みんなのPython勉強会#47 7/10, 2019

データを可視化して数学の理解を助ける

辻 真吾（@tsjshg）

自己紹介

- 東京大学先端科学技術研究センターに所属
- 応用データサイエンスを目指したい
- 日経コンピュータさんに取材してもらいました
- このコミュニティのロゴが拙著なのはゴリ押しではないです

2019/06/24 05:00

Python大全、AI時代の本命言語

AI時代のいまPythonがブレイクする本当の理由

白井 良、清嶋 直樹、浅川 直輝、中田 敦＝日経 xTECH／日経コンピュータ　日経 XTECH

Python人気の原動力はAI（人工知能）だ。AIシステムを開発するプログラミング言語としてPythonが最も使いやすいからだ。データ分析やインフラ管理など、Pythonの用途は広がっている。

2020年春、IPA（情報処理推進機構）が運営する国家試験「基本情報技術者試験」で選択できるプログラミング言語からCOBOLが消え、Pythonが新たに採用される。前身の第二種情報処理技術者試験が産声を上げた1970年からCOBOLは50年間にわたって出題され続けた。だが2020年春の試験からはPythonに入れ替わる。

FE

平成 31 年度　春期
基本情報技術者試験
午後　問題

試験時間	13:00 ～ 15:30（2 時間 30 分）

注意事項
1. 試験開始及び終了は，監督員の時計が基準です。監督員の指示に従ってください。
2. 試験開始の合図があるまで，問題冊子を開いて中を見てはいけません。
3. 答案用紙への受験番号などの記入は，試験開始の合図があってから始めてください。
4. 問題は，次の表に従って解答してください。

問題番号	問 1	問 2 ～ 問 7	問 8	問 9 ～ 問 13
選択方法	必須	4 問選択	必須	1 問選択

午後試験のうち1問がプログラミング言語に関する問題

5. 答案用紙の記入に当たっては，次の指示に従ってください。
(1) 答案用紙は光学式読取り装置で読み取った上で採点しますので，B 又は HB の黒鉛筆で答案用紙のマークの記入方法のとおりマークしてください。マークの濃度がうすいなど，マークの記入方法のとおり正しくマークされていない場合は読み取れないことがあります。特に
に十分注意してください。打
に消し，消しくずを残さない

●問 9～問 13（5 問中 1 問選択）

2020年春期試験からCOBOLの出題を廃止してPythonの出

https://tech.nikkeibp.co.jp/atcl/nxt/column/18/00818/061100001/

お知らせ

- RとStanではじめるベイズ統計モデリングによるデータ分析入門
  - データ分析の基礎から解説
  - フルカラーで見やすい
- 講談社【実践Data Scienceシリーズ】第1弾！
- ほんとは、別の本が第1弾の予定だった・・・

もくじ

- シグモイド関数の形
- Pythonで3D描画をするには
- Mayaviを使った描画
- 4次元空間へ

シグモイド関数

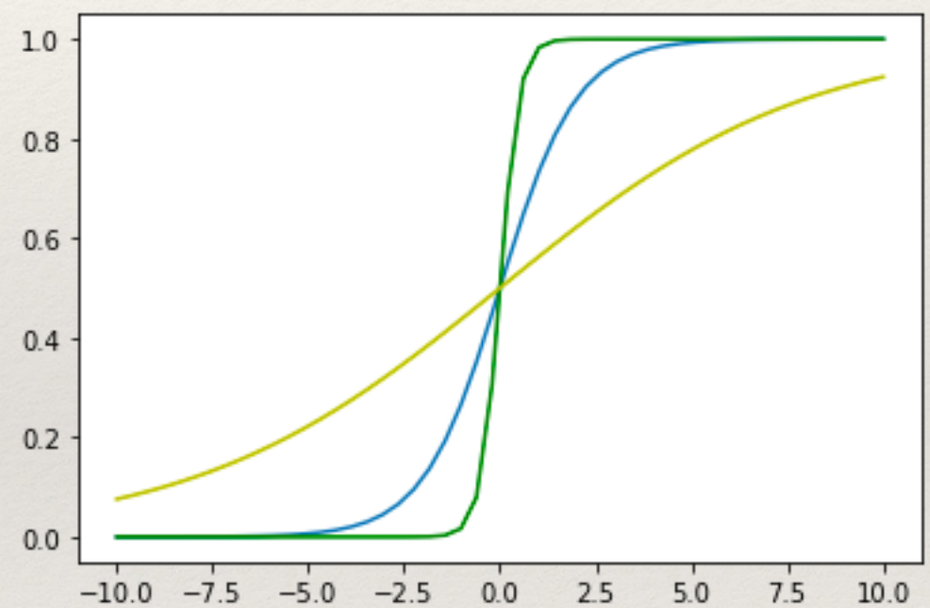
$$y = \frac{1}{1 + e^{-ax}} \quad (a > 0)$$

はじめて見たらどんな形かわからない・・・

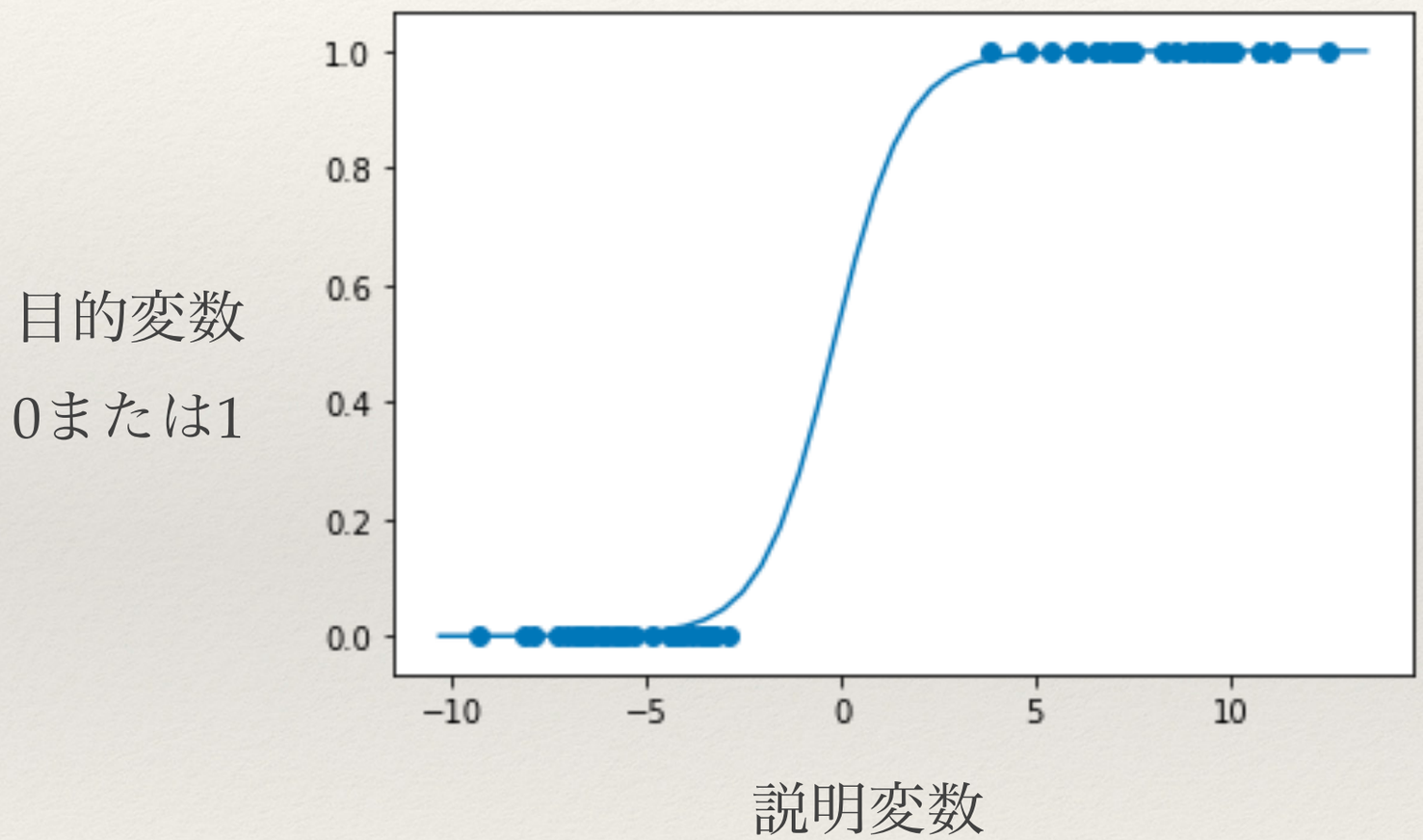
描いてみればいい

```
def sigmoid_f(x, a=1):
    return 1.0 / (1.0 + np.exp(-a*x))

x = np.linspace(-10,10)
plt.plot(x, sigmoid_f(x))
plt.plot(x, sigmoid_f(x, 4), c='g')
plt.plot(x, sigmoid_f(x, 0.25), c='y')
```



ロジステック回帰

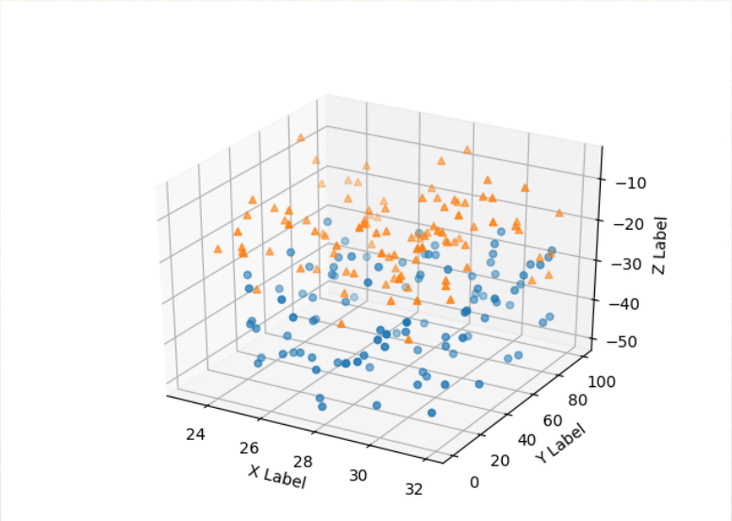


説明変数が1つと目的変数で描画には2次元が必要

3次元までは活用したい

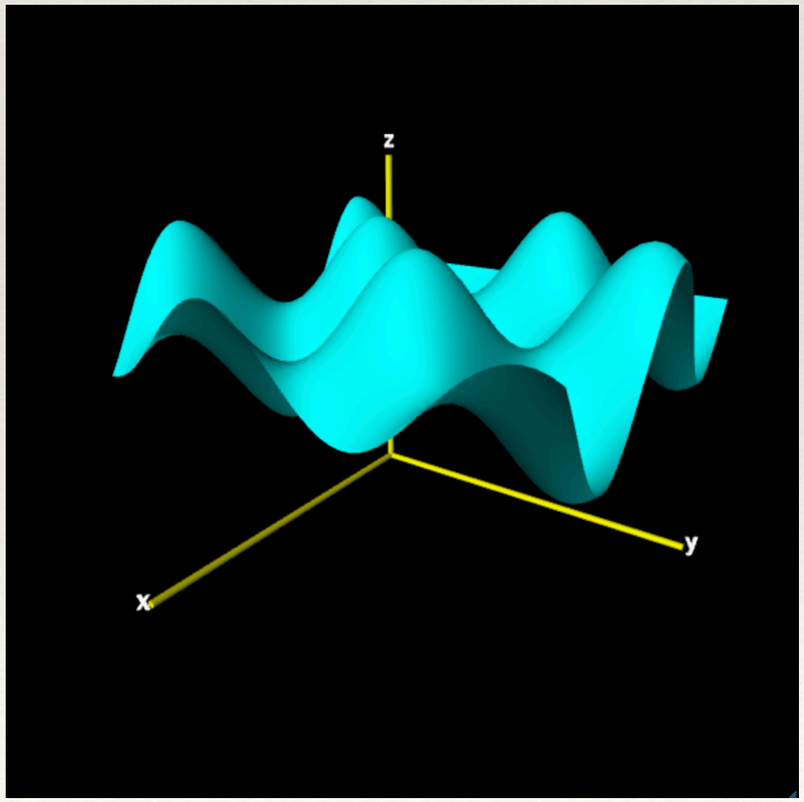
3D描画

- Matplotlibでも描ける
- ゴリゴリ動かしたい？
- VPythonも面白そう
  - GlowScript（よく知らない）と連携してWebとの相性も良い



Matplotlib

https://matplotlib.org/3.1.1/gallery/mplot3d/scatter3d.html



VPython

https://www.glowscript.org/#/user/GlowScriptDemos/folder/Examples/program/Plot3D

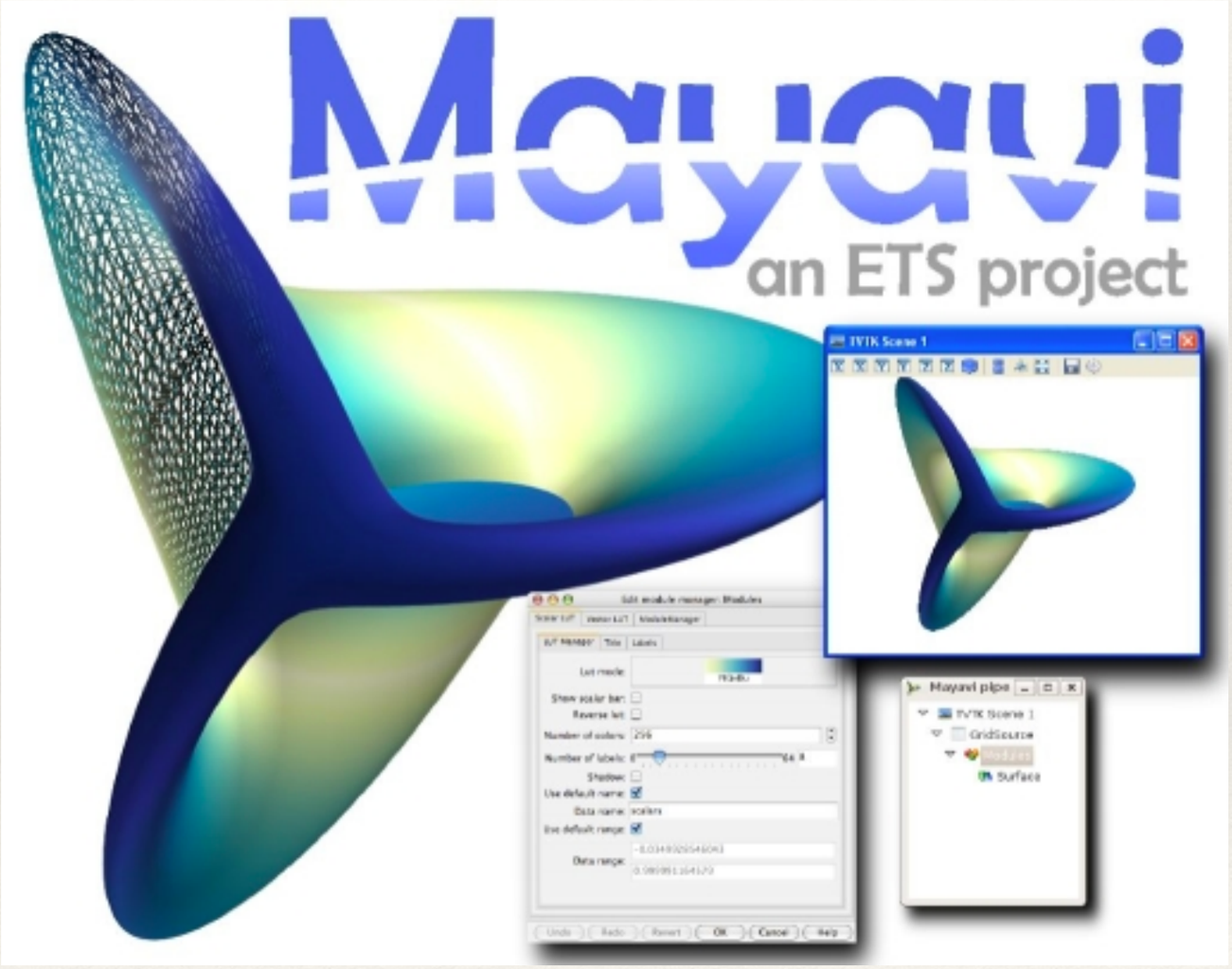
Panda3D

- C++で開発されPythonのAPIがある本格的3D描画ライブラリ
- 2002年が最初のリリース
  - もともとDisneyで使われていた？
- カッコイイけど、ちょっと違う気もする・・・

https://www.panda3d.org/

Mayavi

- 科学計算のための3D描画ライブラリ
- Enthought社が開発している？
- インストールが困難という話もあるが、macのPython3.7.3にはpipで普通に入った



https://docs.enthought.com/mayavi/mayavi/

Jupyter Notebookで動く

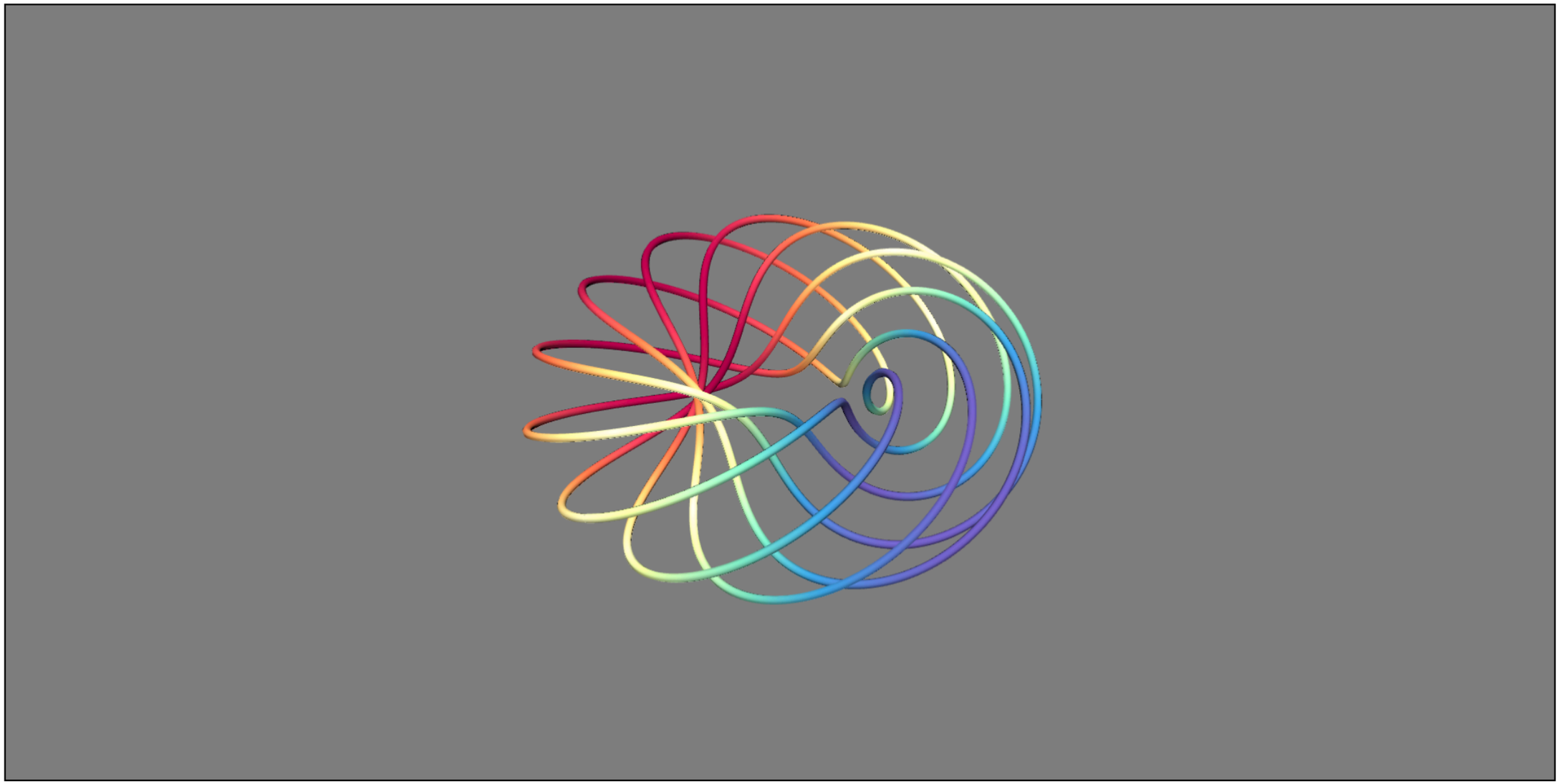
In [1]:
```
from mayavi import mlab
```

In [2]:
```
mlab.init_notebook()
```

Notebook initialized with x3d backend.

In [3]:
```
mlab.test_plot3d()
```

Out[3]:

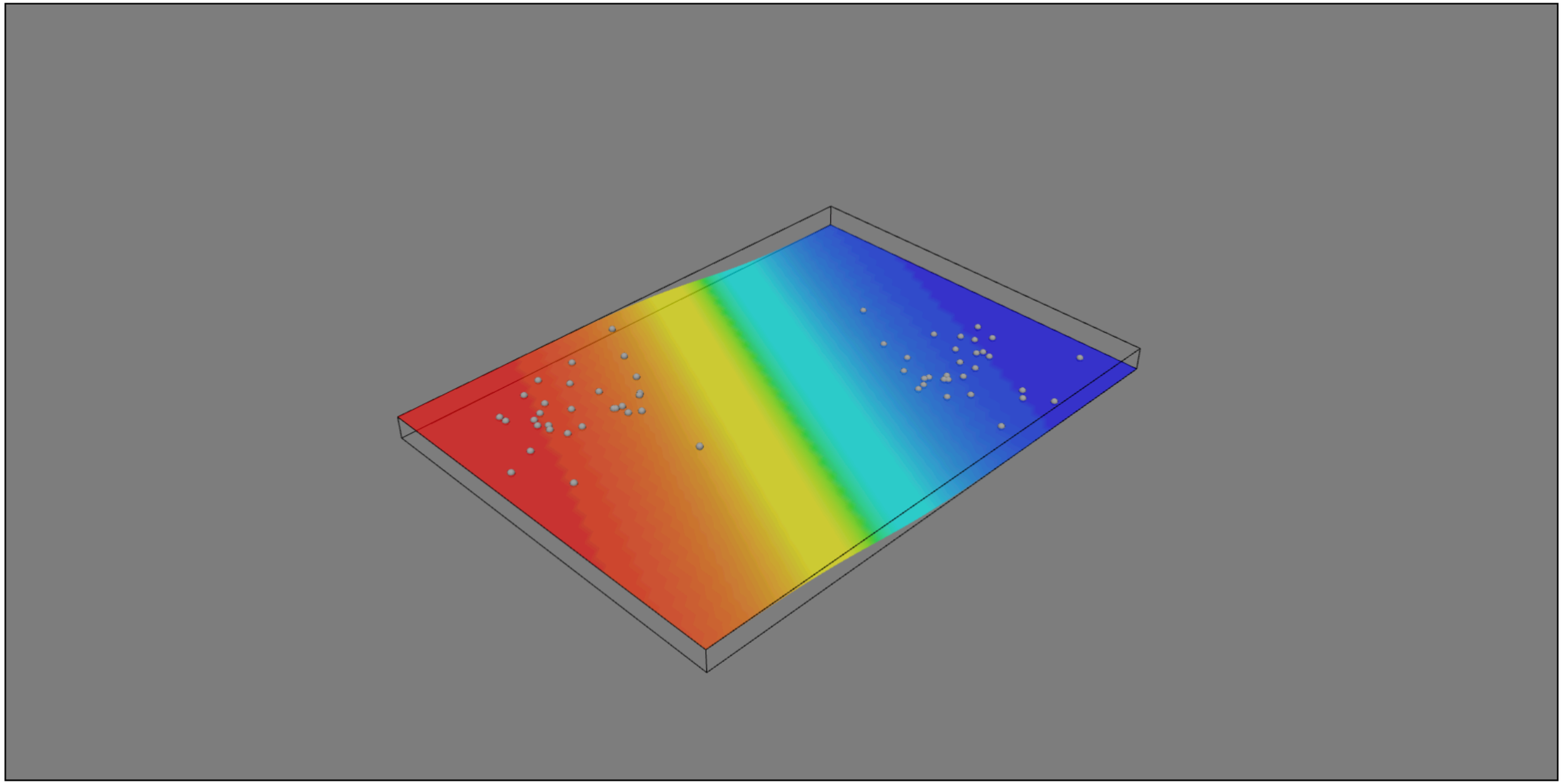


説明変数2＋目的変数1で3次元を使った可視化

```
@np.vectorize
def sigmoid2(x,y,reg):
    _x = np.array([x,y])
    w = np.dot(_x , reg.coef_[0])
    return 1.0 / (1.0 + np.exp(-w))
```

```
x0m, x1m = np.mgrid[min(X[:,0])-1:max(X[:,0])+1:64j, min(X[:,1])-1:max(X[:,1])+1:64j]
```

```
mlab.clf()
_ = mlab.points3d(X[:,0], X[:,1], y, scale_factor=0.3, color=(0.5,0.5,0.5))
mlab.mesh(x0m, x1m, sigmoid2(x0m, x1m, reg=reg2), opacity=0.6)
mlab.outline()
```



数式はイメージが難しい

1.1.11. Logistic regression

Logistic regression, despite its name, is a linear model for classification rather than regression. Logistic regression is also known in the literature as logit regression, maximum-entropy classification (MaxEnt) or the log-linear classifier. In this model, the probabilities describing the possible outcomes of a single trial are modeled using a logistic function.

Logistic regression is implemented in LogisticRegression . This implementation can fit binary, One-vs-Rest, or multinomial logistic regression with optional ℓ_1, ℓ_2 or Elastic-Net regularization. Note that regularization is applied by default.

As an optimization problem, binary class ℓ_2 penalized logistic regression minimizes the following cost function:

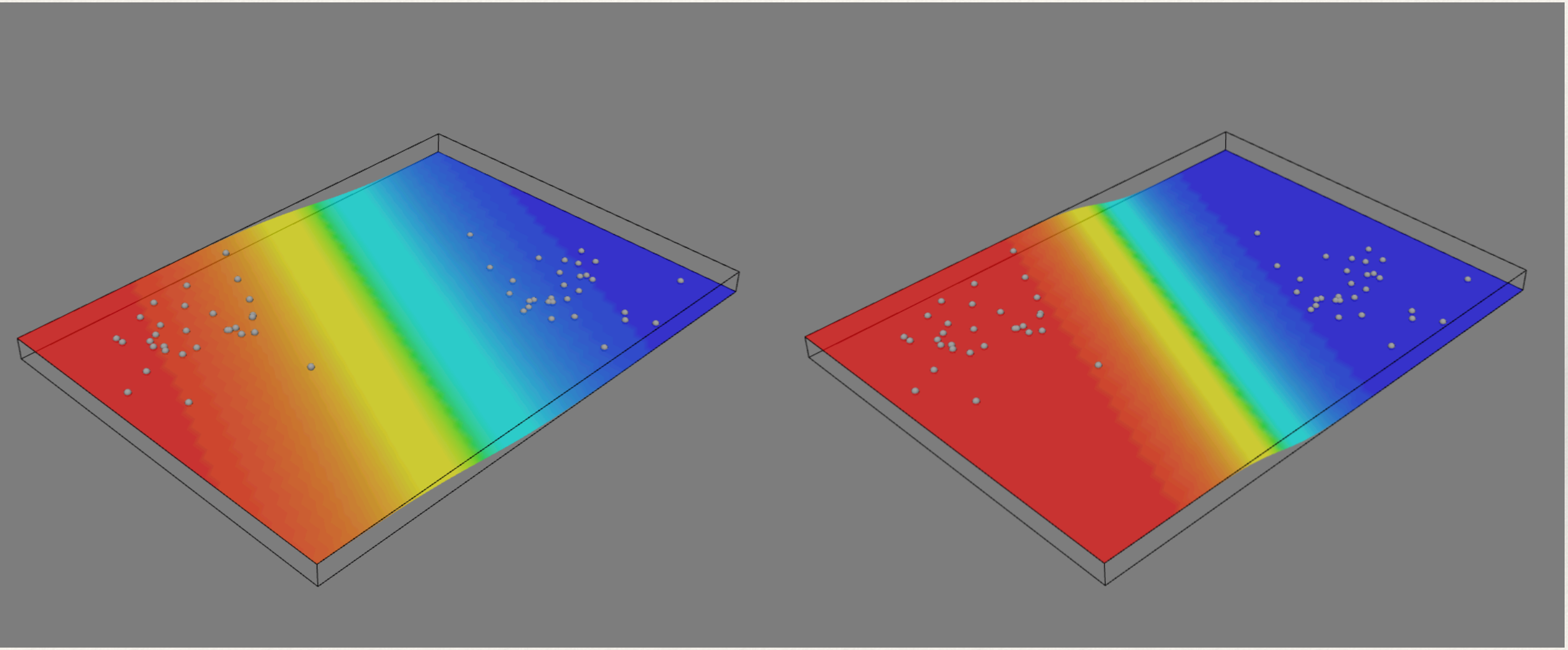
$$\min_{w,c} \frac{1}{2} w^T w + C \sum_{i=1}^{n} \log(\exp(-y_i (X_i^T w + c)) + 1).$$

Similarly, ℓ_1 regularized logistic regression solves the following optimization problem:

$$\min_{w,c} \|w\|_1 + C \sum_{i=1}^{n} \log(\exp(-y_i (X_i^T w + c)) + 1).$$

Cを大きくするとどうなるか？

傾斜が変わる

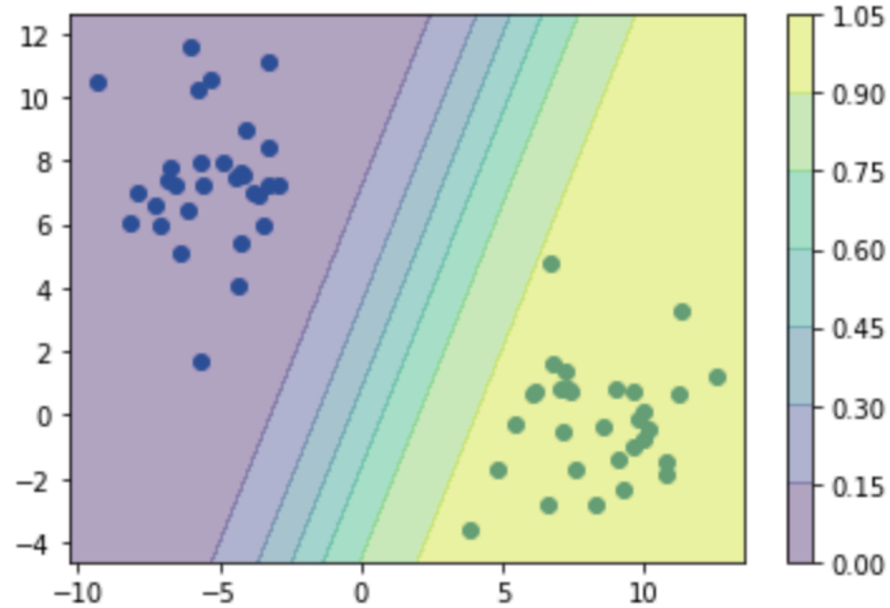


C=0.1　　C=10

2Dで良くね？

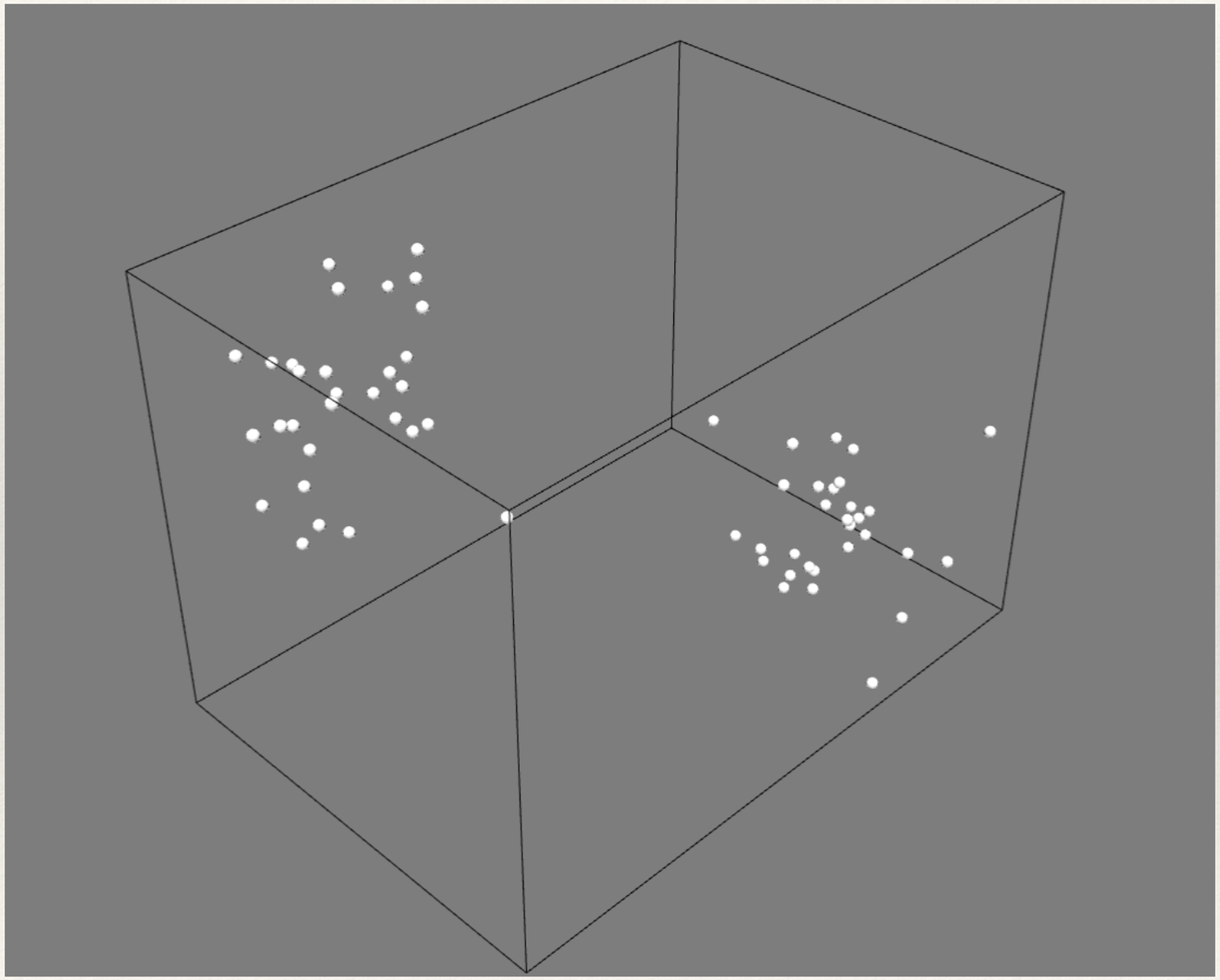
```
fig, ax = plt.subplots()
ax.scatter(X[:, 0], X[:,1])
img = ax.contourf(x0m, x1m, vsig2(x0m, x1m, reg=reg2), alpha=0.4)
fig.colorbar(img)
```

```
<matplotlib.colorbar.Colorbar at 0x14ee7ad30>
```



3次元（説明変数2＋目的変数1）を2次元のコンター図にしている

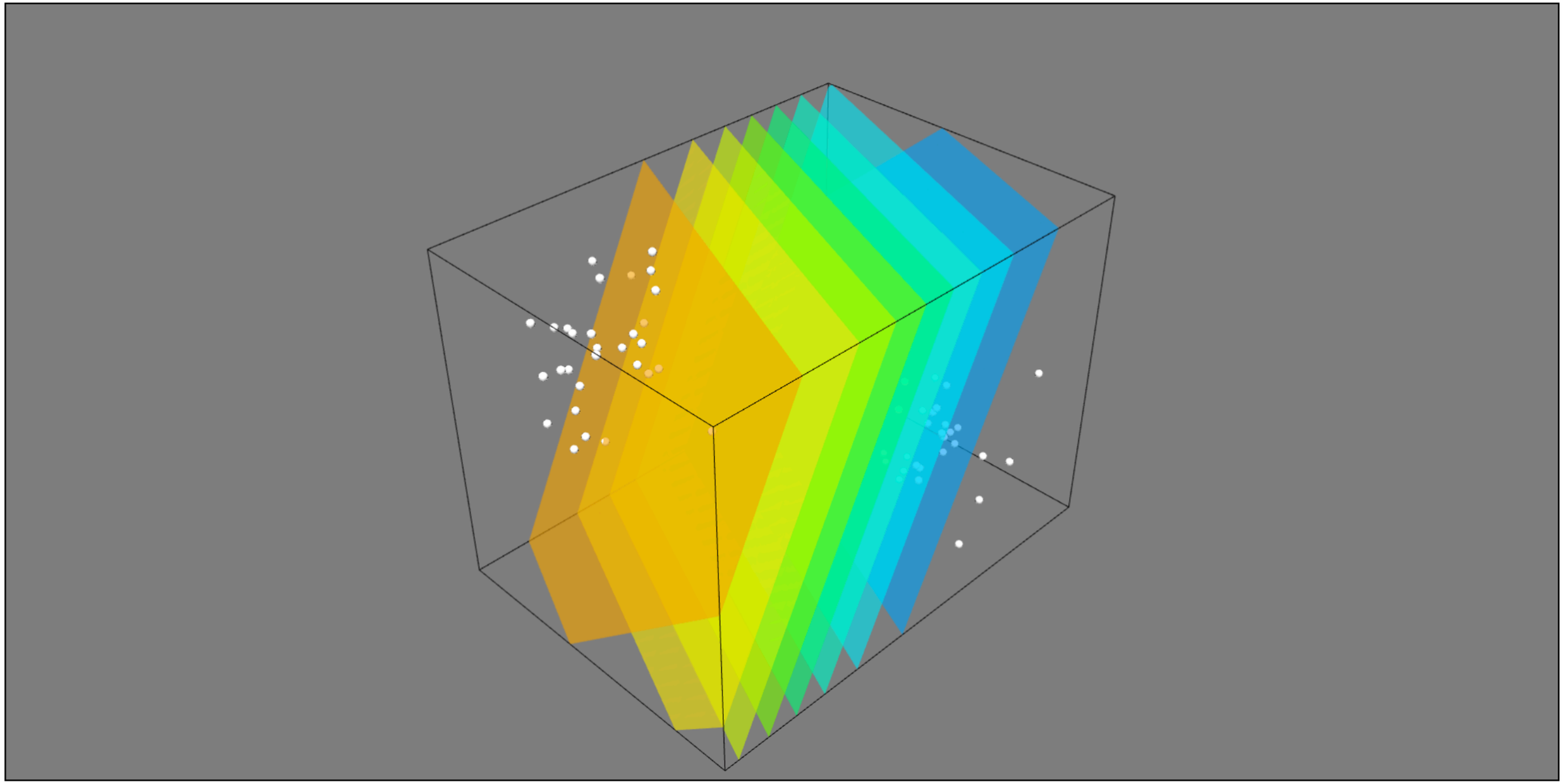
説明変数が3つ



4次元空間で実行されたロジステック回帰の結果を3次元のコンター図にしてみる！

4次元を垣間見た気がする・・・

```
mlab.clf()
_ = mlab.points3d(X[:,0], X[:,1], X[:,2], scale_factor=0.4)
mlab.contour3d(x,y,z, sigmoid3(x,y,z,reg=reg3), opacity=0.6, contours=10)
mlab.outline()
```



まとめ

- 2D, 3Dグラフ描画が数学の理解を助ける
- 科学計算関連の3DプロットはMayaviがおすすめ
- Mayaviはnumpyのよい練習になりそう