

1. はじめに

1. はじめに

1. はじめに

制約条件を考慮で
きる多目的遺伝的
アルゴリズム

今後の展望

データの解析

データの前処理

SMOTE による
データ増強

ロジスティック回帰
分析

終わり

進捗状況

堀 由隆

December 10, 2024

1. はじめに

1. はじめに

1. はじめに

制約条件を考慮で
きる多目的遺伝的
アルゴリズム

今後の展望

データの解析

データの前処理

SMOTE による
データ増強

ロジスティック回帰
分析

終わり

したこと

ポスター頑張った

チェックボックスの結果を送信した後の csv ファイル作成

ロジスティック回帰分析のためのデータセット

csv ファイルの解析

1. はじめに

1. はじめに

1. はじめに

制約条件を考慮できる多目的遺伝的アルゴリズム

今後の展望

データの解析

データの前処理

SMOTE によるデータ増強

ロジスティック回帰分析

終わり

ロジスティック回帰分析の例

Pseudo R-squ.

擬似決定係数. 値が高いほど説明変数が目的変数を説明できている.

coef

ロジスティック回帰分析により推定された偏回帰係数.

P

P 値. 一般的に 0.05 未満なら説明変数が目的変数に関係があり, 0.05 以上なら関係がない.

```
=====
Logit Regression Results
=====
Dep. Variable:      registration    No. Observations:    48
Model:              Logit          Df Residuals:        36
Method:             MLE            Df Model:            3
Date:               Mon, 01 Jan 2024    Pseudo R-squ.:      0.2588
Time:               20:42:48           Log-likelihood:     -19.615
converged:          True              LL-Null:            -26.463
Covariance Type:    nonrobust          LLR p-value:        0.003350
=====
```

	coef	std err	z	P> z	[0.025	0.975]
const	-4.2982	1.555	-2.764	0.006	-7.346	-1.250
sex	0.5251	0.842	0.624	0.533	-1.124	2.174
student	2.0147	0.841	2.396	0.017	0.367	3.663
stay time	0.0390	0.019	2.087	0.037	0.002	0.076

sex (男女)	student (学生か)	stay time (閲覧時間)	会員登録をする確率
0	1	67	0.581287
0	0	80	0.235077
1	0	19	0.045979
0	1	92	0.786263
1	1	56	0.604531

ロジスティック回帰分析の例

偏回帰係数の結果から、会員登録をする確率は以下の式で推定できる。

$$\text{登録する確率} = -4.2982 + 0.5251 \times (\text{男女}) + 2.0147 \times (\text{学生か}) + 0.0390 \times (\text{閲覧時間}) \quad (1)$$

sex (男女)	student (学生か)	stay time (閲覧時間)	会員登録をする確率
0	1	67	0.581287
0	0	80	0.235077
1	0	19	0.045979
0	1	92	0.786263
1	1	56	0.604531

1. はじめに

1. はじめに

1. はじめに

制約条件を考慮で
きる多目的遺伝的
アルゴリズム

今後の展望

データの解析

データの前処理

SMOTE による
データ増強

ロジスティック回帰
分析

終わり

ロジスティック回帰は、目的変数が二値（0 または 1）を取る場合に適した回帰モデル。説明変数と目的変数の関係をロジスティック関数を通じて表現し、確率を求める。献立の UX に関する質問は、yes or no の二値を扱うので説明変数の分析にロジスティック回帰分析を扱うことになった

ロジスティック回帰モデルは次の式で表される：

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}} \quad (2)$$

ここで、 $P(Y = 1|X)$ は目的変数 Y が 1 になる確率。 β_0 は切片で、 $\beta_1, \beta_2, \dots, \beta_k$ は各説明変数 X_1, X_2, \dots, X_k に対応する回帰係数。

3.1 多目的最適化とパレート最適解

6/13

多目的最適化は、ある制約条件のもと、複数の目的関数を最大化、あるいは最小化する手法である。全ての目的関数を最大化、あるいは最小化するような最適解が存在するとは言えないため、パレート最適という概念を導入する必要がある。

多目的最適化の定式化

minimize
 x

$\{f_1(x), f_2(x), \dots, f_n(x)\}$

subject to

$g_k(x) \leq 0$

$k = 1, 2, \dots, m$

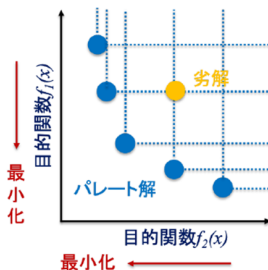


図12 パレート解のイメージ

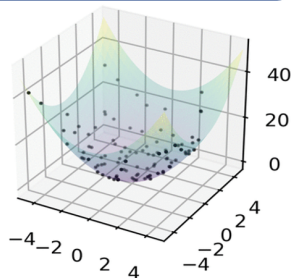


図13 解探索のイメージ (粒子群最適化)

1. はじめに

1. はじめに

1. はじめに

制約条件を考慮できる多目的遺伝的アルゴリズム

今後の展望

データの解析

データの前処理

SMOTE によるデータ増強

ロジスティック回帰分析

終わり

結果の分析

ポスター発表で使う数値実験は以下のような使われ方を想定した csv ファイルの説明を行う

1. 家族

2. 父疾患持ち

3. 子供アレルギー持ち

4. 残り二人は健康者

で献立作成行なった結果 csv ファイルに蓄積していく

name	たんぱく質	炭水化物	糖質	脂質	食塩相当量	食物繊維	ビタミンA	ビタミンB1	ビタミンB2	ビタミンB6	ビタミンB12	ビタミンC
水上sun 男9.4g	109.7g	108.2g	1.8g	2.2g	1.5g	1.5g	1 μg	0.17mg	0.05mg	0.2mg	0 μg	2mg
水上sun 男15.9g	6.4g	5.8g	8.5g	1.5g	0.6g	4 μg	0.12mg	0.09mg	0.13mg	0.9 μg	2mg	
水上sun 男4.7g	4.1g	1.8g	3.9g	0.4g	2.3g	5 μg	0.04mg	0.15mg	0.09mg	0 μg	1mg	
水上sun 男2.7g	4.2g	1.5g	1.5g	1g	2.7g	111 μg	0.08mg	0.05mg	0.05mg	1.8 μg	5mg	
水上sun 男14.9g	12.6g	8.5g	20.5g	1.2g	4.1g	224 μg	0.24mg	0.48mg	0.34mg	0.9 μg	107mg	
水上sun 男23.6g	2.9g	2.6g	4.3g	1.3g	0.3g	8 μg	0.1mg	0.04mg	0.12mg	1.1 μg	1mg	
水上sun 男9.6g	60.7g	60.2g	1g	1.4g	0.5g	0 μg	0.03mg	0.14mg	0.07mg	9.1 μg	12mg	
高血圧のノ9.4g	109.7g	108.2g	1.8g	2.2g	1.5g	1 μg	0.17mg	0.05mg	0.2mg	0 μg	2mg	
高血圧のノ22.3g	121.3g	120.2g	10.8g	2.8g	1.1g	77 μg	0.15mg	0.31mg	0.39mg	0.6 μg	4mg	
高血圧のノ9.9g	20.6g	7.8g	16.5g	0.8g	12.8g	91 μg	0.13mg	0.26mg	0.15mg	0.3 μg	23mg	
高血圧のノ19.4g	4.2g	3.6g	25.4g	1.6g	0.6g	294 μg	0.11mg	0.55mg	0.15mg	2.1 μg	8mg	
高血圧のノ14.8g	77.1g	75.9g	7g	0.2g	1.2g	1 μg	0.09mg	0.14mg	0.09mg	0.4 μg	12mg	
高血圧のノ2.7g	18.8g	15.3g	8.3g	1.2g	3.5g	145 μg	0.08mg	0.04mg	0.11mg	0 μg	24mg	
高血圧のノ9.8g	19.8g	18.6g	12.1g	0.6g	1.2g	8 μg	0.02mg	0.11mg	0.03mg	0.8 μg	11mg	
アレルギー-15g	12.7g	10.9g	7.1g	0.8g	1.8g	7,020 μg	0.3mg	1.15mg	0.5mg	23.3 μg	16mg	
アレルギー-8.7g	18.3g	15.1g	4.6g	0.9g	3.2g	82 μg	0.12mg	0.14mg	0.34mg	0.1 μg	19mg	
アレルギー-9.7g	28.4g	25.8g	9.9g	1.1g	2.6g	6 μg	0.12mg	0.31mg	0.14mg	1.4 μg	1mg	

1. はじめに

1. はじめに

1. はじめに

制約条件を考慮できる多目的遺伝的アルゴリズム

今後の展望

データの解析

データの前処理

SMOTE によるデータ増強

ロジスティック回帰分析

終わり

願望

献立の表示→ロジスティック回帰分析とその結果により、ユーザに合わせた献立をお勧めする（UX 項目の推定）
→出力された献立の評価の繰り返しでユーザごとにパーソナライズされた献立の表示



1. はじめに

1. はじめに

1. はじめに

制約条件を考慮できる多目的遺伝的アルゴリズム

今後の展望

データの解析

データの前処理

SMOTE によるデータ増強

ロジスティック回帰分析

終わり

普通にロジスティック回帰分析

サンプルに偏りがあり（不均衡なデータ）だったので、分析の途中でエラーが出ていた。

そこで,smote を使い、オーバーサンプリングを行った。

```
File "C:\Users\hori\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.11_qbz5n2kfp\numpy\numpy\linalg\linalg.py", line 112, in _raise_linalgerror_singular
    raise LinAlgError("Singular matrix")
umpy.linalg.LinAlgError: Singular matrix
```

1. はじめに

1. はじめに

1. はじめに

制約条件を考慮で
きる多目的遺伝的
アルゴリズム

今後の展望

データの解析

データの前処理

SMOTE による
データ増強

ロジスティック回帰
分析

終わり

不均衡データへの対応

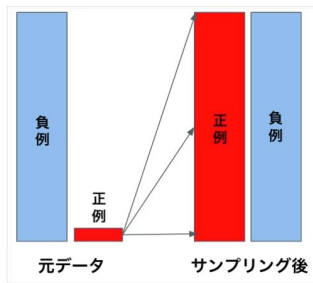
10/13

SMOTE を利用したオーバーサンプリング

サンプルに偏りがある不均衡データの場合、少数クラスの情報が十分に活用されない問題が発生する。

SMOTE (Synthetic Minority Over-sampling Technique) は、少数クラスのデータを合成することで、この問題に対処する手法である。

この手法により、分析モデルが少数クラスをより正確に学習できるようになる。



1. はじめに

1. はじめに

1. はじめに

制約条件を考慮できる多目的遺伝的アルゴリズム

今後の展望

データの解析

データの前処理

SMOTE によるデータ増強

ロジスティック回帰分析

終わり

SMOTE (Synthetic Minority Over-sampling Technique) の概要

SMOTE は、少数クラスのデータ点間で線形補間を行い、新たなサンプルを生成することで、不均衡データを改善する手法である。

以下の手順で新しいデータ点を作成する：

- 1 少数クラスの各サンプルに対して、最も近い k 個の少数クラスのサンプルを計算。
- 2 ランダムに 1 つの近傍サンプルを選択。
- 3 選択したサンプルと元のサンプルの間で線形補間を実行。

線形補間の数式

2 つの少数クラスサンプル \mathbf{x}_i と \mathbf{x}_j の間で新しいデータ点 \mathbf{x}_{new} を生成する際の式は次の通り：

$$\mathbf{x}_{\text{new}} = \mathbf{x}_i + \lambda(\mathbf{x}_j - \mathbf{x}_i)$$

ここで、 $\lambda \in [0, 1]$ は一様乱数として生成される重み。

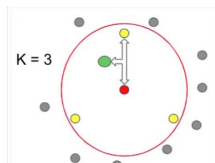
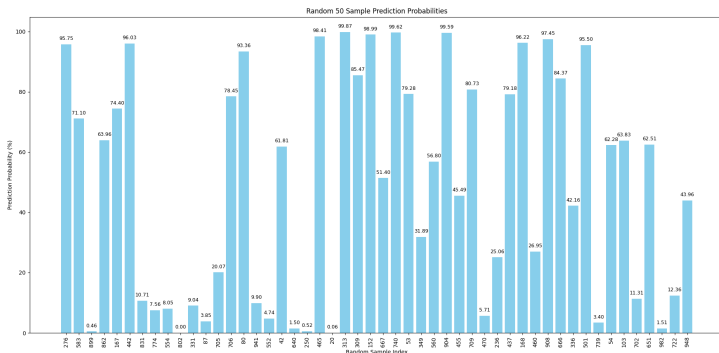


Figure 1: SMOTE のサンプル生成プロセス

smote を用いたロジスティック回帰分析

エラーがなくなり、サンプルレシピ1000個に対して UX 項目（よく使う食材は含まれているか？）の推定を行ったその結果を示す。



1. はじめに

1. はじめに

1. はじめに

制約条件を考慮できる多目的遺伝的アルゴリズム

今後の展望

データの解析

データの前処理

SMOTE によるデータ増強

ロジスティック回帰分析

終わり

1. はじめに

1. はじめに

1. はじめに

制約条件を考慮で
きる多目的遺伝的
アルゴリズム

今後の展望

データの解析

データの前処理

SMOTE による
データ増強

ロジスティック回帰
分析

終わり

すること

smote を使うことで、正しい結果が得られているかの検証

終わりに

得られた結果が正しければ、評価関数に組み込む