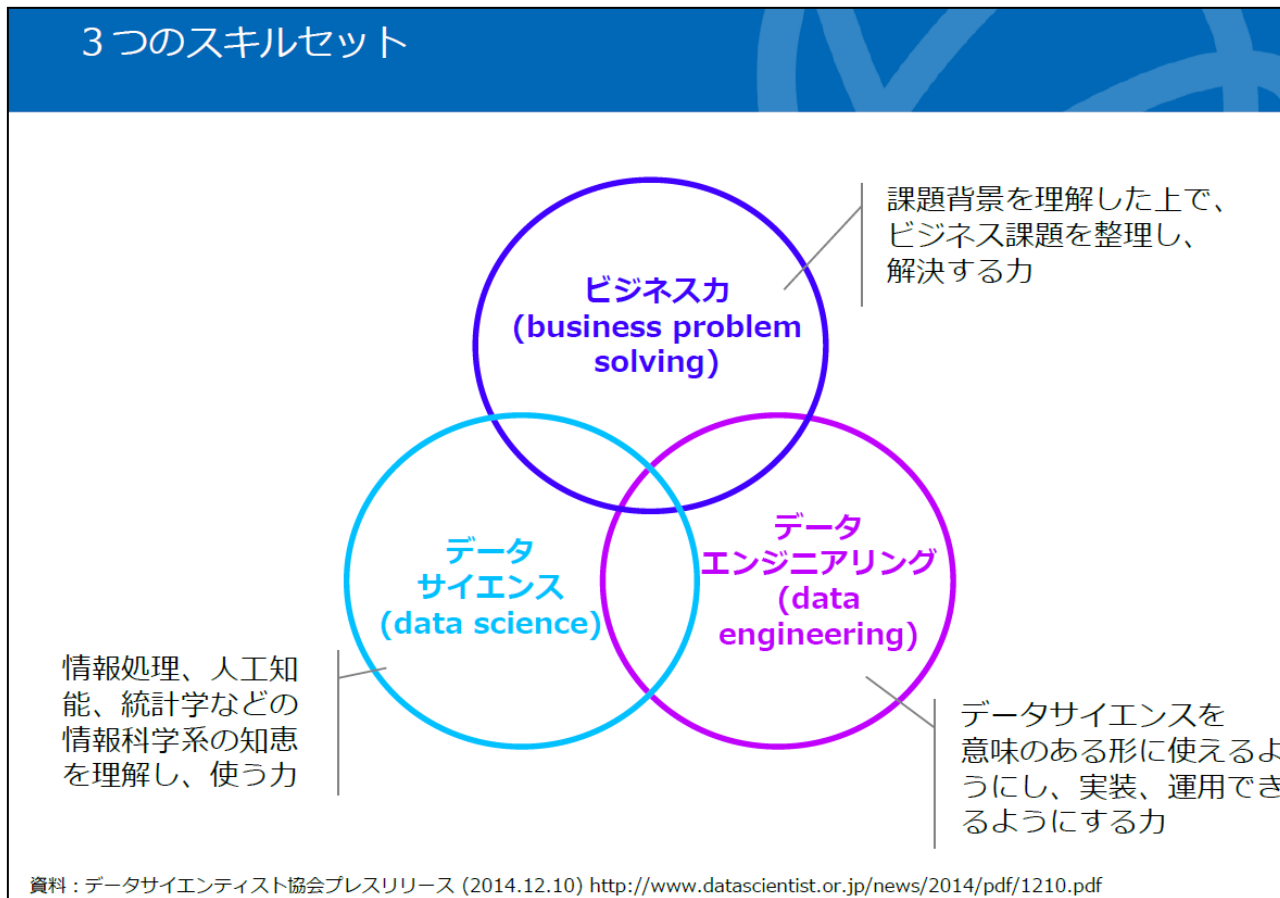
## 2. データサイエンスについて

- (1) データサイエンスの位置付け
- (2) AI、機械学習、ディープラーニングの違い
- (3) 今回の研究について

## 2. データサイエンスについて

### (1) データサイエンスの位置付け

- ・一般社団法人データサイエンティスト協会においては3つのスキルセットを提唱。
- ・このうち、データサイエンスはAIや機械学習を理解し使う力（スキル）とされている。



## 2. データサイエンスについて

### (2) AI、機械学習、ディープラーニングの違い

- AIという言葉は非常に幅広い概念であり、「人のようにふるまうもの全般」として使われる。
- AIの中でも、特に「機械が学習して分類や予測を行うこと」を**機械学習**という。
- また、機械学習のうち、特徴量を自動的に機械が見つかるものを**ディープラーニング**という。

**AI** 人のようにふるまうもの全般（明確な定義はなし）

#### 機械学習

- ✓ 分類や認識、予測を行う
- ✓ 特徴量については人間が指定

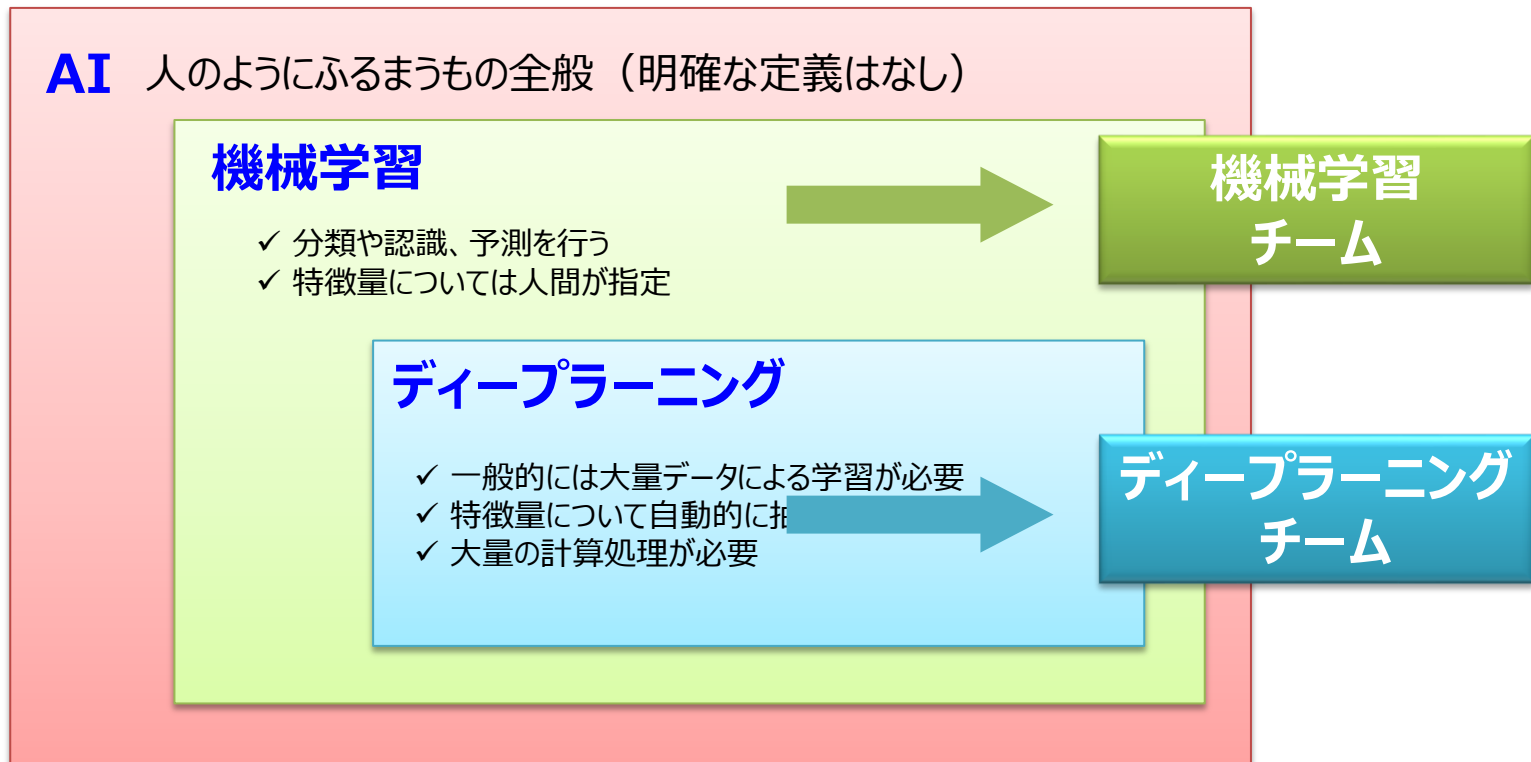
#### ディープラーニング

- ✓ 一般的には大量データによる学習が必要
- ✓ 特徴量について自動的に抽出
- ✓ 大量の計算処理が必要

## 2. データサイエンスについて

### (3) 今回の研究について

- ・今回は「機械学習チーム」と「ディープラーニングチーム」に分かれて研究を行った。
- ・以降、それぞれのチームの研究内容を報告する。



## 2. データサイエンスについて

### (3) 今回の研究について

- ・今回は「機械学習チーム」と「ディープラーニングチーム」に分かれて研究を行った。
- ・以降、それぞれのチームの研究内容を報告する。

#### 機械学習 チーム

- ・台風が襲来した場合の、**当社契約の支払総額を予測（マクロ的視点）**する。
- ・台風毎に支払総額を予測するため、**データ数は数十件程度**。

#### ディープラーニング チーム

- ・台風が襲来した場合の、**当社契約の個別支払額を予測（ミクロ的視点）**する。
- ・契約毎に個別支払額を予測するため、**データ数は数十万件程度**。

# AGENDA

## 3. 機械学習の概要と台風分析について

- (1) 主な機械学習の手法
- (2) 機械学習を用いた台風損害予測の概要
- (3) 予測結果
- (4) 今後の課題

# 3. 機械学習の概要と台風分析について

## (1) 主な機械学習の手法

・主な機械学習の手法は以下のとおりである。

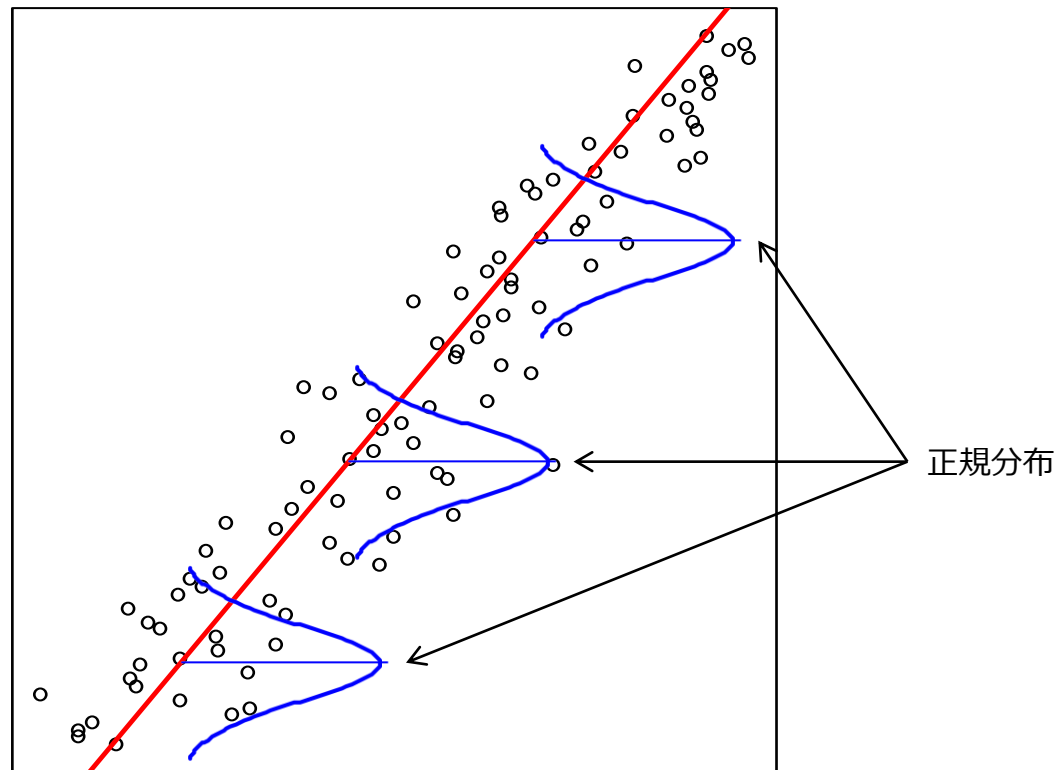
手法	概要	備考
① 線形モデル	<ul style="list-style-type: none"><li>✓ 目的変数の期待値は説明変数の線形結合で表わされる。</li><li>✓ 目的変数は正規分布に従い、その分散は一定。</li></ul>	<ul style="list-style-type: none"><li>✓ 分かりやすいが、比較的単純すぎるモデルである。</li></ul>
② 一般化線形モデル (GLM)	<ul style="list-style-type: none"><li>✓ 線形モデルを拡張したもの。</li><li>✓ 目的変数の期待値はリンク関数を用いることで説明変数の線形結合で表わされる。</li><li>✓ 目的変数は指数型分布族に従い、その分散は平均の関数で表わされる。</li></ul>	<ul style="list-style-type: none"><li>✓ 線形モデルは、リンク関数を<math>g(x)=x</math>、指数型分布族を正規分布とした特別な場合である。</li><li>✓ 代表的なものとして、ロジスティック回帰がある。</li></ul>
③ 決定木	<ul style="list-style-type: none"><li>✓ 予測に役立つ情報が最も効率的に得られるように、特徴量とその境界値を繰り返し定めることによって木の構造をした選択枝群を作成し、木の末端のそれぞれに予測値を当てはめる手法。</li></ul>	<ul style="list-style-type: none"><li>✓ 視覚的に分かりやすいが、比較的単純すぎるモデルである。</li></ul>
④ ランダムフォレスト	<ul style="list-style-type: none"><li>✓ 決定木を組み合わせたモデル。</li><li>✓ 訓練データの中からランダムに復元抽出で選んだ観測対象の集合と、全特徴量の中からランダムに選んだ特徴量の集合で（小さな）決定木を作る。そうした決定木を多数、並列に作り、集めた決定木の予測値の平均をモデル全体の予測値とする。</li></ul>	<ul style="list-style-type: none"><li>✓ 比較的複雑なモデルである。</li><li>✓ 設定するハイパーパラメータの数は比較的少ない。</li></ul>
⑤ ブースティング木	<ul style="list-style-type: none"><li>✓ 予測の方法はランダムフォレストほぼ同様。</li><li>✓ ただし、ブースティング木の場合は各決定木を1つひとつ順番に作る。その際、既に作った決定木において予測が外れたデータを優先的に正しく予測できるように次の決定木を構築する、ということを繰り返す。</li></ul>	<ul style="list-style-type: none"><li>✓ 比較的複雑なモデルである。</li><li>✓ 設定するハイパーパラメータの数は比較的多い。</li><li>✓ 代表的なものとして、勾配ブースティング、XGBoost、LightGBMがある。</li></ul>

# 3. 機械学習の概要と台風分析について

## (1) 主な機械学習の手法

### ① 線形モデル

- ・データの関係を直線で表したもの。
- ・目的変数は正規分布に従い、その分散は一定である必要がある。



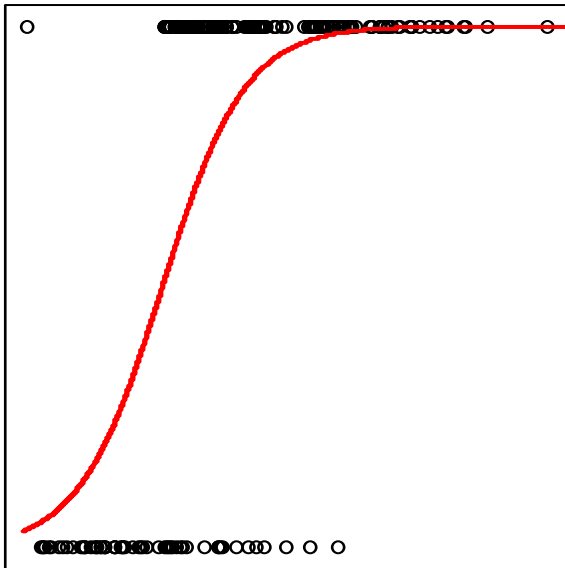


# 3. 機械学習の概要と台風分析について

## (1) 主な機械学習の手法

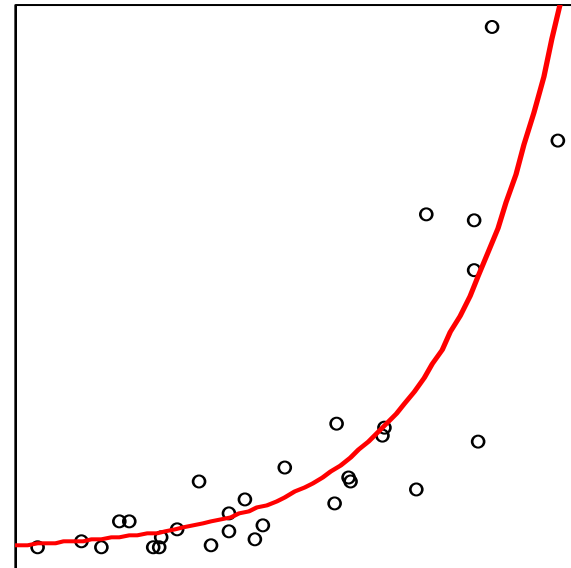
### ② 一般化線形モデル

- ・線形モデルを拡張したものである。
- ・リンク関数を用いて、データの関係を直線で表したものの。
- ・目的変数の従う分布は指数型分布族（正規分布・二項分布・ポアソン分布等）である必要がある。



ロジスティック回帰

(指数型分布族：二項分布、リンク関数：ロジット関数)



ポアソン回帰

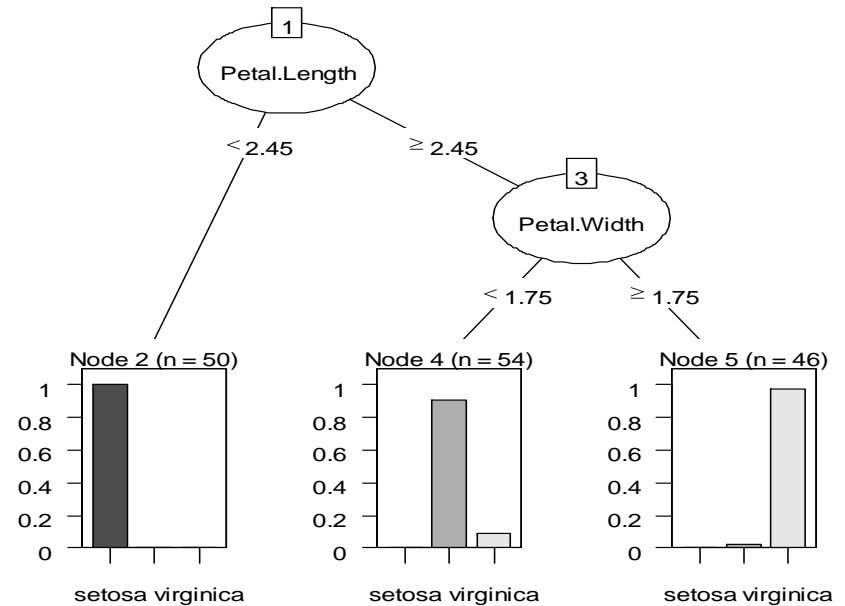
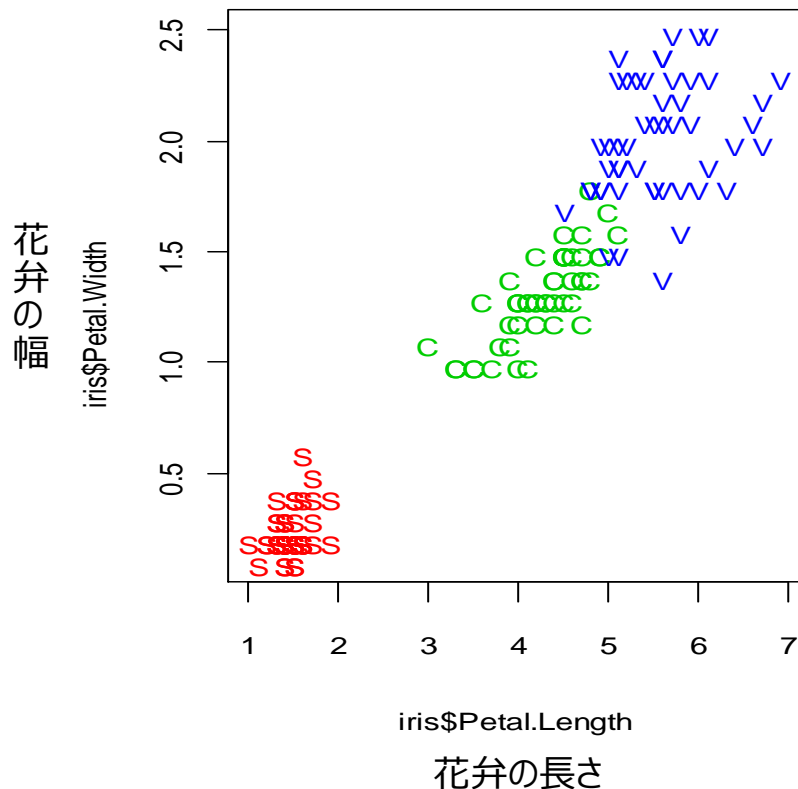
(指数型分布族：ポアソン分布、リンク関数：対数関数)

# 3. 機械学習の概要と台風分析について

## (1) 主な機械学習の手法

### ③ 決定木

- ・特徴量とその境界値を繰り返し定めることによって木の構造をした選択肢群を作成する。
- ・主にジニ不純度という指標を用いて決定木を作成していく。

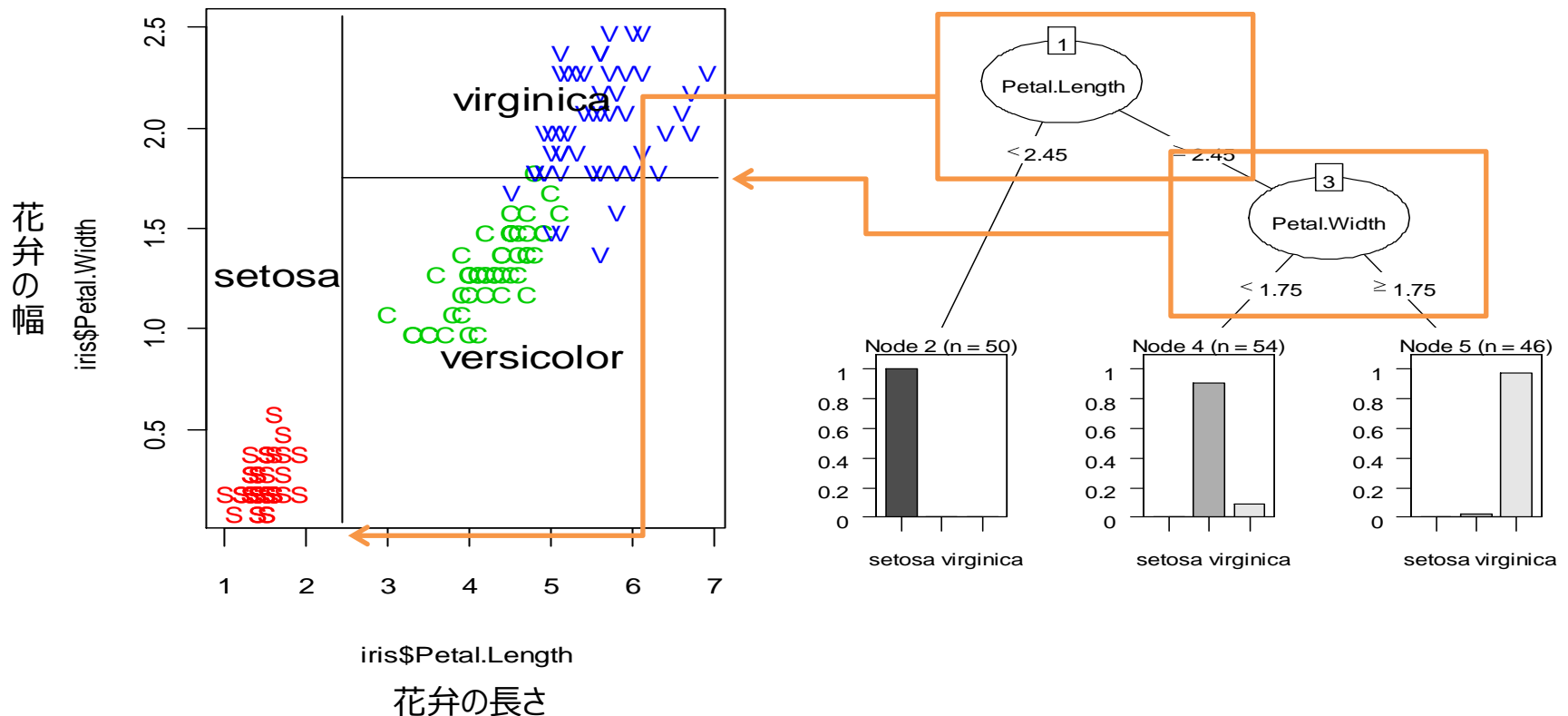


# 3. 機械学習の概要と台風分析について

## (1) 主な機械学習の手法

### ③ 決定木

- ・特徴量とその境界値を繰り返し定めることによって木の構造をした選択枝群を作成する。
- ・主にジニ不純度という指標を用いて決定木を作成していく。

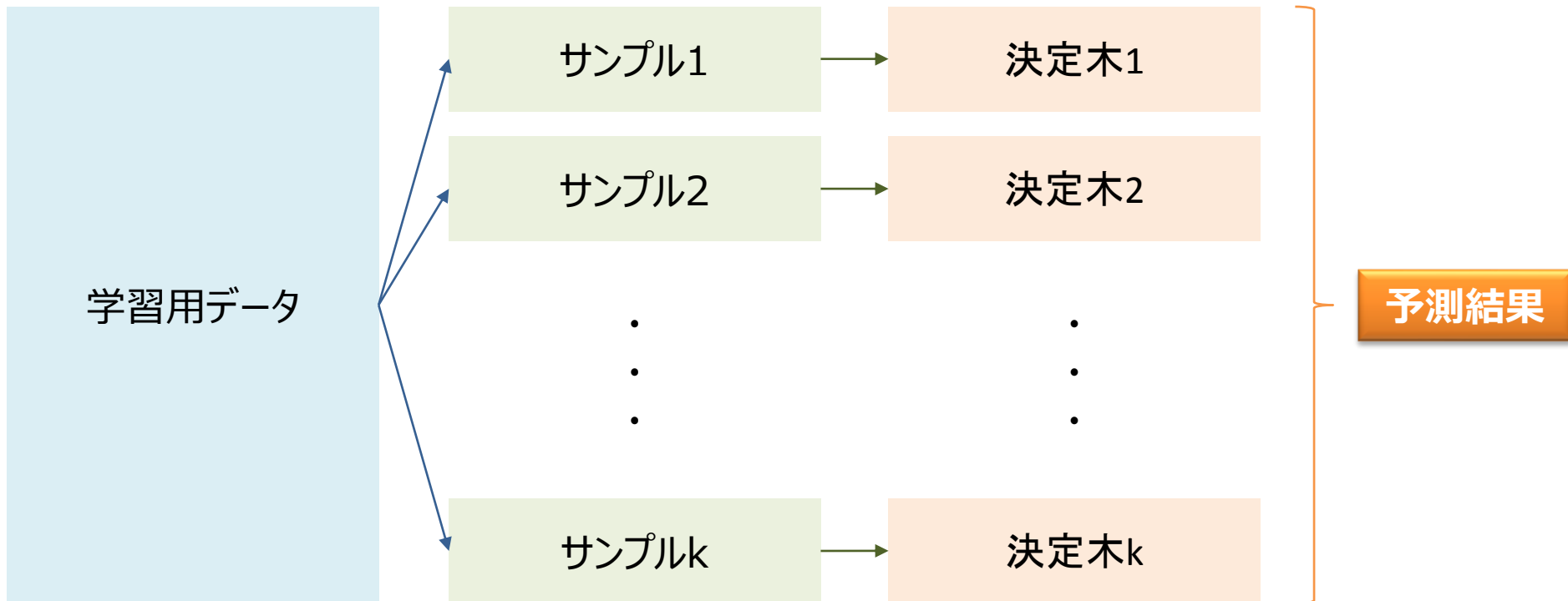


# 3. 機械学習の概要と台風分析について

## (1) 主な機械学習の手法

### ④ ランダムフォレスト

- ・決定木を組み合わせたモデル。
- ・（小さな）決定木を多数、並列に作り、集めた決定木の予測値の平均をモデル全体の予測値とする。

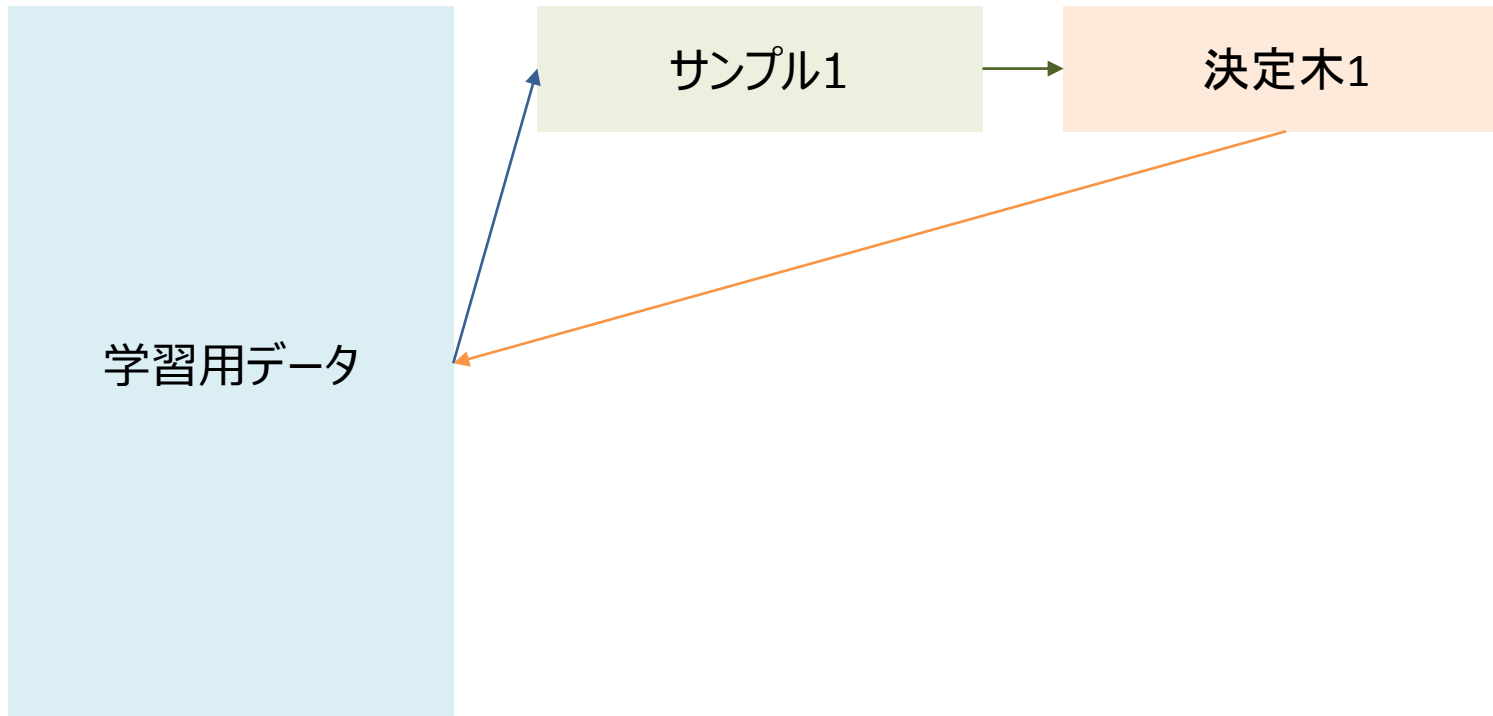


# 3. 機械学習の概要と台風分析について

## (1) 主な機械学習の手法

### ⑤ ブースティング木

- ・ランダムフォレストと同じく、決定木を組み合わせたモデル。
- ・ランダムフォレストは各決定木を並列に作成していくが、ブースティング木では1つひとつ作成していき、前に作成した決定木の影響を反映していく。

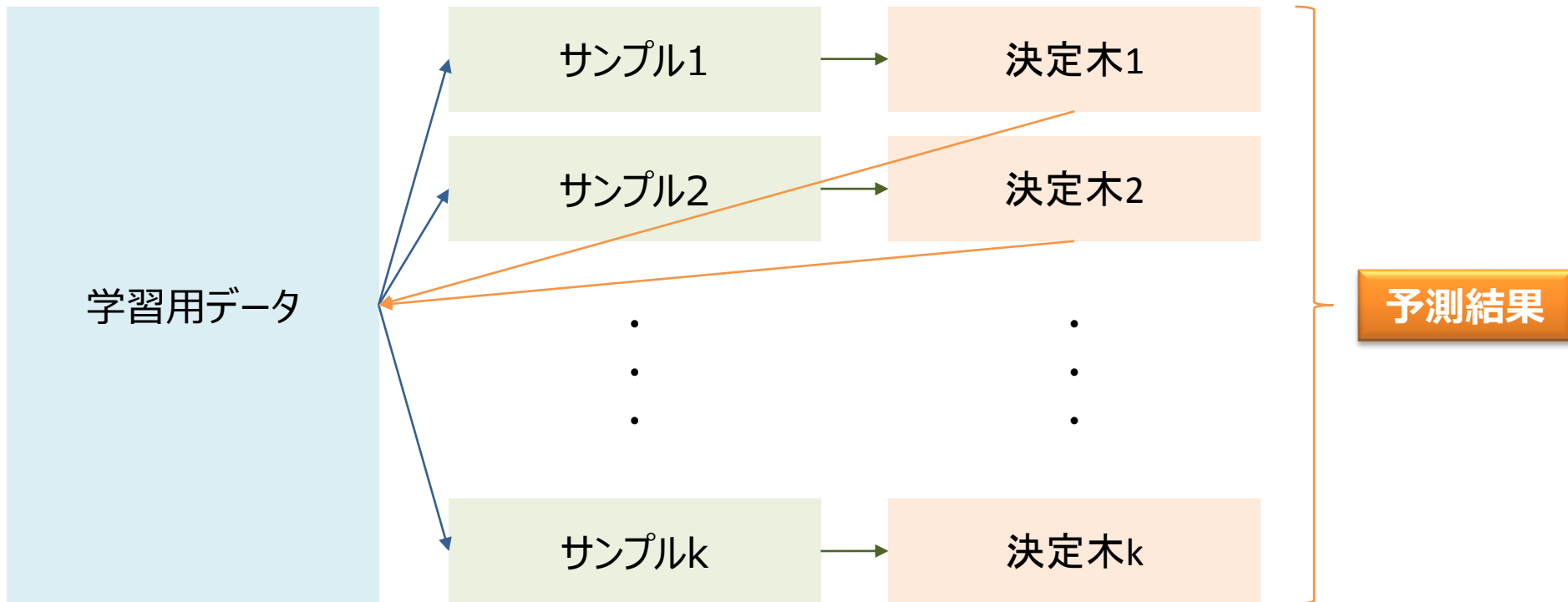


# 3. 機械学習の概要と台風分析について

## (1) 主な機械学習の手法

### ⑤ ブースティング木

- ・ランダムフォレストと同じく、決定木を組み合わせたモデル。
- ・ランダムフォレストは各決定木を並列に作成していくが、ブースティング木では1つひとつ作成していき、前に作成した決定木の影響を反映していく。



# 3. 機械学習の概要と台風分析について

## (2) 機械学習を用いた台風損害予測の概要

### ① 予測概要

- ・台風が襲来した場合に、いくら保険金が支払われるか予測を行う。
- ・機械学習を用いて、台風情報から支払総額を求めるモデルを構築する。



# 3. 機械学習の概要と台風分析について

## (2) 機械学習を用いた台風損害予測の概要

### ② データセット

・2009年～2019年（2019年は台風17号まで）に襲来した台風を対象に、当社保険契約と支払データ、外部データとして沖縄気象台HP (<https://www.jma-net.go.jp/okinawa/data/kencho/typntoppu.html>) で公表されている各台風の情報（人的被害、最低海面気圧等）を用いた。

項目（全62項目）	備考
1 台風における支払総額	内部データ：当社の支払総額（全種目）
1 週間の受付件数	内部データ：当社の事故受付件数
保有保険金額	内部データ：当社の県内全体の補償金額
人的被害	外部データ
住家・非住家被害	外部データ
停電戸数	外部データ
最低海面気圧	外部データ
最大風速	※ 8 観測地点（那覇・名護・久米島・南大東・宮古・石垣・西表・与那国）の情報。なお、観測地点の情報が無い場合は、気象庁のデータを使用した。
最大瞬間風速	
最大日降水量	
暴風域時間	
強風域時間	
最接近距離	

対象地域：沖縄県地方

（ 速 報 ）

令和元年台風第9号について

目 次

資料 1 気象概況

資料 2 台風経路図

資料 3 地上天気図、沿岸波浪実況図

資料 4 気象衛星画像、気象レーダー

資料 5 暴風域・強風域入出日時

資料 6 台風期間における観測値、極値

資料 7 気象警報・注意報発表状況

資料 8 時系列図（降水量、風向・風速、現地気圧）

資料 9 観測値表（降水量、風向・風速）

資料 10 被害状況

注）この報告は、8月9日12時までの資料（一部を除く）により速報としてまとめたものです。

この資料に関する問合せ先：防災調査課調査係（TEL：098-833-2186）

令和元年8月9日

沖 縄 気 象 台

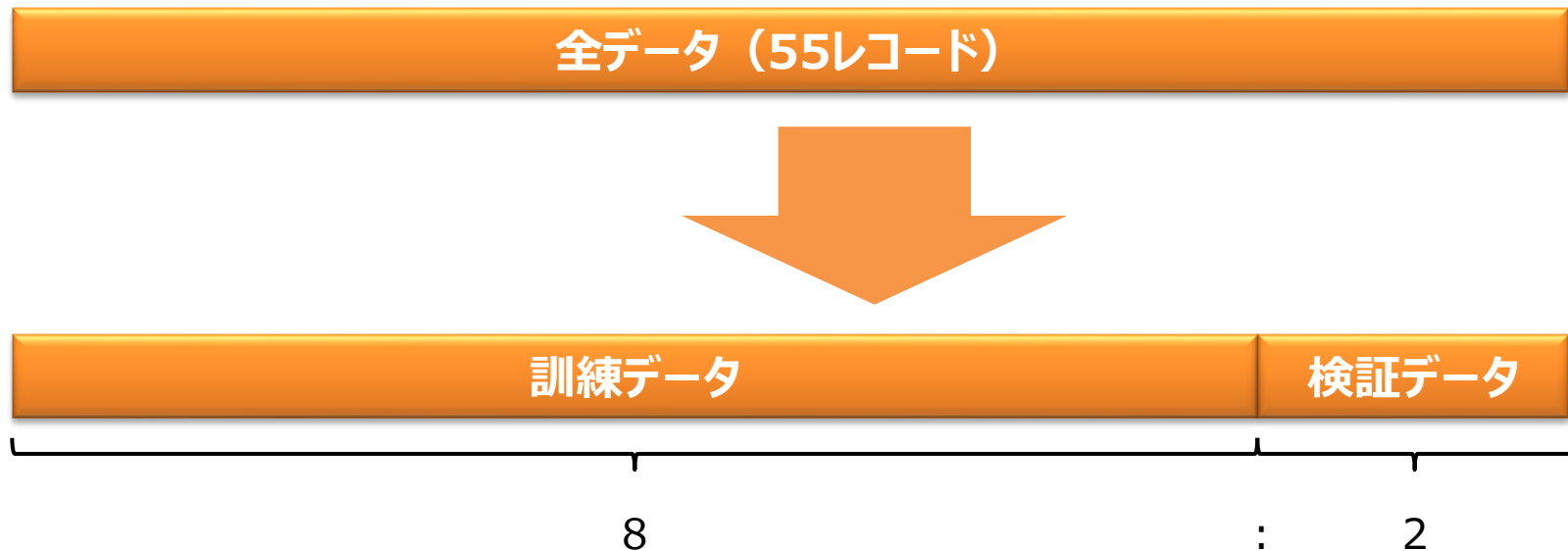


# 3. 機械学習の概要と台風分析について

## (2) 機械学習を用いた台風損害予測の概要

### ③ データセットの分割

- ・データセットを 8 : 2 の割合でランダムに「訓練データ」と「検証データ」に分割した。
- ・「訓練データ」で学習させ、検証データで予測結果の精度を検証した。



# 3. 機械学習の概要と台風分析について

## (2) 機械学習を用いた台風損害予測の概要

### ④手法

- ・一般化線形モデル、ランダムフォレスト、ブースティング木（XGBoost）を用いる。
- ・どの手法が予測精度が高いか検証を行い、予測精度が高いモデルを採用する。

# 3. 機械学習の概要と台風分析について

## (2) 機械学習を用いた台風損害予測の概要

### ⑤ 手法間の評価指標

- モデルを評価する指標としてRMSE（Root Mean Squared Error：二乗平均平方根誤差）を使用する。
- RMSEは0に近いほど予測精度が高いことを表している。

$$RMSE = \sqrt{\frac{1}{n} \sum (\text{実績値} - \text{予測値})^2}$$

# 3. 機械学習の概要と台風分析について

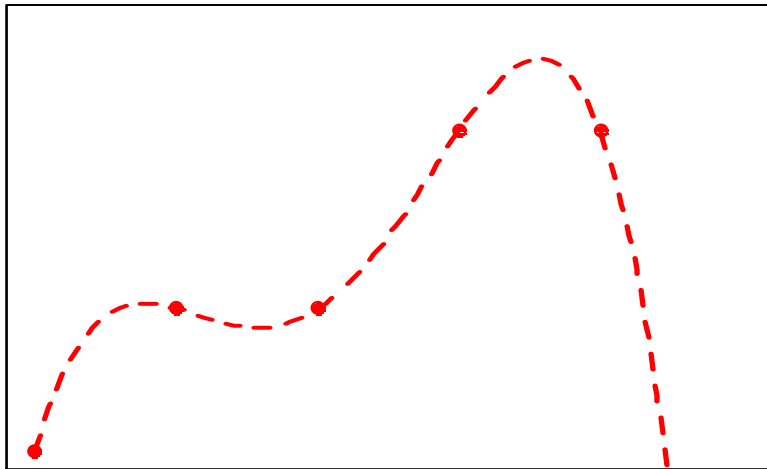
## (2) 機械学習を用いた台風損害予測の概要

### ⑥ 過学習への対応

- ・今回の分析では、レコード数（55）より説明変数の数（61）が多くなっているため、一般化線形モデルは過学習になってしまう。そのため、正則化を行う。

#### 過学習とは

訓練データに適合し過ぎてしまい、実際に予測したいデータに対してあまり適合しないこと。



#### 正則化とは

過学習を抑える方法として正則化がある。正則化は以下の正則化項（罰則項）を追加することで、制約を設ける。

$$\lambda\{\alpha\|\beta\|_1 + (1 - \alpha)\|\beta\|_2\}$$

$$\|\beta\|_1 = |\beta_1| + |\beta_2| + \dots + |\beta_p|$$

$$\|\beta\|_2 = \beta_1^2 + \beta_2^2 + \dots + \beta_p^2$$

※ $\lambda$ および $\alpha$ はパラメータである。特に、 $\alpha=1$ のときLASSO回帰、 $\alpha=0$ のときRidge回帰、それ以外るときElastic Netという。

# 3. 機械学習の概要と台風分析について

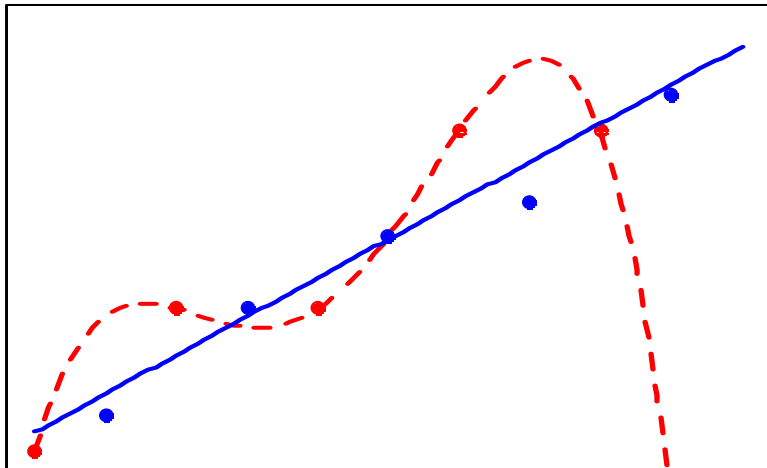
## (2) 機械学習を用いた台風損害予測の概要

### ⑥ 過学習への対応

- ・今回の分析では、レコード数（55）より説明変数の数（61）が多くなっているため、一般化線形モデルは過学習になってしまう。そのため、正則化を行う。

#### 過学習とは

訓練データに適合し過ぎてしまい、実際に予測したいデータに対してあまり適合しないこと。



#### 正則化とは

過学習を抑える方法として正則化がある。正則化は以下の正則化項（罰則項）を追加することで、制約を設ける。

$$\lambda\{\alpha\|\beta\|_1 + (1 - \alpha)\|\beta\|_2\}$$

$$\|\beta\|_1 = |\beta_1| + |\beta_2| + \dots + |\beta_p|$$

$$\|\beta\|_2 = \beta_1^2 + \beta_2^2 + \dots + \beta_p^2$$

※ $\lambda$ および $\alpha$ はパラメータである。特に、 $\alpha=1$ のときLASSO回帰、 $\alpha=0$ のときRidge回帰、それ以外のときElastic Netという。

# 3. 機械学習の概要と台風分析について

## (2) 機械学習を用いた台風損害予測の概要

### ⑦その他

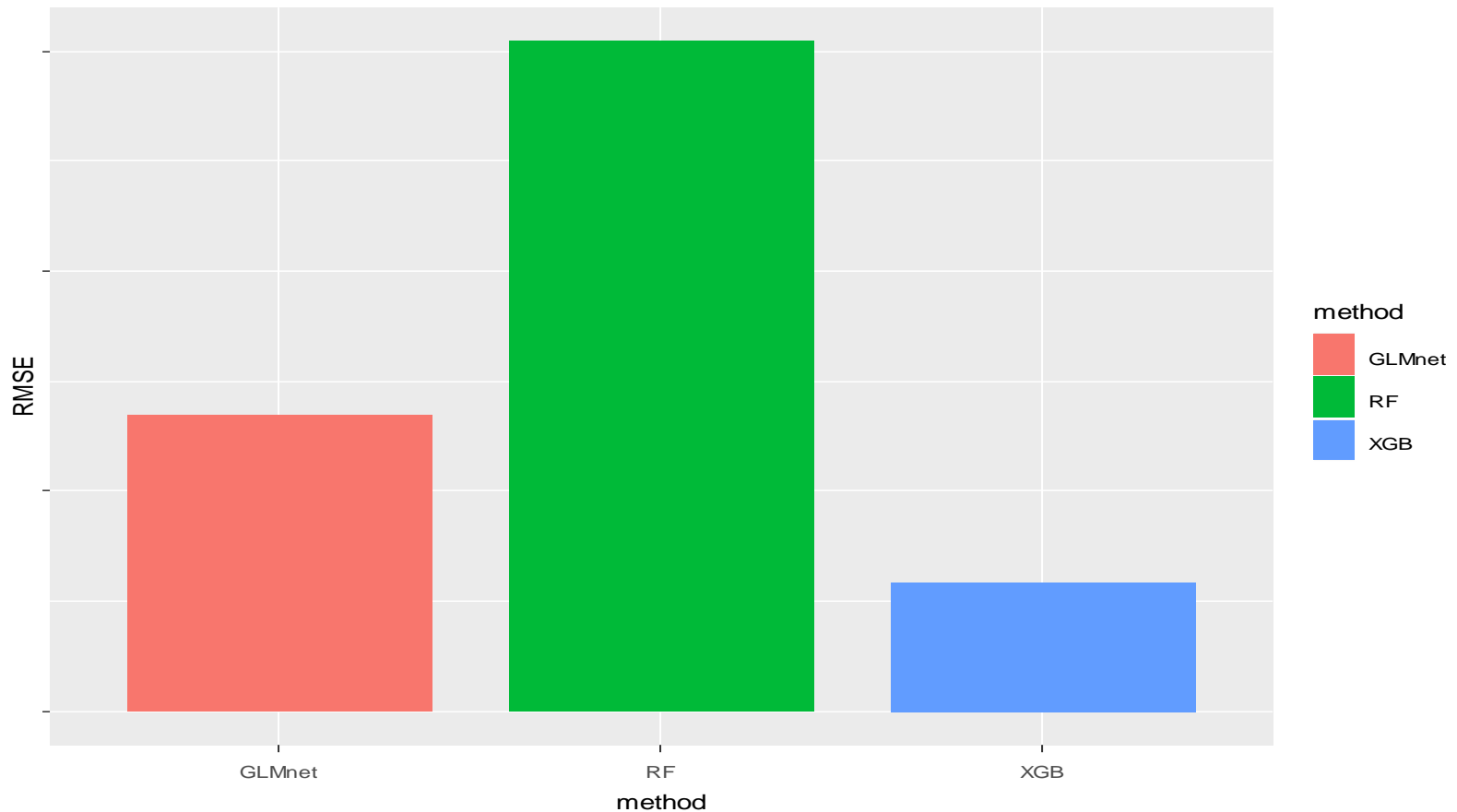
その他のモデル実装の概要は以下の通り・・・（詳細割愛）

- ・ソフトはR言語、caretパッケージを使用した。
- ・caretパッケージを使用し、ハイパーパラメータのチューニング、交差検証を行った。

# 3. 機械学習の概要と台風分析について

## (3) 予測結果

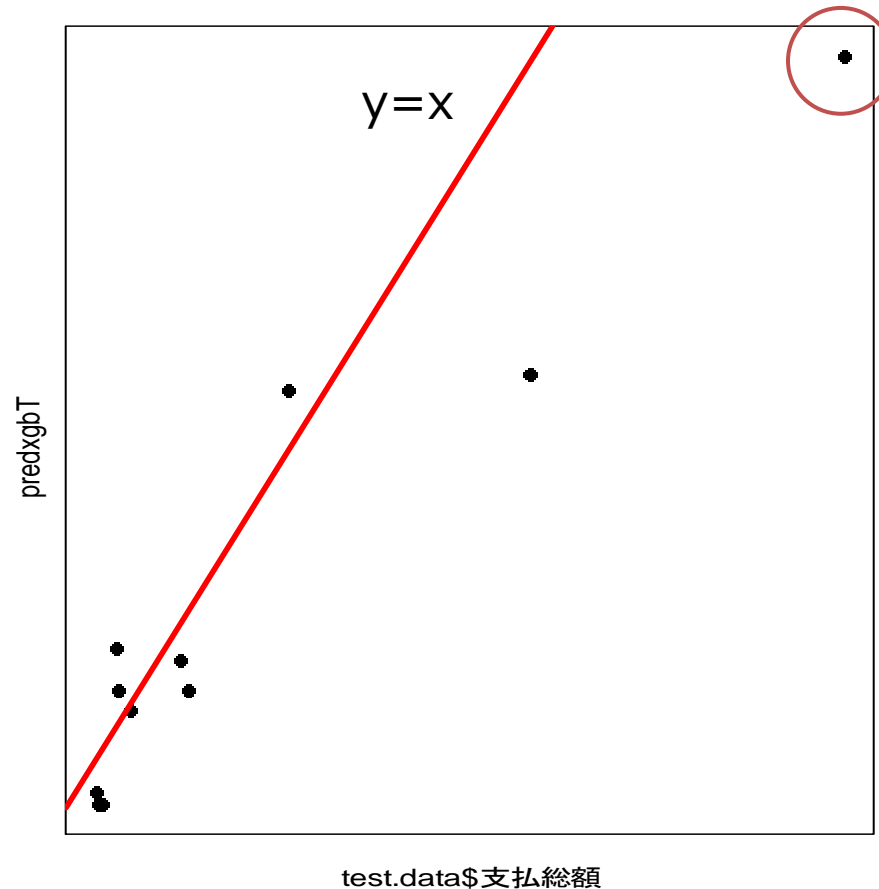
- ・各手法のRMSEは以下のとおり。
- ・ブースティング木（XGBoost）が最小のRMSEとなった。



# 3. 機械学習の概要と台風分析について

## (3) 予測結果

- ・ブースティング木 (XGBoost) の結果と検証データの値をプロットすると、以下のとおり。
- ・検証データの支払総額が大きい台風については、過小評価する結果となった。





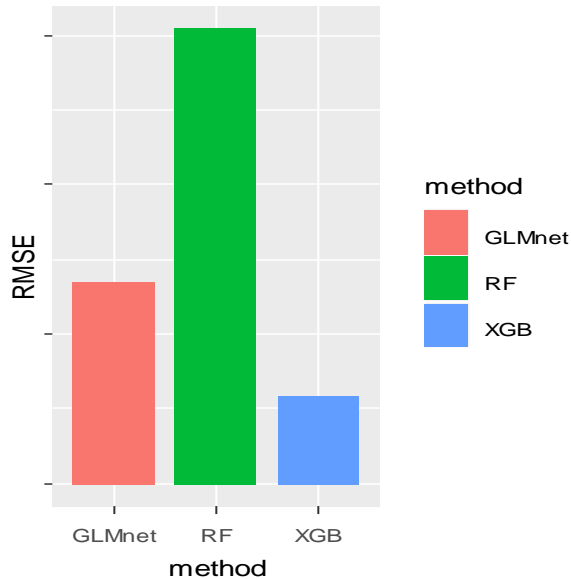
# 3. 機械学習の概要と台風分析について

## (3) 予測結果

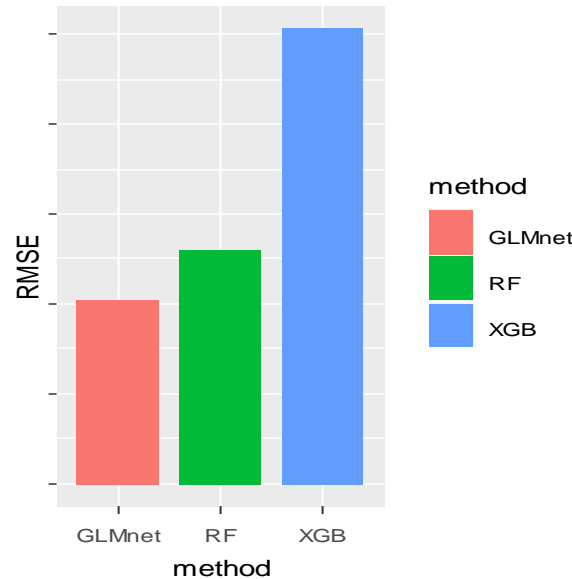
### <参考>

- ・データセットの分割方法を変更したところ、別の手法のRMSEが最小となった。
- ・支払総額が他と比較して大きいものがあり、このデータが訓練データに含まれるかどうかで、検証データのRMSEが大きく変動したと考えられる。

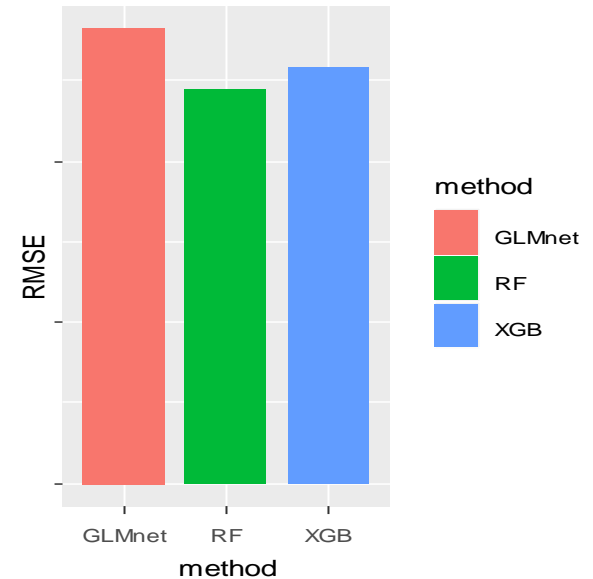
今回採用



別パターン1



別パターン2

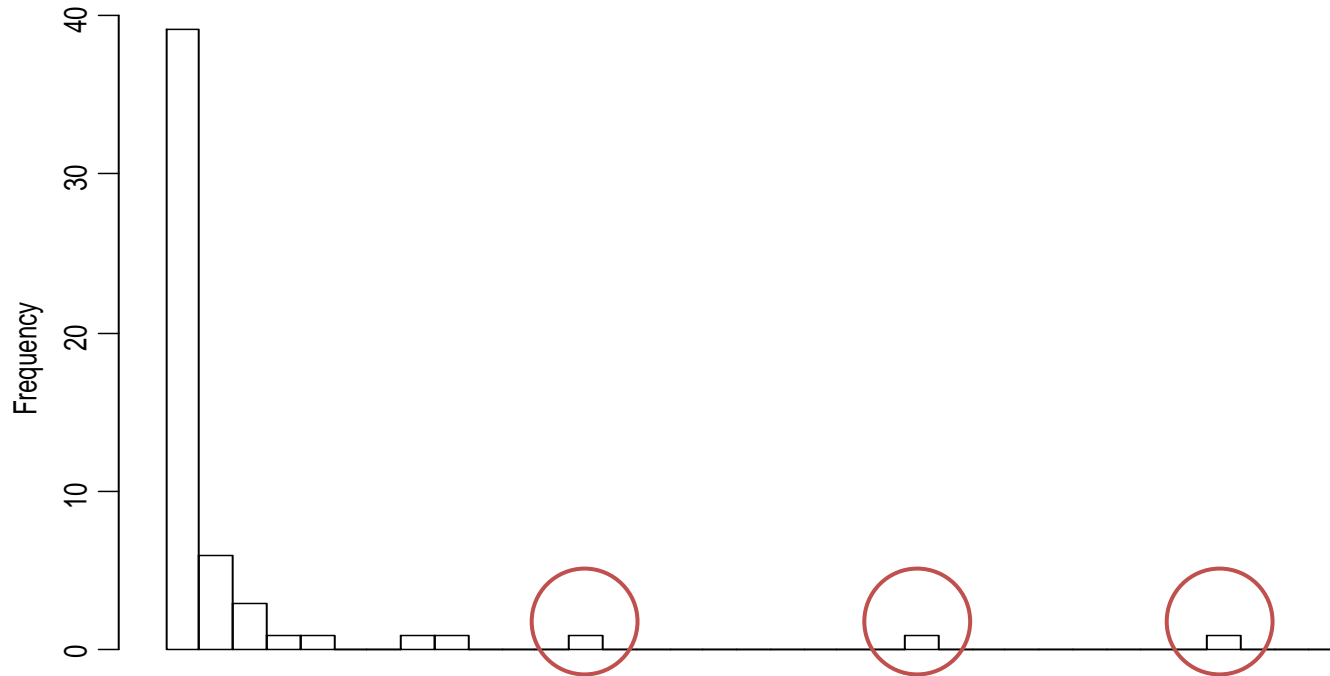


# 3. 機械学習の概要と台風分析について

## (3) 予測結果

### <参考>

- データセットの分割方法を変更したところ、別の手法のRMSEが最小となった。
- 支払総額が他と比較して大きいものがあり、このデータが訓練データに含まれるかどうかで、検証データのRMSEが大きく変動したと考えられる。



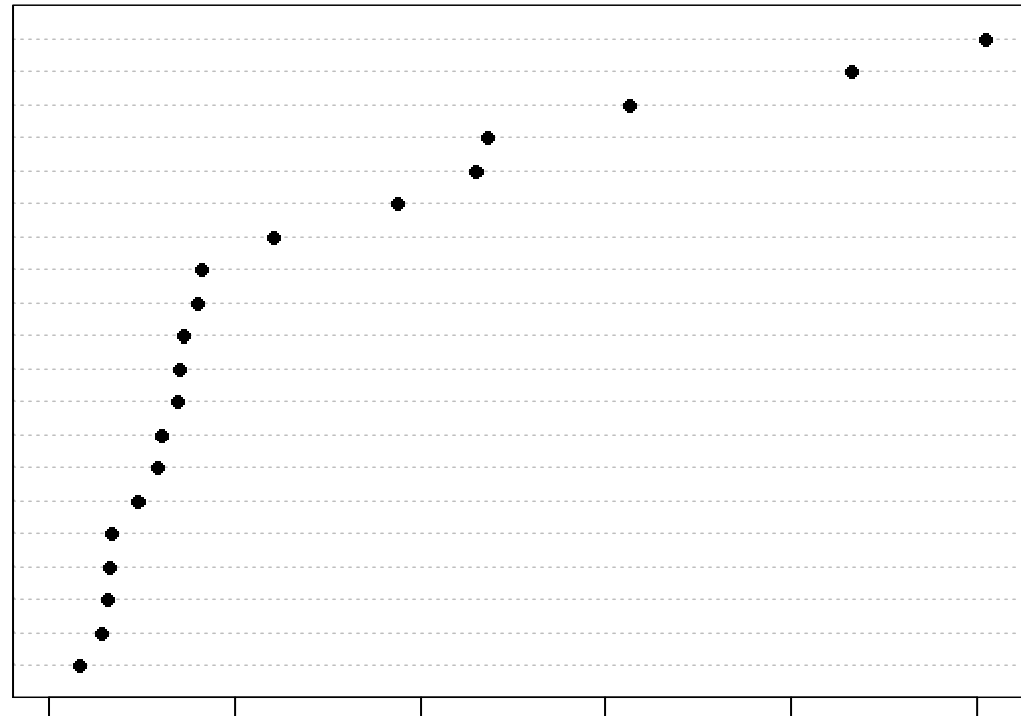
# 3. 機械学習の概要と台風分析について

## (3) 予測結果

### <参考>

・ランダムフォレストでは、各説明変数の重要度の計算が可能であり、以下のとおりとなった。

- 最低海面気圧.那覇.
- 最大風速.那覇.
- 最大瞬間風速.那覇.
- 1週間の受付件数
- 停電戸数
- 最低海面気圧.久米島.
- 人的被害
- 最大日降水量.久米島.
- 最大風速.久米島.
- 最接近距離.那覇.
- 最大日降水量.宮古.
- 最低海面気圧.名護.
- 最大瞬間風速.名護.
- 最大瞬間風速.久米島.
- 最大日降水量.石垣.
- 最大日降水量.西表.
- 強風域時間.与那国.
- 強風域時間.宮古.
- 暴風域時間.宮古.
- 強風域時間.石垣.



IncNodePurity

# 3. 機械学習の概要と台風分析について

## (4) 今後の課題

今後より精緻な予測モデルが開発できるよう、以下の観点からデータ整備やモデル開発に取り組んでいきたい。

- ・過学習への対応
- ・その他の機械学習の使用
- ・過去の台風情報の取得、今後の台風情報の蓄積
- ・他の外部データの使用

# AGENDA

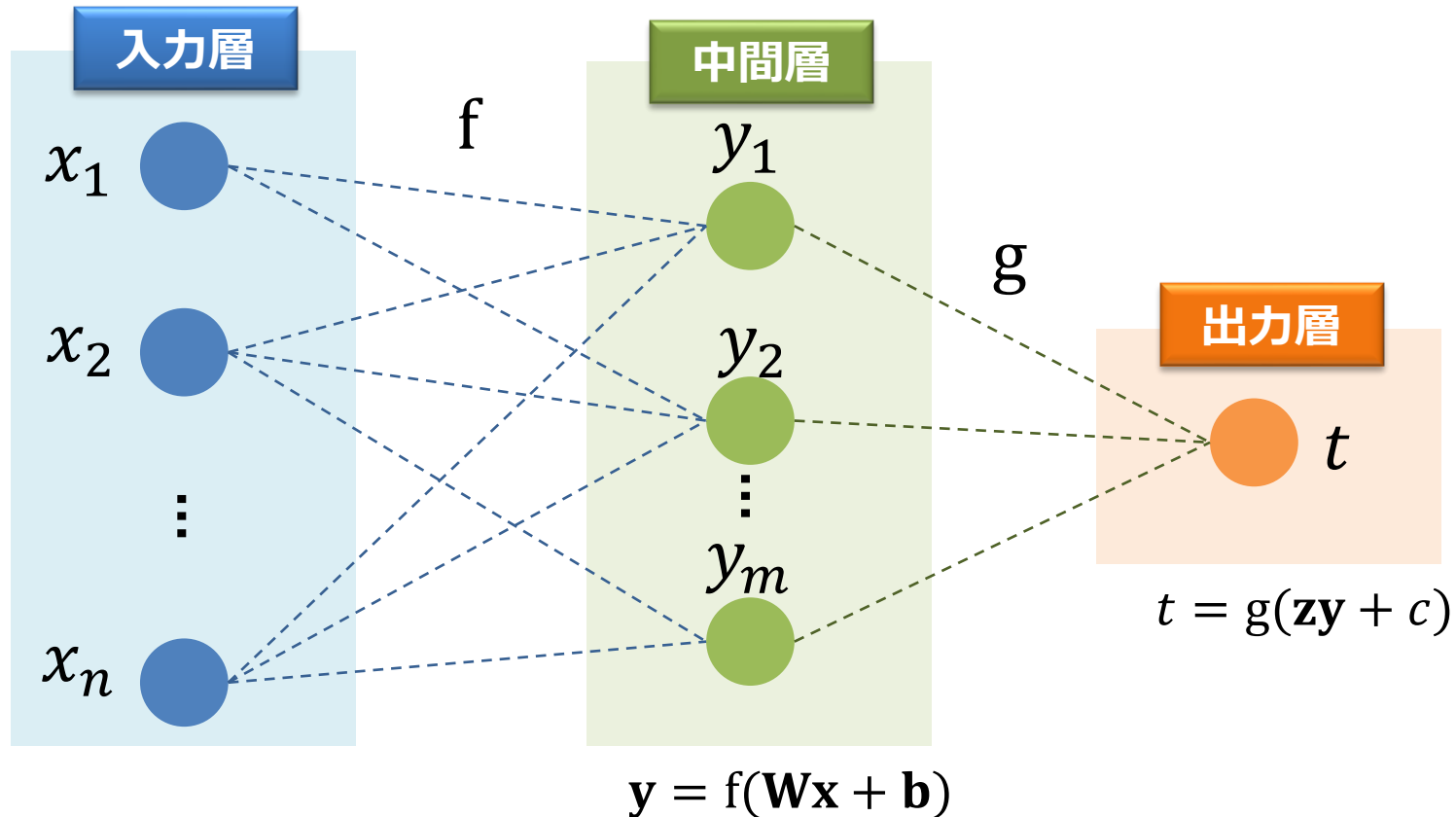
## 4. ディープラーニングの概要と台風分析について

- (1) ニューラルネットワークの概要
- (2) ディープラーニングの概要
- (3) ディープラーニングを用いた台風損害予測の概要
- (4) 予測結果
- (5) 今後の課題

# 4. ディープラーニングの概要と台風分析について

## (1) ニューラルネットワークの概要

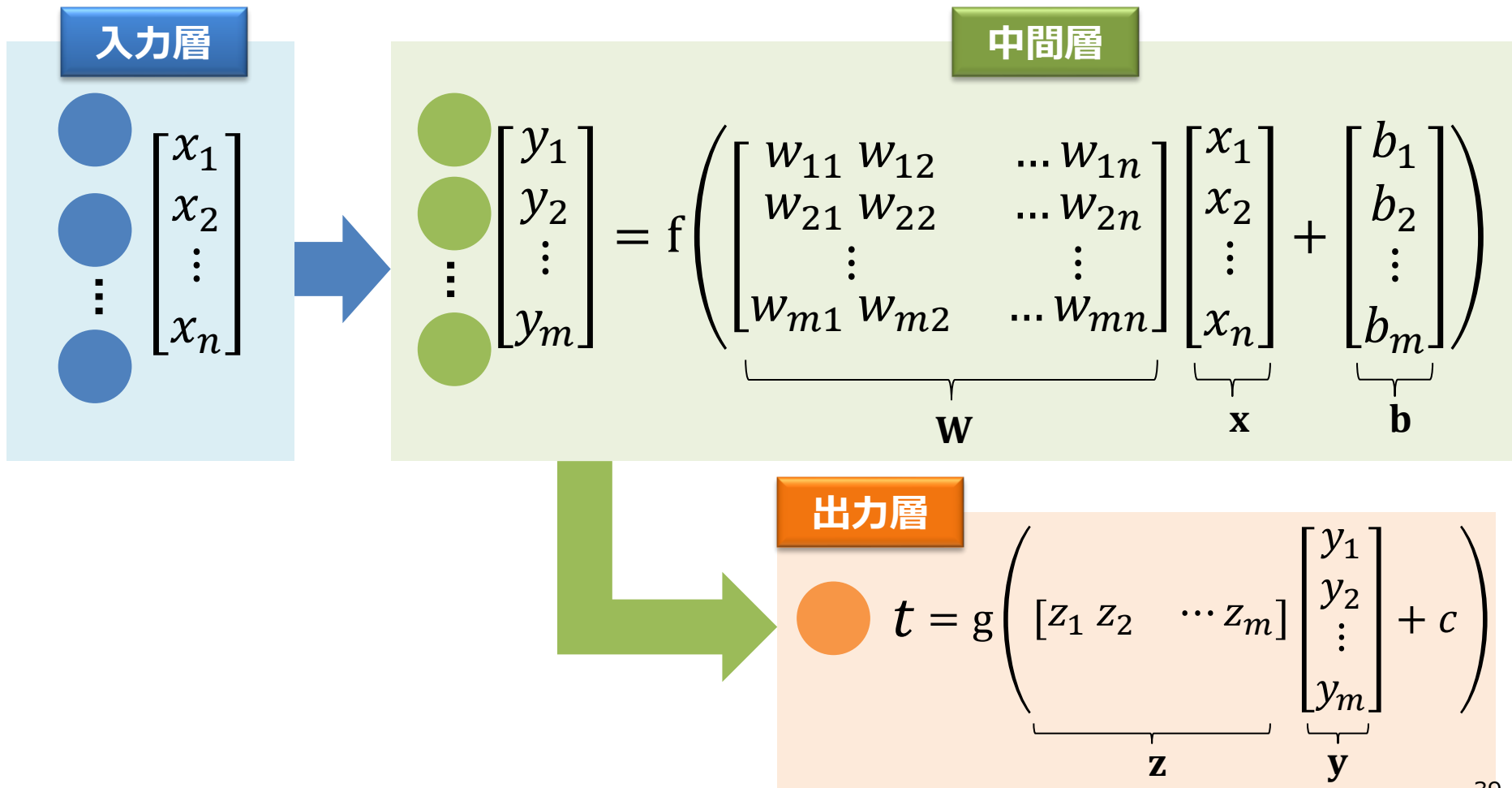
- ・入力層／中間層／出力層の3層で構成
- ・活性化関数 $f, g$ によって各層へ出力され、最終的な予測結果 $t$ が得られる。
- ・活性化関数 $f, g$ に入れる数値は各層の重み $W, z$ とバイアス $b, c$ で調整される。



# 4. ディープラーニングの概要と台風分析について

## (1) ニューラルネットワークの概要

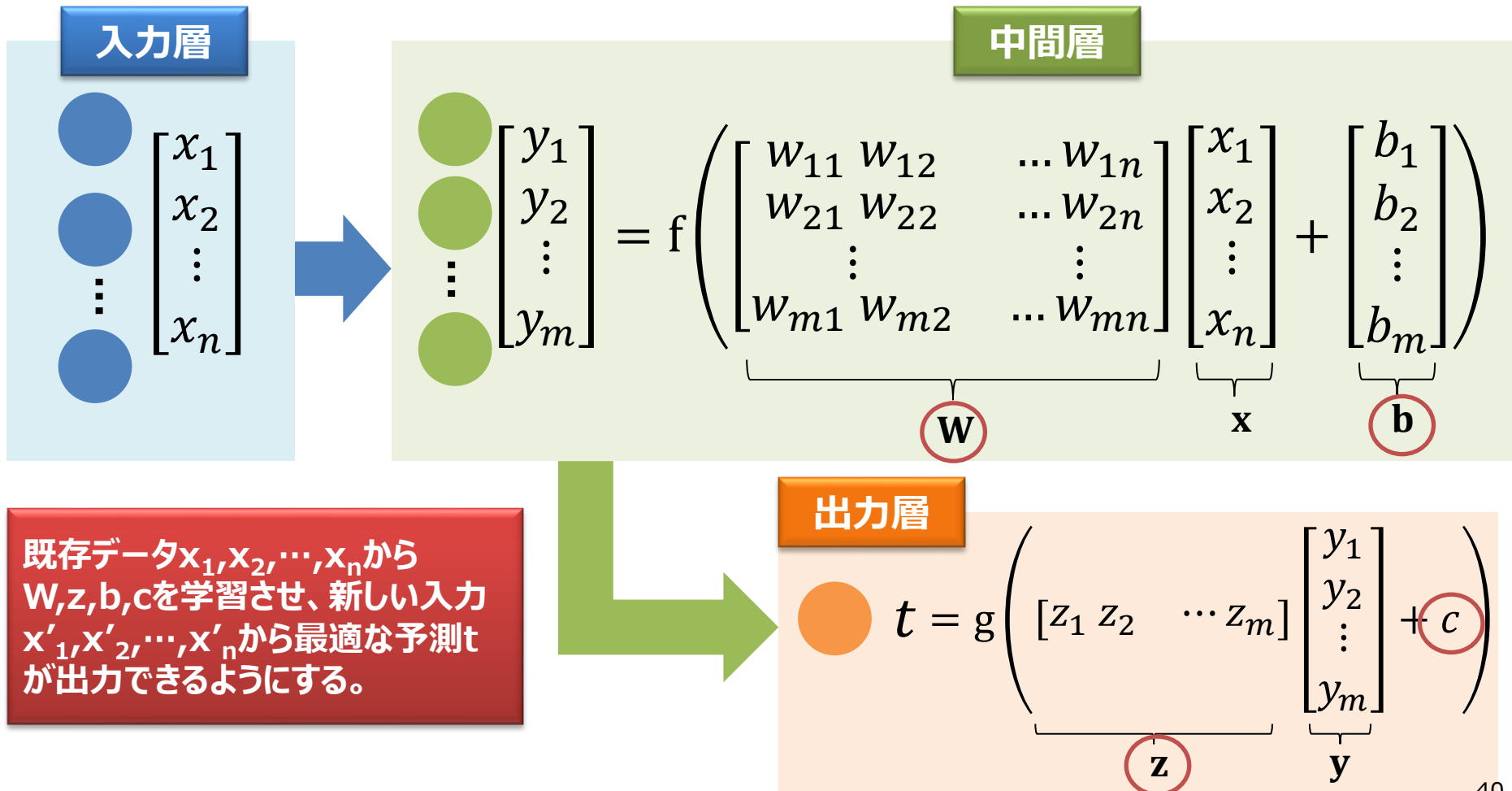
・行列を用いた式で表すと以下の通り。



# 4. ディープラーニングの概要と台風分析について

## (1) ニューラルネットワークの概要

・行列を用いた式で表すと以下の通り。

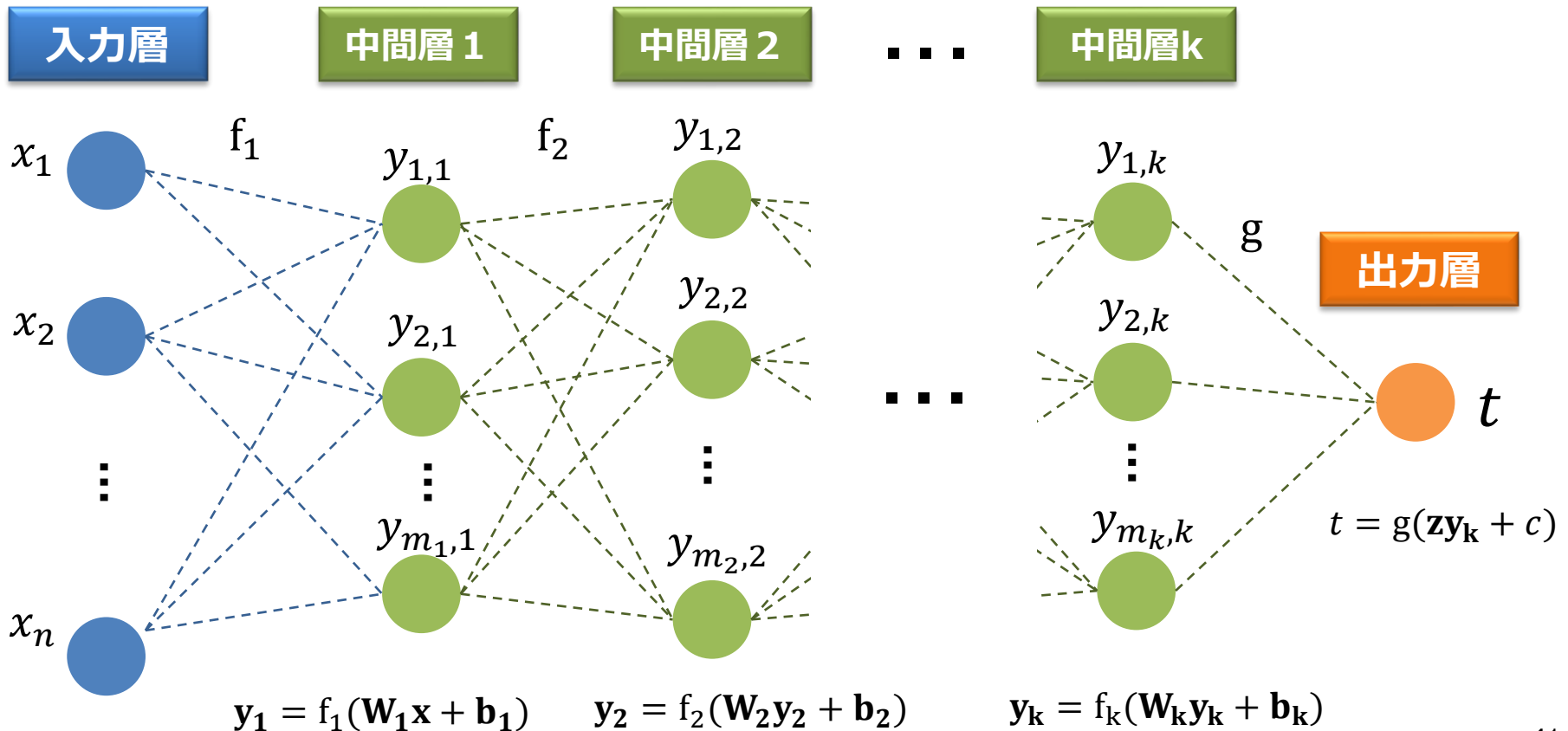




# 4. ディープラーニングの概要と台風分析について

## (2) ディープラーニングの概要 (ディープニューラルネットワーク)

- ・ディープラーニングは、中間層の数を多くしたもの (中間層を厚くしたもの)
- ・以前は $W_1 \sim W_k$ や $b_1 \sim b_k$ などのパラメータを求めるのに大量の計算が必要だったが、近年のPC処理の高速化や効率的な計算手法の開発によって活用されるようになった。

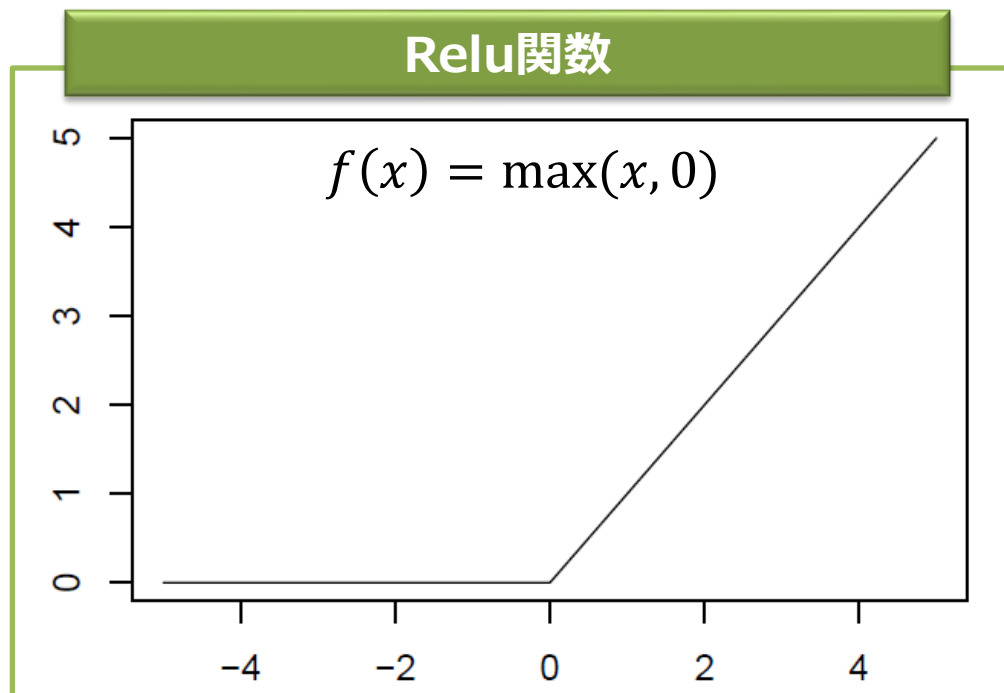


# 4. ディープラーニングの概要と台風分析について

## (2) ディープラーニングの概要 (活性化関数)

### ① 中間層の活性化関数

- ・中間層の活性化関数ではRelu関数を用いることが多い。
- ・一般にRelu関数は他の活性化関数より学習が早く進み、良い結果が得られることが多いとされている。



他の活性化関数の例

$$f(x) = \tanh(x)$$

$$f(x) = \begin{cases} 1 & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

# 4. ディープラーニングの概要と台風分析について

## (2) ディープラーニングの概要 (活性化関数)

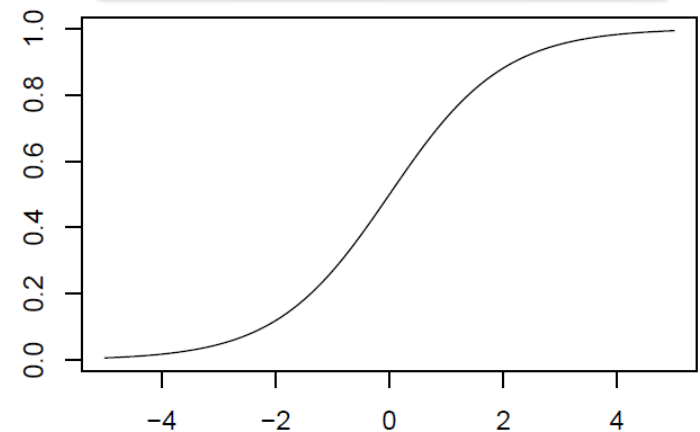
### ② 出力層の活性化関数

- 出力層は、回帰問題では恒等関数、分類問題ではシグモイド関数、ソフトマックス関数を用いることが多い。

#### 出力層の主な活性化関数

$g(x) = x$	恒等関数 : 回帰
$g(x) = \frac{1}{1 + e^{-x}}$	シグモイド関数 : 二値分類
$g(x_k) = \frac{\exp(x_k)}{\sum_i \exp(x_i)}$	ソフトマックス関数 : 多クラス分類

#### 参考 : シグモイド関数のグラフ



# 4. ディープラーニングの概要と台風分析について

## (2) ディープラーニングの概要 (W,b,z,cを最適化する方法)

### ③損失関数 (誤差関数)

- ・学習データにおける「損失関数 (誤差関数)」が出来るだけ小さくなるようにパラメータ W,b,z,cを調整する。
- ・損失関数 (誤差関数) としては、回帰問題では平均二乗誤差、分類問題ではクロスエントロピー誤差を用いることが多い。

#### 主な損失関数 (誤差関数)

平均二乗誤差  
(回帰問題)

$$E = \sum_{i=1}^n (\hat{t}_i - t_i)^2$$

$\hat{t}_i$  : 正解値

$t_i$  : 予測値

$n$  : 学習データ数

クロスエントロピー誤差  
(分類問題)

$$E = - \sum_{i=1}^n \hat{t}_i \log t_i$$

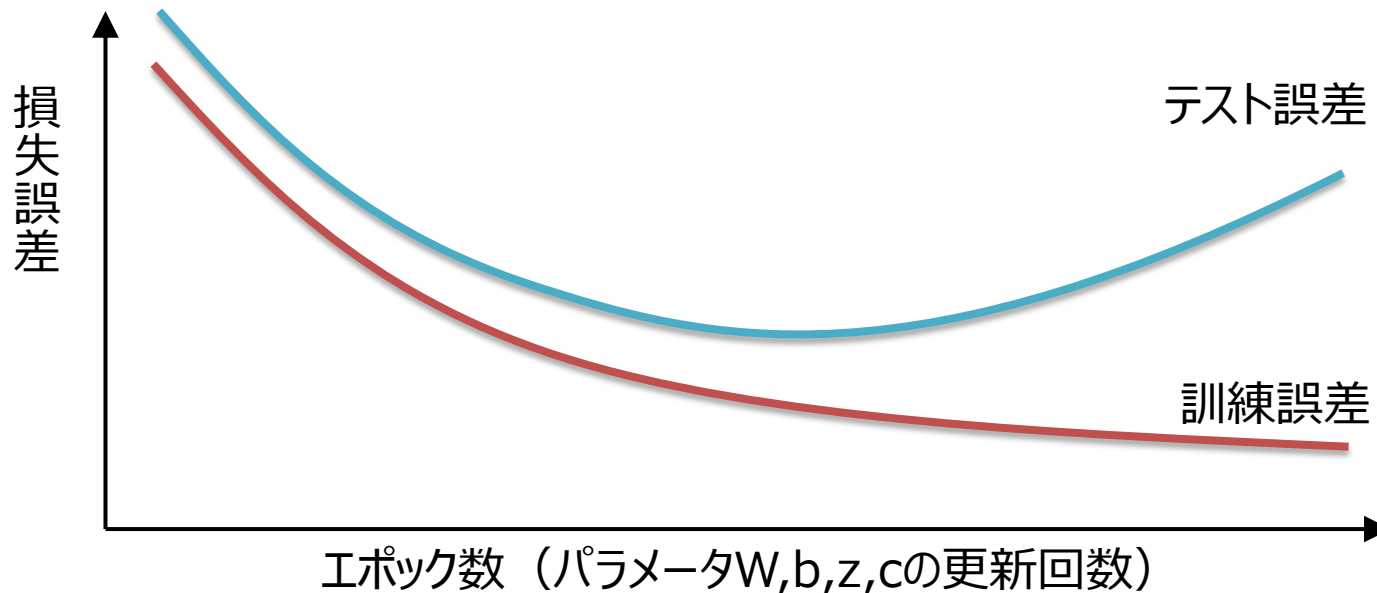
他の損失関数 (誤差関数) の例 : 平均絶対誤差、Huber Loss等

## 4. ディープラーニングの概要と台風分析について

### (2) ディープラーニングの概要 (W,b,z,cを最適化する方法)

#### ④パラメータW,b,z,cの更新

- ・損失（誤差）が最小になるようにパラメータW,b,z,cを繰り返し更新していく。
- ・テスト誤差が大きくなると過学習の傾向にあるため、適切な更新回数で終了させる必要がある。

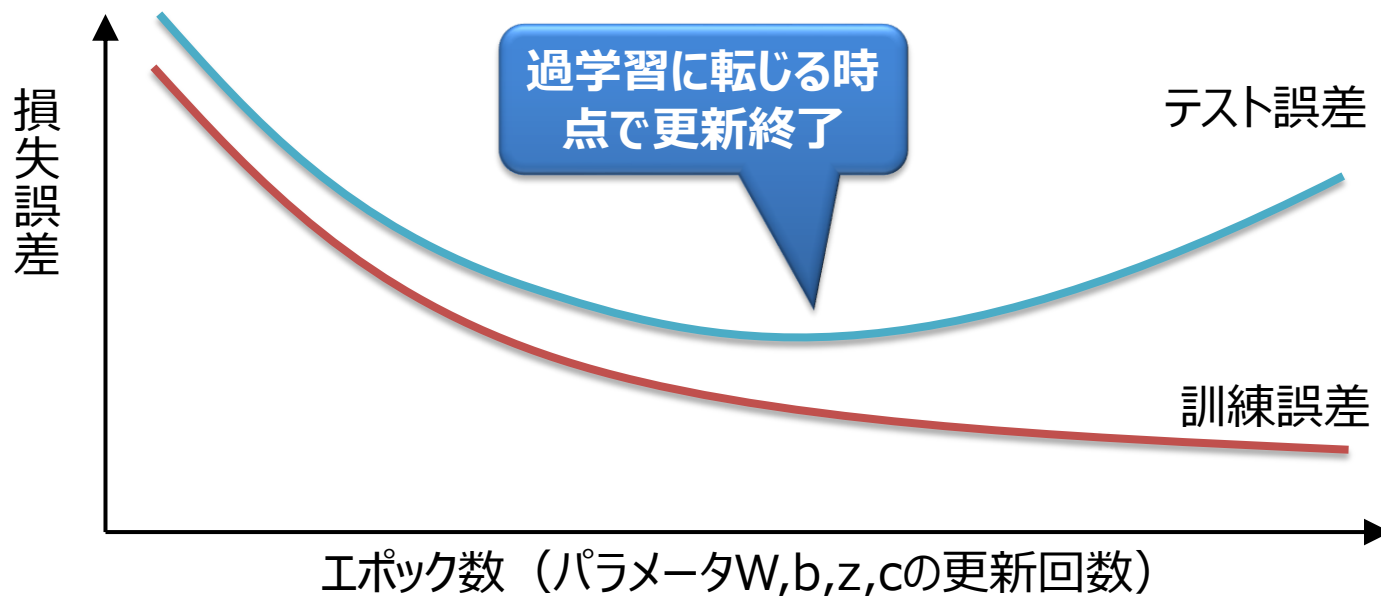


## 4. ディープラーニングの概要と台風分析について

### (2) ディープラーニングの概要 (W,b,z,cを最適化する方法)

#### ④パラメータW,b,z,cの更新

- ・損失（誤差）が最小になるようにパラメータW,b,z,cを繰り返し更新していく。
- ・テスト誤差が大きくなると過学習の傾向にあるため、適切な更新回数で終了させる必要がある。

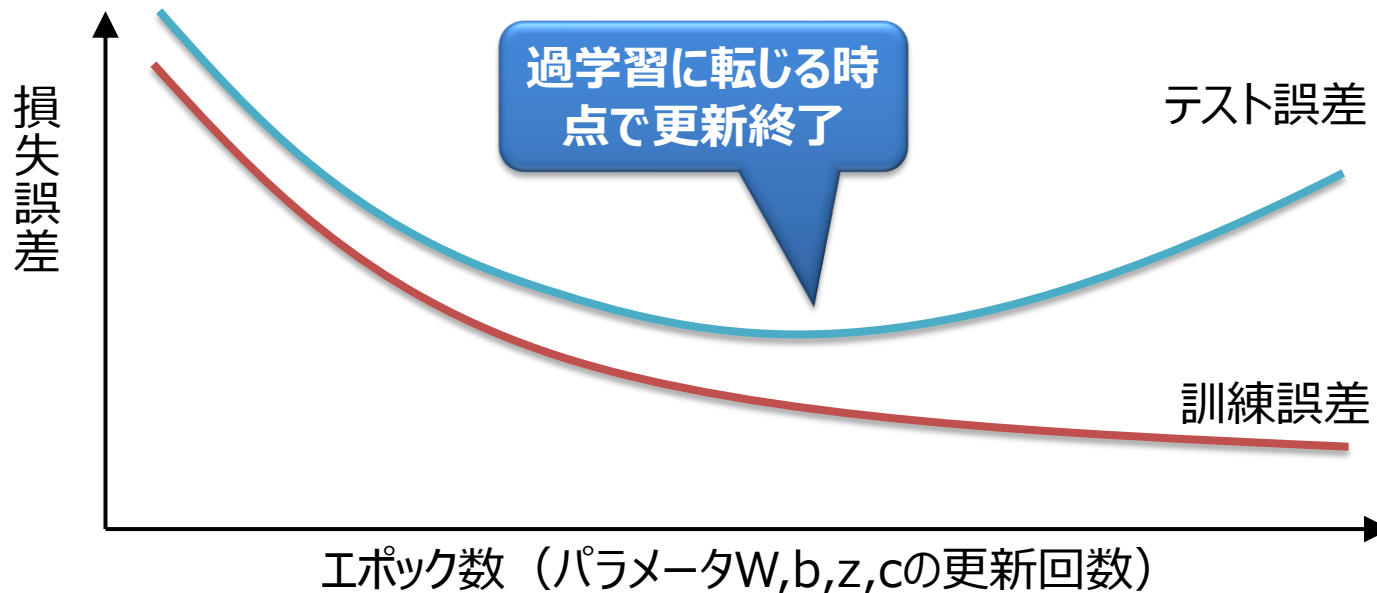


## 4. ディープラーニングの概要と台風分析について

### (2) ディープラーニングの概要 (W,b,z,cを最適化する方法)

#### ④パラメータW,b,z,cの更新

- ・損失（誤差）が最小になるようにパラメータW,b,z,cを繰り返し更新していく。
- ・テスト誤差が大きくなると過学習の傾向にあるため、適切な更新回数で終了させる必要がある。



その他、確率的勾配降下法や交差検証など、効率的かつ精緻に学習させるためのテクニックが活用されている(説明は省略)。

# 4. ディープラーニングの概要と台風分析について

## (2) ディープラーニングの概要 (事例)

<You tube動画レコメンド機能>

- You tubeの動画レコメンド機能ではディープラーニングを「候補抽出」と「再生時間の予測」の2段階で活用している。

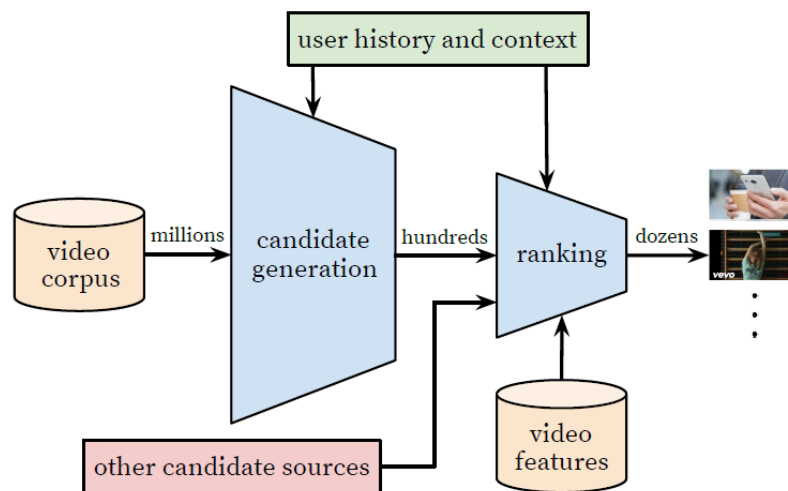


Figure 2: Recommendation system architecture demonstrating the “funnel” where candidate videos are retrieved and ranked before presenting only a few to the user.

出典 : Deep Neural Networks for YouTube Recommendations  
(Google Research HPより)



# 4. ディープラーニングの概要と台風分析について

## (2) ディープラーニングの概要 (事例)

<You tube動画レコメンド機能>

- You tubeの動画レコメンド機能ではディープラーニングを「候補抽出」と「再生時間の予測」の2段階で活用している。

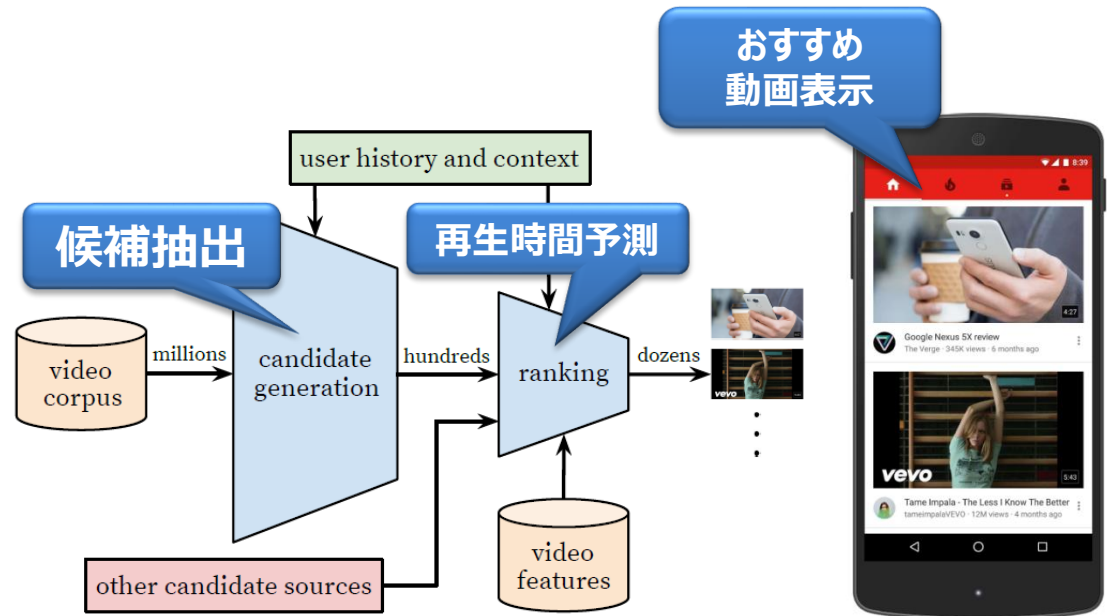


Figure 2: Recommendation system architecture demonstrating the “funnel” where candidate videos are retrieved and ranked before presenting only a few to the user.

出典 : Deep Neural Networks for YouTube Recommendations  
(Google Research HPより)

# 4. ディープラーニングの概要と台風分析について

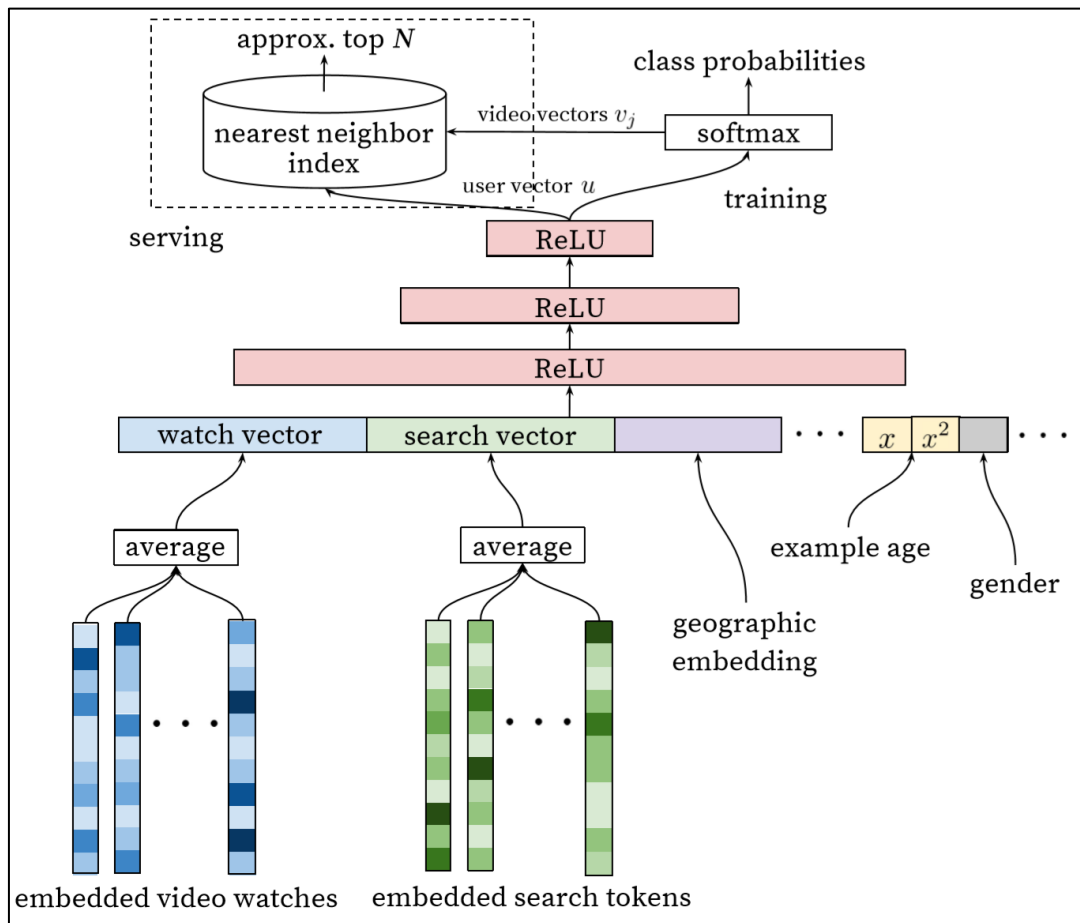
## (2) ディープラーニングの概要 (事例)

<You tube動画レコメンド機能>

• You tubeの動画レコメンド機能ではディープラーニングを「候補抽出」と「再生時間の予測」の2段階で活用している。

• まず「候補抽出」でユーザーが再生してくれそうな動画リストを数百程度に絞込をする。

### ■ 候補抽出のネットワーク



出典：Deep Neural Networks for YouTube Recommendations  
(Google Research HPより)

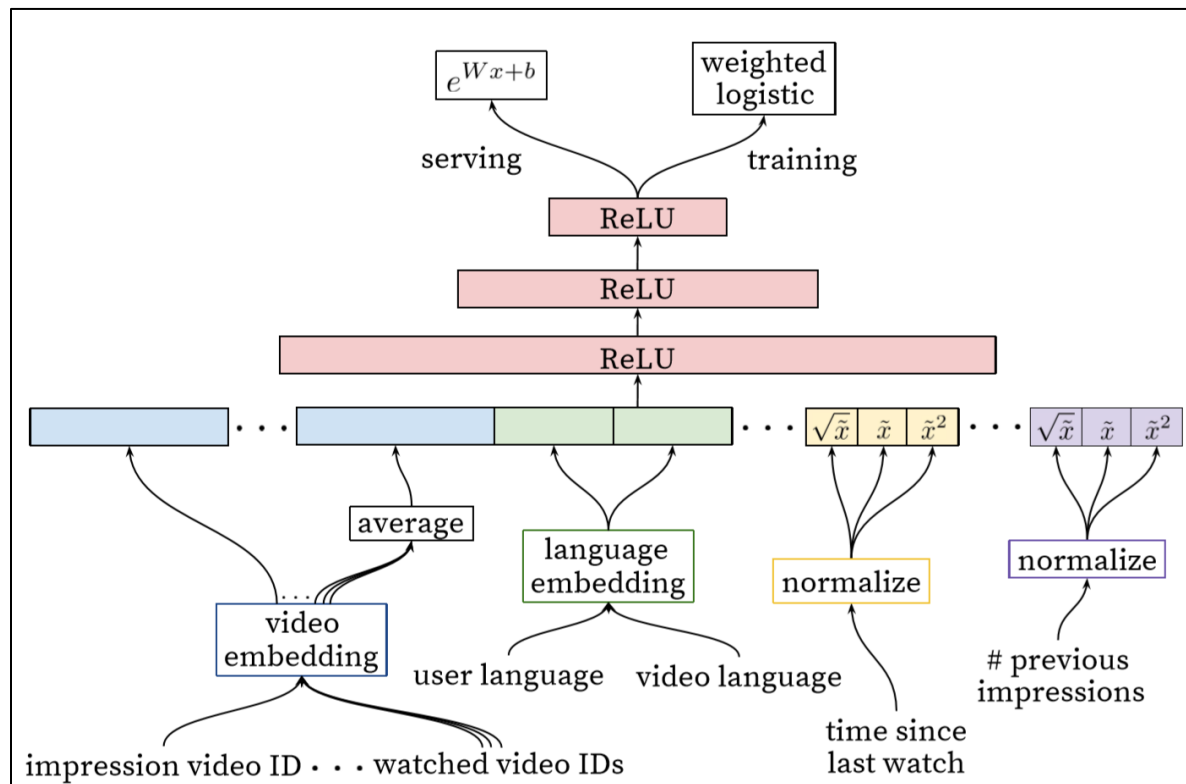
# 4. ディープラーニングの概要と台風分析について

## (2) ディープラーニングの概要 (事例)

<You tube動画レコメンド機能>

- You tubeの動画レコメンド機能ではディープラーニングを「候補抽出」と「再生時間の予測」の2段階で活用している。
- まず「候補抽出」でユーザーが再生してくれそうな動画リストを数百程度に絞込をする。
- 次に「候補抽出」で候補となった動画の「再生時間を予測」する。これによって釣り動画（タイトルやサムネイルで誇張して中身がそぐわない動画）を排除することが可能となる。

### ■ 再生時間予測のネットワーク



出典：Deep Neural Networks for YouTube Recommendations (Google Research HPより)

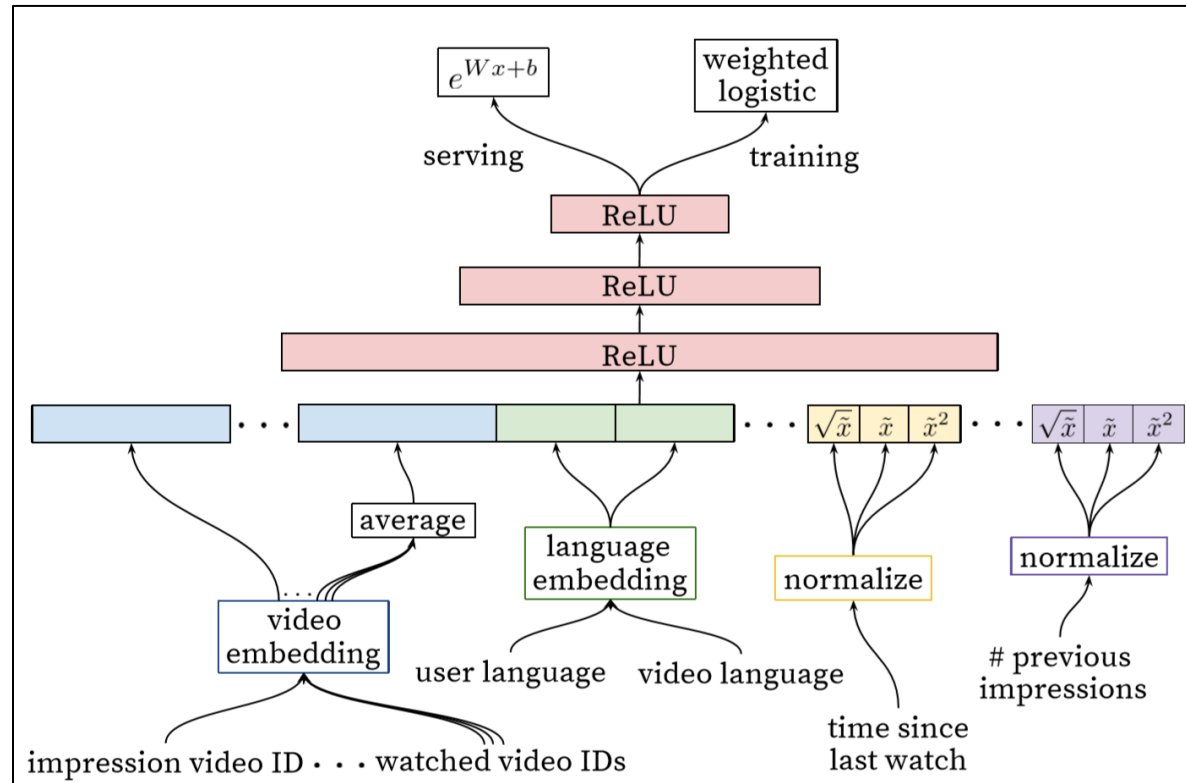
# 4. ディープラーニングの概要と台風分析について

## (2) ディープラーニングの概要 (事例)

<You tube動画レコメンド機能>

- You tubeの動画レコメンド機能ではディープラーニングを「候補抽出」と「再生時間の予測」の2段階で活用している。
- まず「候補抽出」でユーザーが再生してくれそうな動画リストを数百程度に絞込をする。
- 次に「候補抽出」で候補となった動画の「再生時間を予測」する。これによって釣り動画（タイトルやサムネイルで誇張して中身がそぐわない動画）を排除することが可能となる。

### ■ 再生時間予測のネットワーク



出典：Deep Neural Networks for YouTube Recommendations (Google Research HPより)

その他、自動運転、不良品検出、医療画像診断、需要予測など、社会のいろいろな場面でディープラーニングが活用されている。

## 4. ディープラーニングの概要と台風分析について

### (3) ディープラーニングを用いた台風損害予測の概要

#### ① データセット

- ・過去10年間（2010～2019年）のうち、沖縄県内の損害の大きい5台風に限定した。
- ・データセットは各台風通過時点の当社火災保険契約と支払データ、外部データとして各台風の情報（風速、経路等）を用いた。

データセット（入力層に入れるデータ）	
2018年台風24号	<b>■ 台風通過時点の 保有契約 + 支払データ</b>  <b>■ 台風情報（風速、経路等）</b>
2012年台風17号	
2014年台風8号	
2011年台風2号	
2011年台風9号	

# 4. ディープラーニングの概要と台風分析について

## (3) ディープラーニングを用いた台風損害予測の概要

### ① データセット

- ・過去10年間（2010～2019年）のうち、沖縄県内の損害の大きい5台風に限定した。
- ・データセットは各台風通過時点の当社火災保険契約と支払データ、外部データとして各台風の情報（風速、経路等）を用いた。

データセット（入力層に入れるデータ）

イメージ図

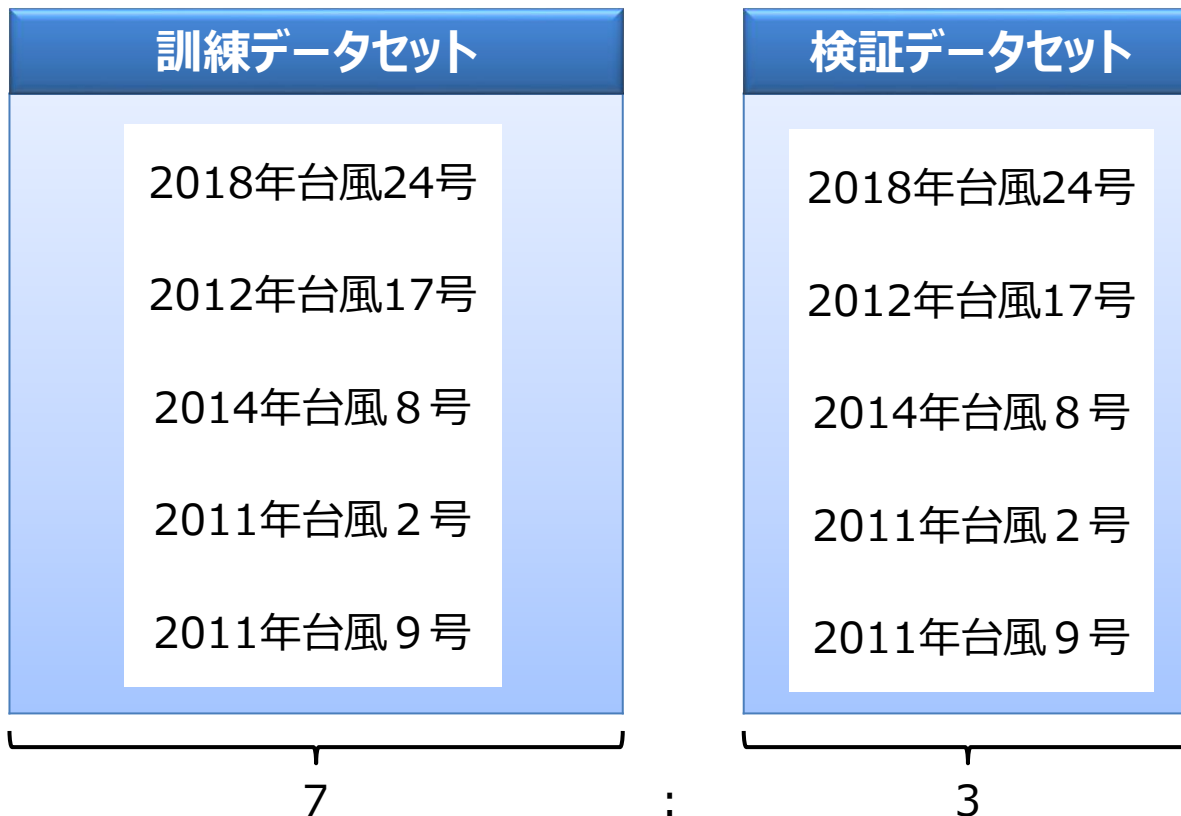
x1	x2	x3	x4		x20	x21	x22	
台風号数	始期日	物件種類	構造		所在地	最大風速	最大瞬間風速	
2018年24号	2011/5/1	住宅	T構造		那覇市	34	53	
2018年24号	...	...	...	...	...	...	...	...
2018年24号	...	...	...	...	...	...	...	...
2018年24号	...	...	...	...	...	...	...	...
2018年24号	...	...	...	...	...	...	...	...
2014年8号	...	...	...		...	...	...	
2014年8号	...	...	...		...	...	...	
2014年8号	...	...	...		...	...	...	
2012年17号	...	...	...		...	...	...	
2012年17号	...	...	...		...	...	...	
2012年17号	...	...	...		...	...	...	
2011年9号	...	...	...		...	...	...	
2011年9号	...	...	...		...	...	...	
2011年9号	...	...	...		...	...	...	

# 4. ディープラーニングの概要と台風分析について

## (3) ディープラーニングを用いた台風損害予測の概要

### ② データセットの分割

- ・データセットを7：3の割合で「訓練データ」と「検証データ」に分割した。
- ・「訓練データ」で学習させ、検証データで予測結果の精度を検証した。



# 4. ディープラーニングの概要と台風分析について

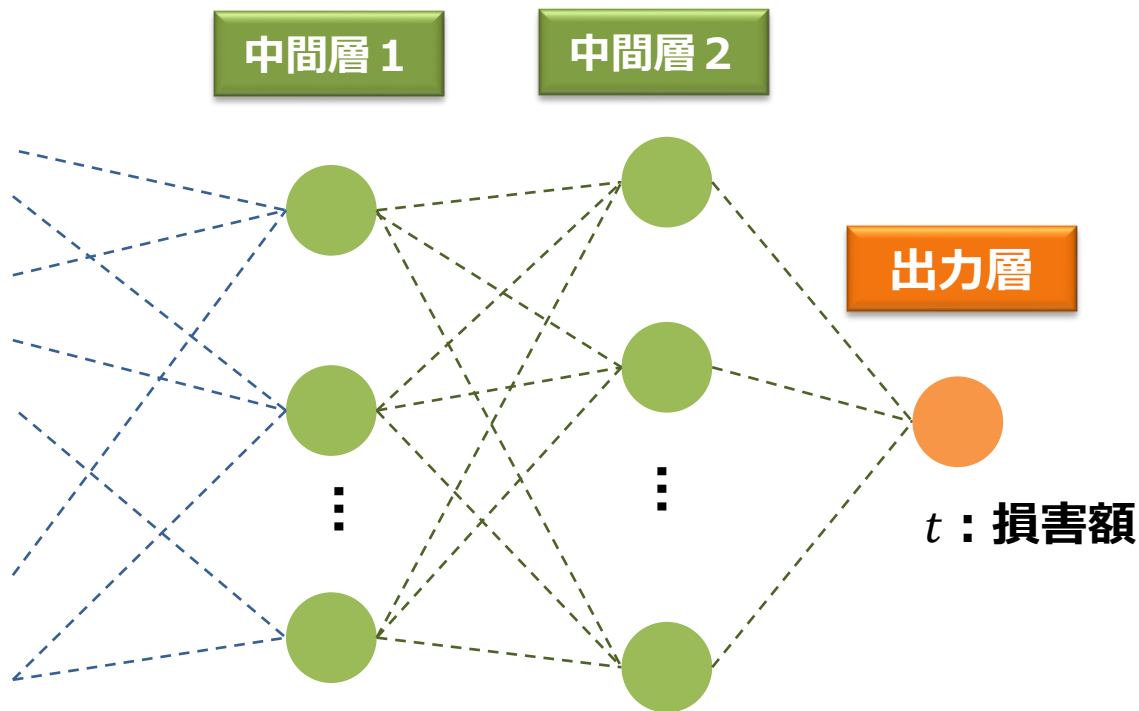
## (3) ディープラーニングを用いた台風損害予測の概要

### ③入力データ

- ・契約 1 件ごとの情報を入力層に入れて損害額を予測する。

#### 入力層

- $x_1$  : 台風号数
- $x_2$  : 始期日
- $x_3$  : 物件種類  
(住宅、工場…)
- $x_4$  : 構造級別  
(M構造、T構造…)
- $x_5$  : 自己負担額
- $x_6$  : 保険金額 (支払限度額)
- $x_7$  : 瞬間最大風速
- ⋮



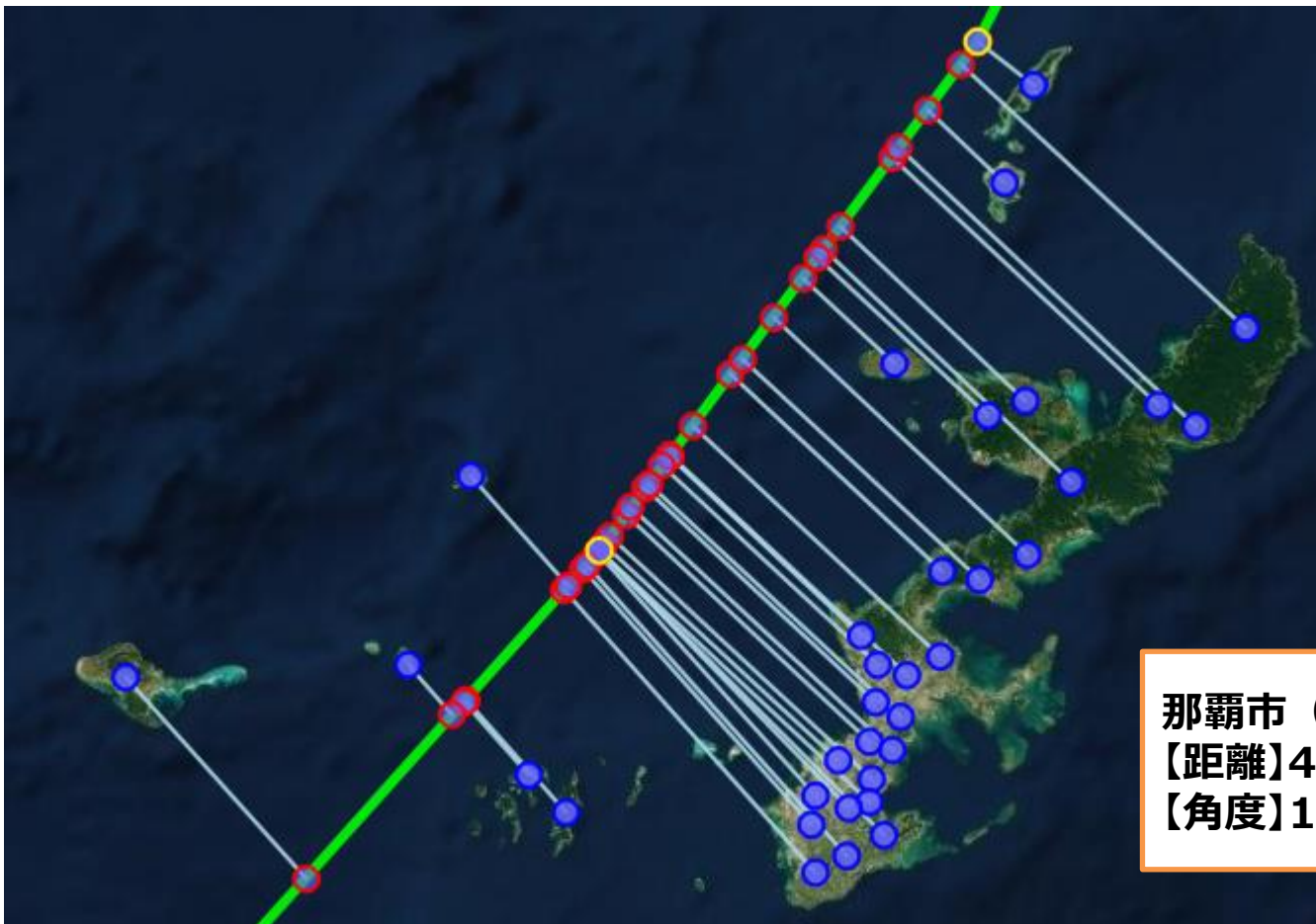


# 4. ディープラーニングの概要と台風分析について

## (3) ディープラーニングを用いた台風損害予測の概要

### ④ 台風経路と所在地の関係

- ・市町村別に台風中心経路との最接近距離・角度の情報を契約データに紐付けた。



那覇市 (台風2018年24号)  
【距離】42666.88 [m]  
【角度】137.92 [度]

# 4. ディープラーニングの概要と台風分析について

## (3) ディープラーニングを用いた台風損害予測の概要

### ⑤その他

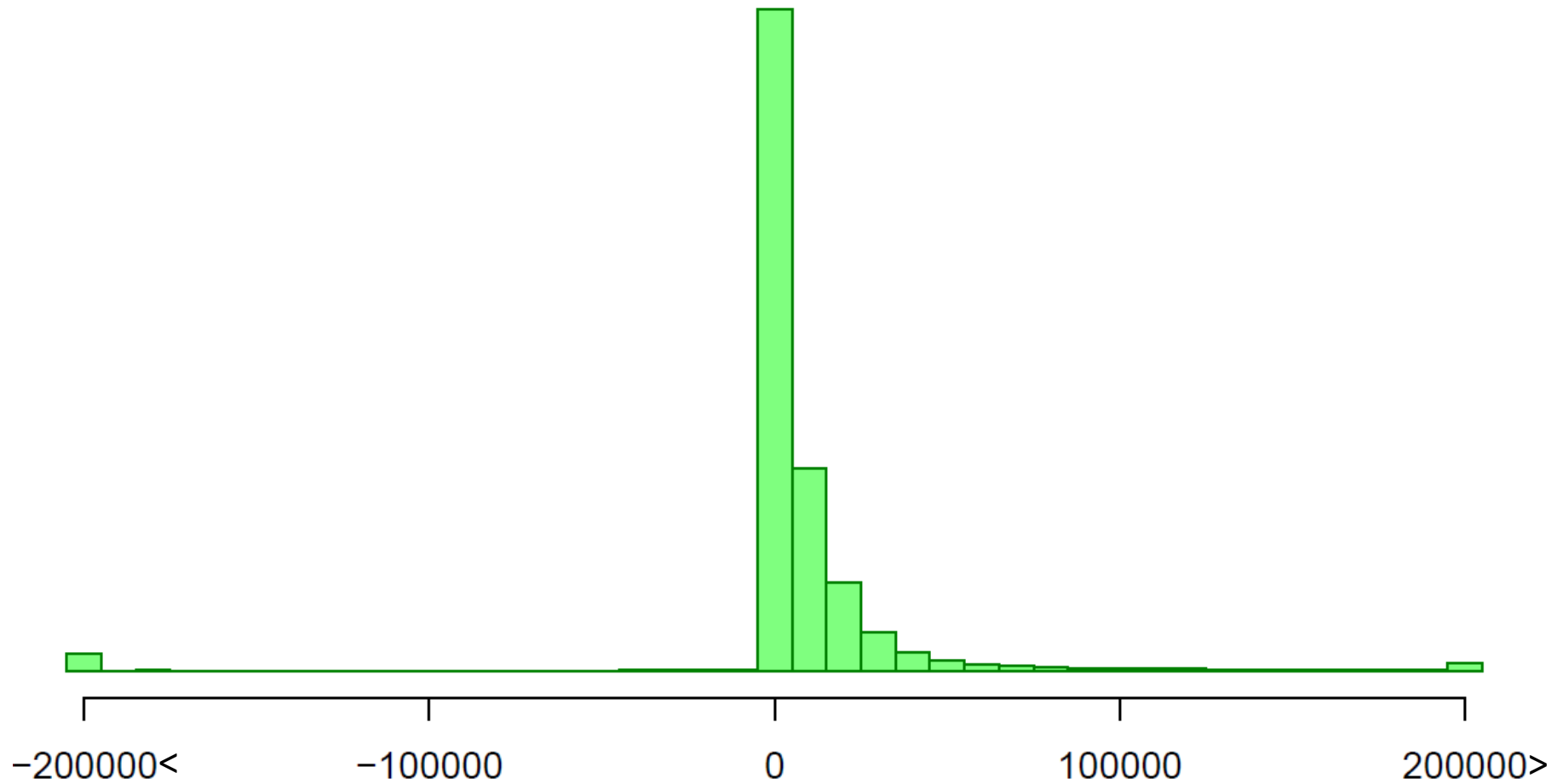
その他のモデル実装の概要は以下の通り・・・（詳細割愛）

- ・ソフトはR言語、kerasパッケージを使用した。
- ・損失関数は平均2乗誤差を用いた。
- ・訓練データを8:2で分割し、交差検証によって学習を行った。
- ・文字列はone-hot-encodingを行った。
- ・量的変数は標準化、カテゴリ変数は中心化と行った。
- ・その他、学習データのクレンジングや活性化関数、層の数、エポック数、バッチサイズを変化させたり、正則化項やドロップアウトを取り入れたりして様々なパターンで検証を行った。

# 4. ディープラーニングの概要と台風分析について

## (4) 予測結果

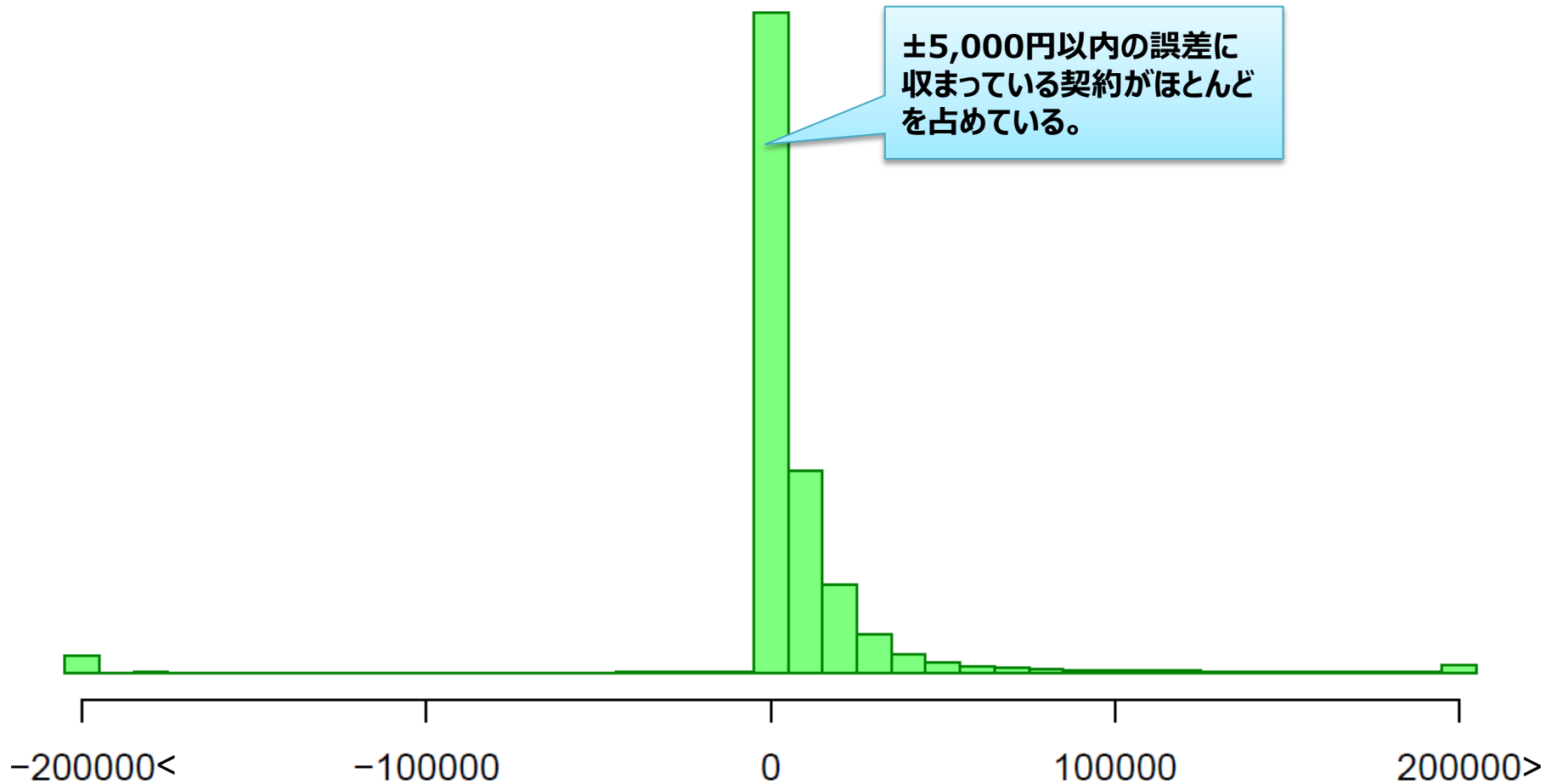
・検証データについて、予測値と実績値の誤差（予測－実績）をヒストグラムに表した。



# 4. ディープラーニングの概要と台風分析について

## (4) 予測結果

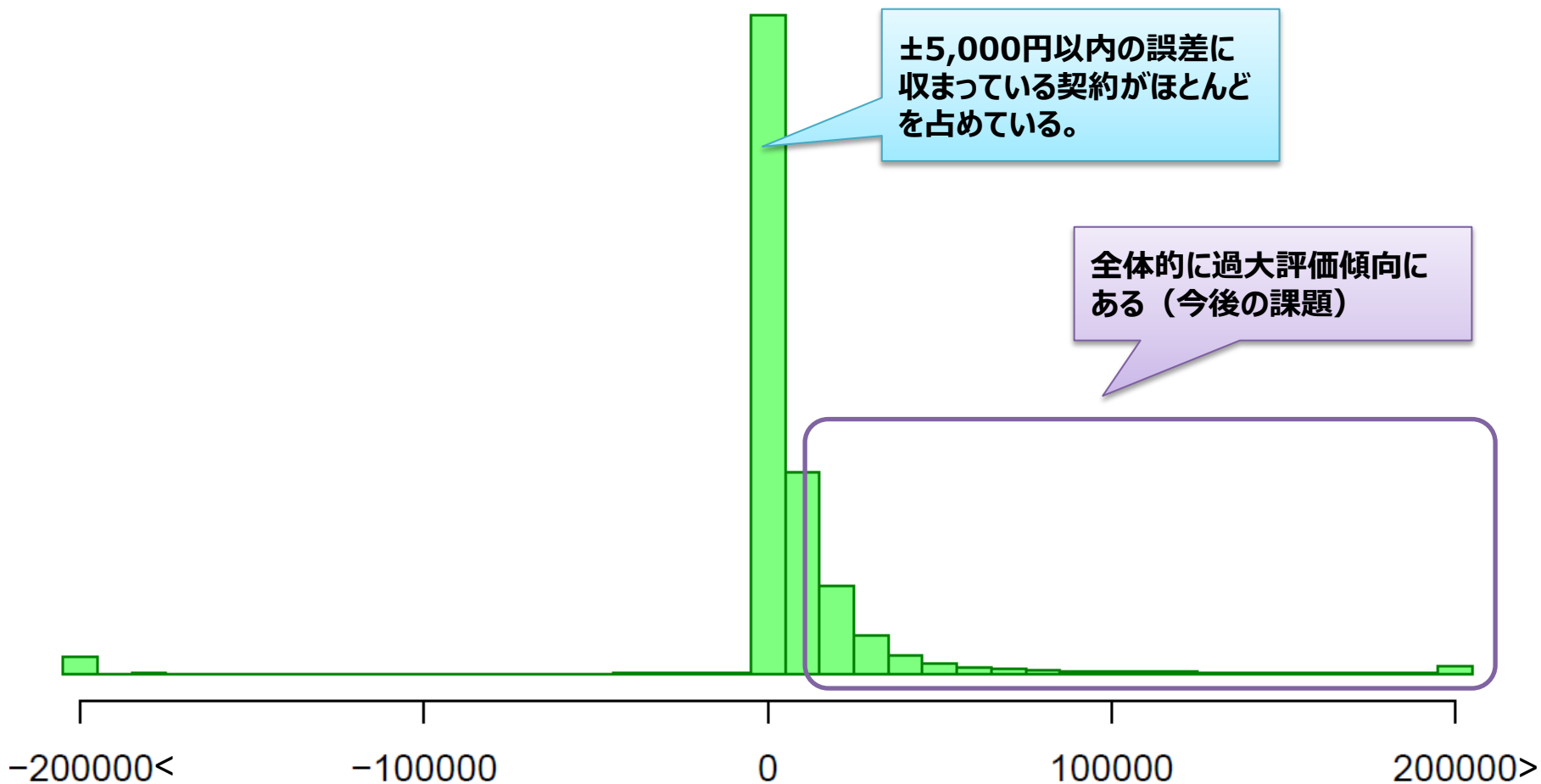
・検証データについて、予測値と実績値の誤差（予測－実績）をヒストグラムに表した。



# 4. ディープラーニングの概要と台風分析について

## (4) 予測結果

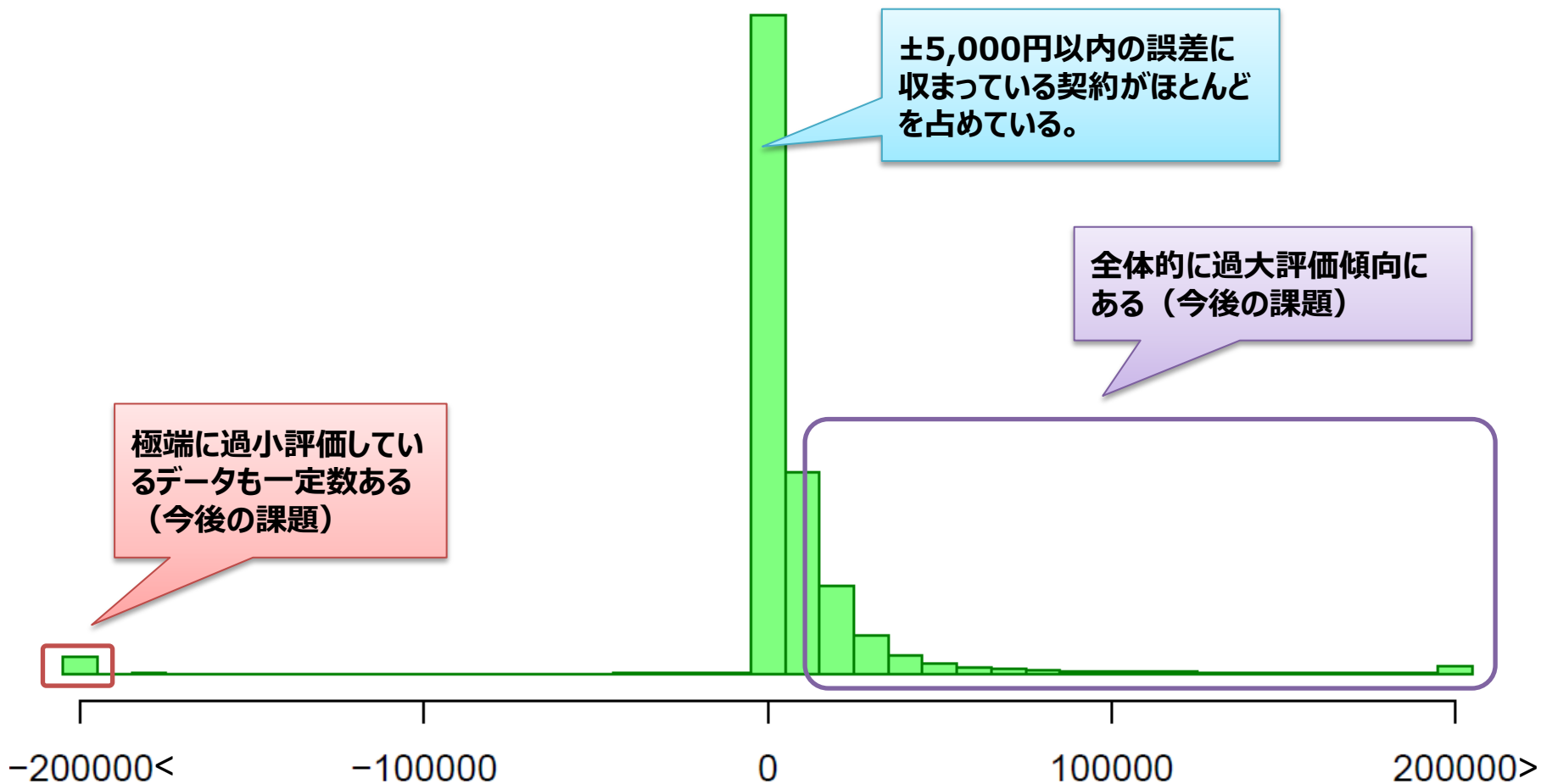
・検証データについて、予測値と実績値の誤差（予測－実績）をヒストグラムに表した。



# 4. ディープラーニングの概要と台風分析について

## (4) 予測結果

・検証データについて、予測値と実績値の誤差（予測－実績）をヒストグラムに表した。



# 4. ディープラーニングの概要と台風分析について

## (5) 今後の課題

今後より精緻な予測モデルが開発できるよう、以下の観点からデータ整備やモデル開発に取り組んでいきたい。

- 1件ごとの所在地に応じた台風被害に関連する情報の取り込み  
例：高度、建物密集度、海面からの距離など
- 建物損害額に影響する経済指標の取り込み  
例：台風通過時点の物価水準、インフレ率、土地価格、消費税など
- 被害の少なかった台風データの取り込み（5台風以外）
- データクレンジング（不明データ・欠損値の補正）
- 特徴量に関する説明（どの特徴量がより大きな情報を持っているか、説明可能な A I）

# AGENDA

## 5. まとめ

- (1) 機械学習とディープラーニングの比較・考察
- (2) 今後の展望



# 5. まとめ

## (1) 機械学習とディープラーニングの比較・考察

機械学習とディープラーニングの主な特徴は以下の通り。

### 機械学習

- ☑学習用データの量はディープラーニングに比べ少ない
- ☑特徴量の選択や様々なモデルの組合せを行うことが可能  
⇒ただし、データが多くなるほどパラメータ調整に時間がかかる。

### ディープラーニング

- ☑大量データから最適な特徴量を学習できる。
- ☑モデルに精通しなくとも、高い精度を得られる。  
⇒ただし、予測結果の理由を説明することが難しい。

# 5. まとめ

## (1) 機械学習とディープラーニングの比較・考察

機械学習とディープラーニングの主な特徴は以下の通り。

### 機械学習

- ☑学習用データの量はディープラーニングに比べ少ない。
- ☑特徴量の選択や様々なモデルの組合せを行うことが可能  
⇒ただし、データが多くなるほどパラメータ調整に時間がかかる。

### ディープラーニング

- ☑大量データから最適な特徴量を学習できる。
- ☑モデルに精通しなくとも、高い精度を得られる。  
⇒ただし、予測結果の理由を説明することが難しい。

**実務ではどのようなデータ（種類、量、ハードウェアの性能）を使用するか、どのような分析を行いたいかで使い分けることが重要**

# 5. まとめ

## (2) 今後の展望

今後より正確性の高いモデルを開発し、下記のような活用を行っていきたい。

☑ 台風襲来前後における被害予測シミュレーション

☑ I B N R 備金の算出

☑ 収支分析

☑ 個々の契約者における台風リスクを評価し、防災対策情報として提供

等

# 5. まとめ

## 参考文献

1. 『入門 Rによる予測モデリング-機械学習を用いたリスク管理のために』 岩沢 宏和・平松 雄司 著 東京図書
2. 『scikit-learnとTensorFlowによる実践機械学習』 Aurélien Géron 著、下田 倫大 監訳、長尾 高弘 訳 O'Reilly Japan, Inc.
3. 『深層学習』 岡谷貴之 講談社サイエンティフィック
4. 『RとKerasによるディープラーニング』 François Chollet、J. J. Allaire 著、瀬戸山 雅人 監訳、長尾 高弘 訳 O'Reilly Japan, Inc.

ご清聴ありがとうございました。

