

表面科学セミナー2020（実践編）

# 実践! マテリアルインフォマティクス 実際の作業を通して身につける材料データ科学

## 応用編

産業技術総合研究所 機能材料コンピューショナルデザイン研究センター  
主任研究員 安藤 康伸

# 実習で利用する環境

## OS



- Linux distributionのひとつ
- フリーのOS
- Windows 10上でも利用可能
- Windowsのバージョンごとに対応することが大変
- USBからブートすることができる  
(BIOS設定変更が必要)
- Macでも同様にできる

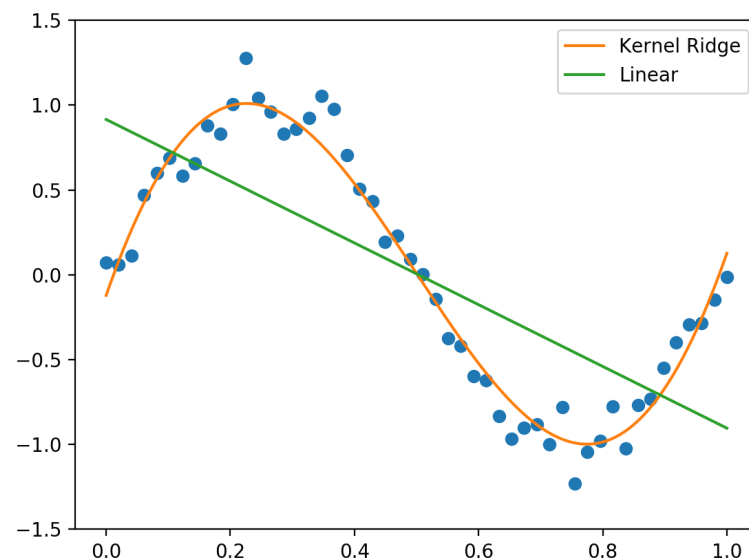
## プログラミング言語



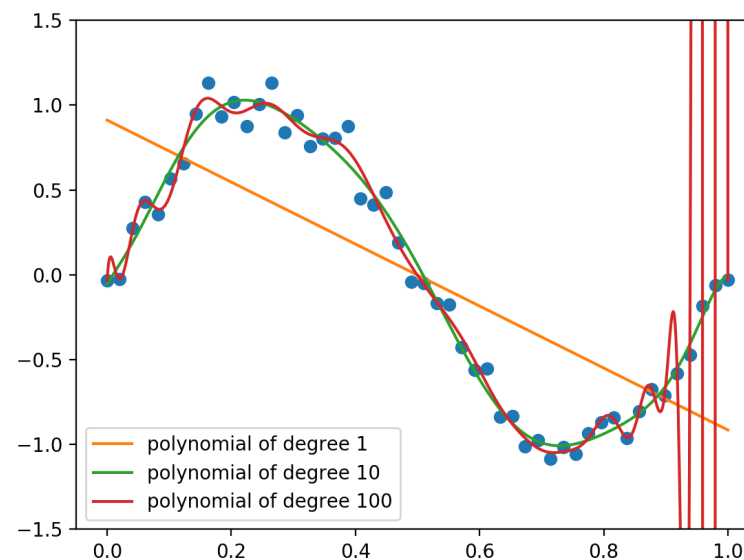
- スクリプト言語のひとつ
- データ入力-処理-可視化がシームレスに実行できる
- 数々のライブラリがフリーで利用可能
- その他、R言語やMatlabもよく使われる
- Anacondaというdistributionが便利
- Jupyter notebookが標準開発環境
- プロットに関してはQiita.com「早く知っておきたかったmatplotlibの基礎知識、あるいは見た目の調整が捗るArtistの話」が詳しい

# テーマ1：予測問題

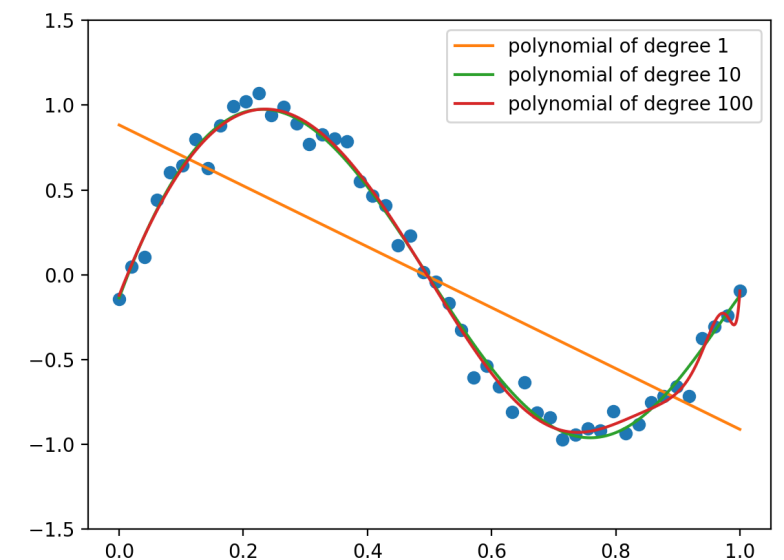
予測問題を取り扱う際に必要な「モデル」と「損失関数」をまず知ろう



カーネル法



最小二乗法

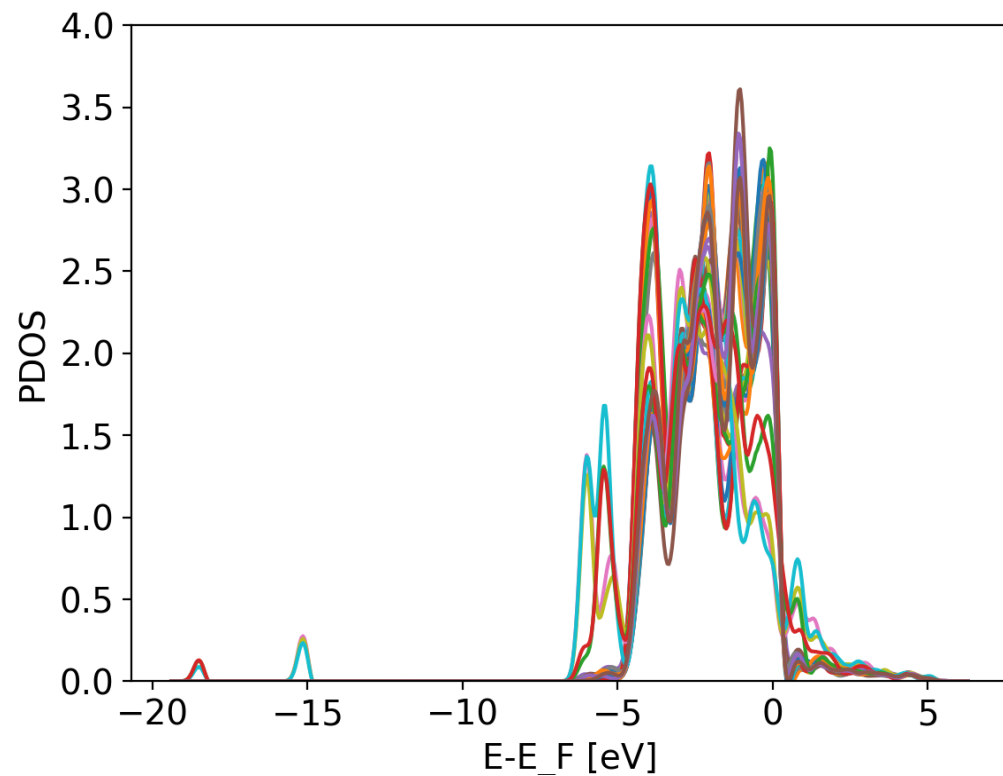


Ridge回帰

## 講義のポイント

- 予測問題で重要な「モデル」と「損失関数」の種類と違いを理解する
- カーネル法と線形最小二乗法の違い
- Ridge回帰やLASSOを使うメリットと交差検証法の使い方について

# テーマ2：分類問題



元データ

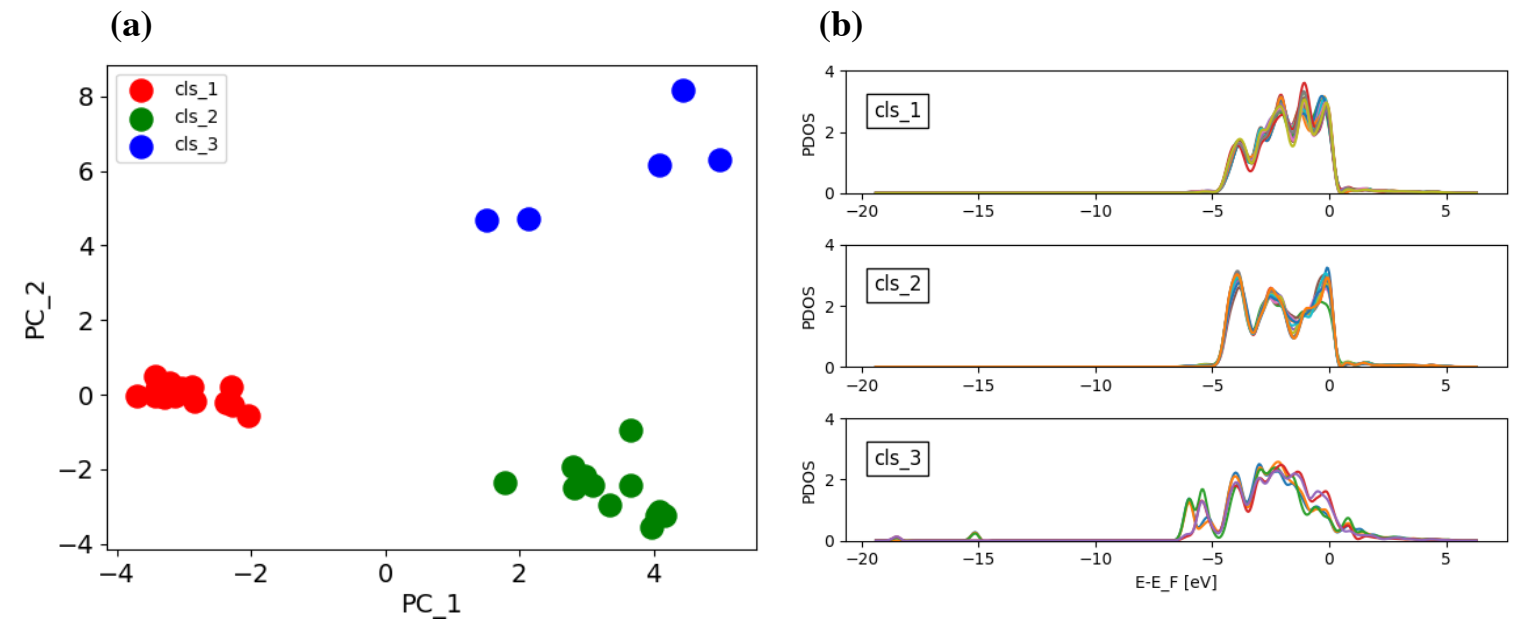
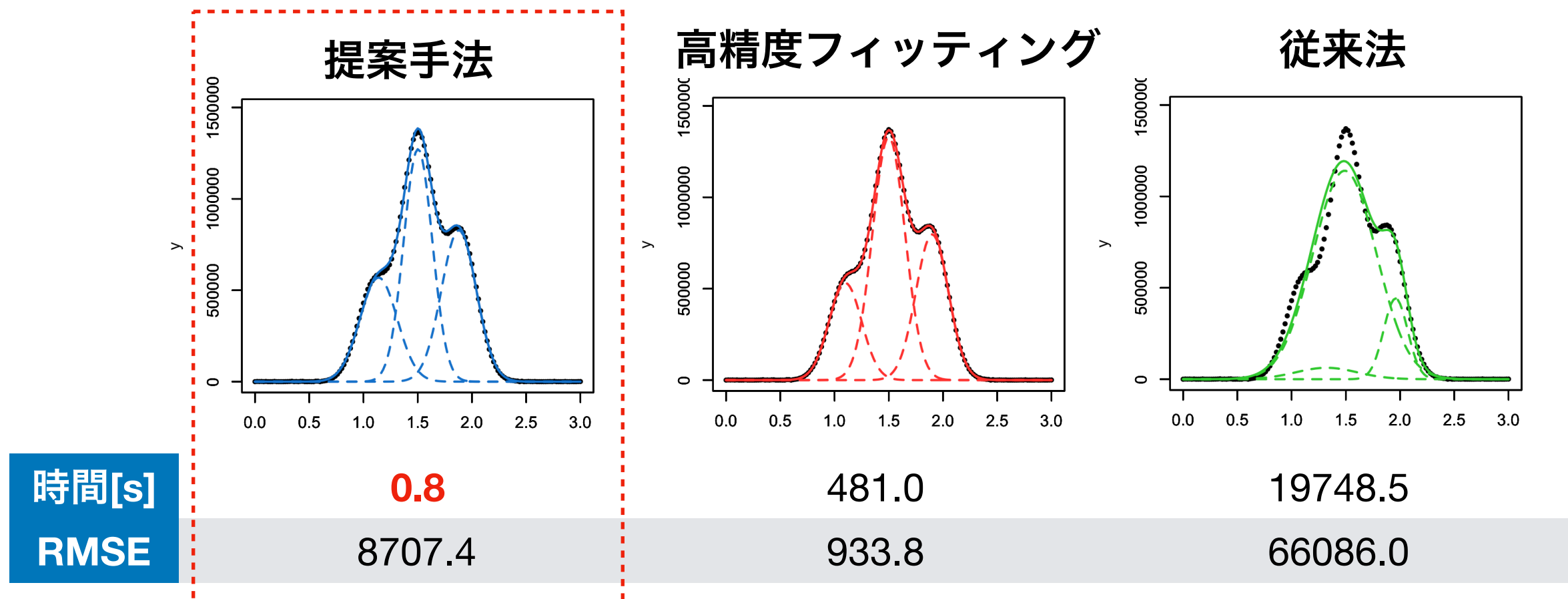


図2-8 (a) K-means法による主成分空間上での分類。(b) K-means法によるPDOSの分類結果

## 講義のポイント

- 様々なスペクトルデータを「理解」「整理する」ための分類問題
- 主成分解析によるスペクトルデータの次元圧縮（低次元化）
- K-means法による機械的な分類アルゴリズムを理解する

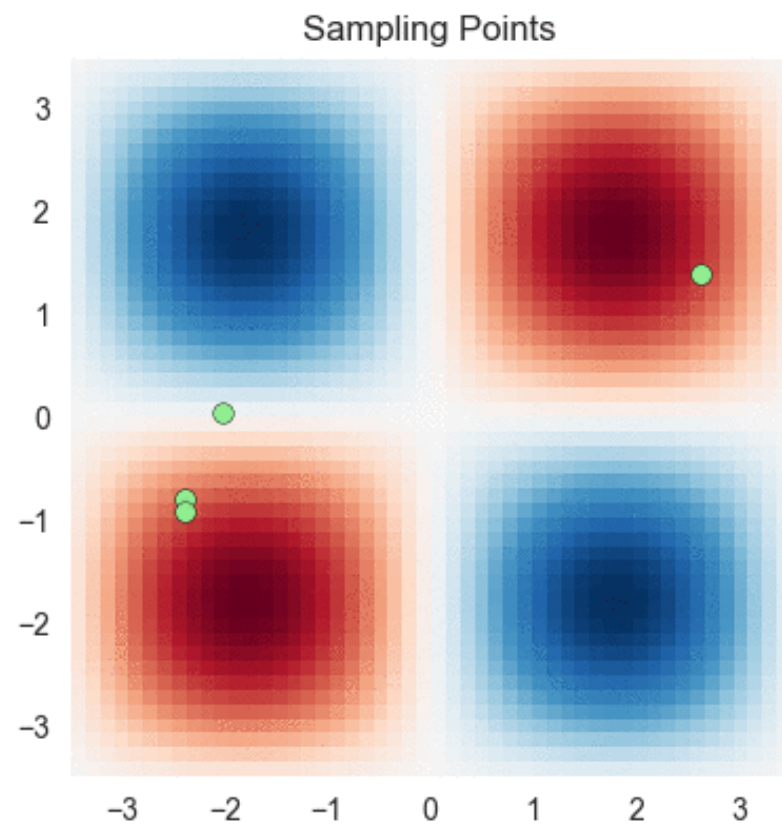
# テーマ3：スペクトル解析



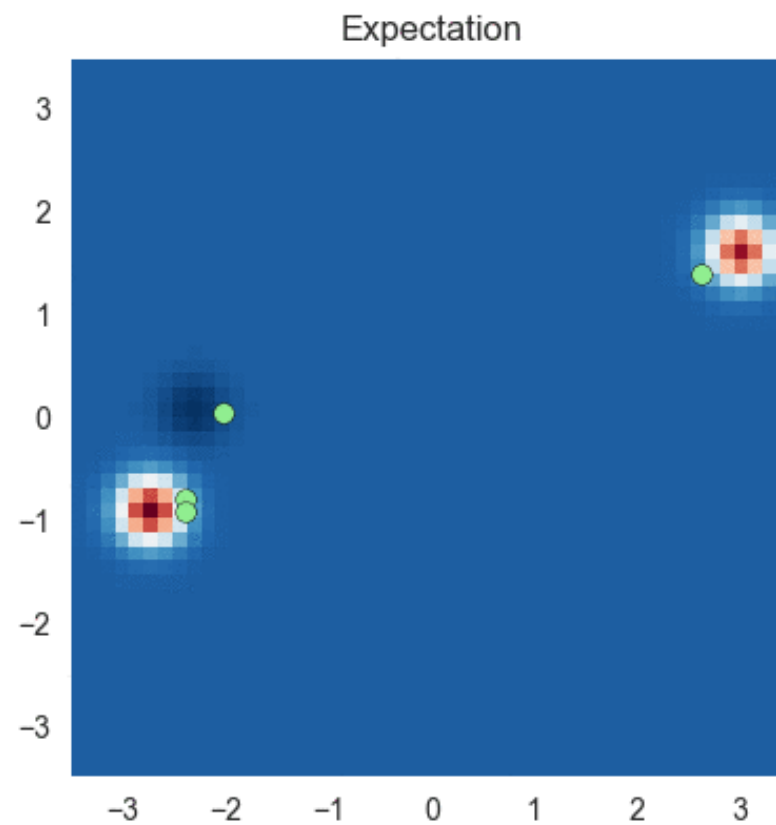
## 講義のポイント

- 非線形最小二乗法を用いない多ピークフィッティング問題
- EMアルゴリズムの理解と実際のスペクトルデータへの適用する際の問題点
- スペクトルに適したEMアルゴリズムによる高速化と自動化を体験してみよう

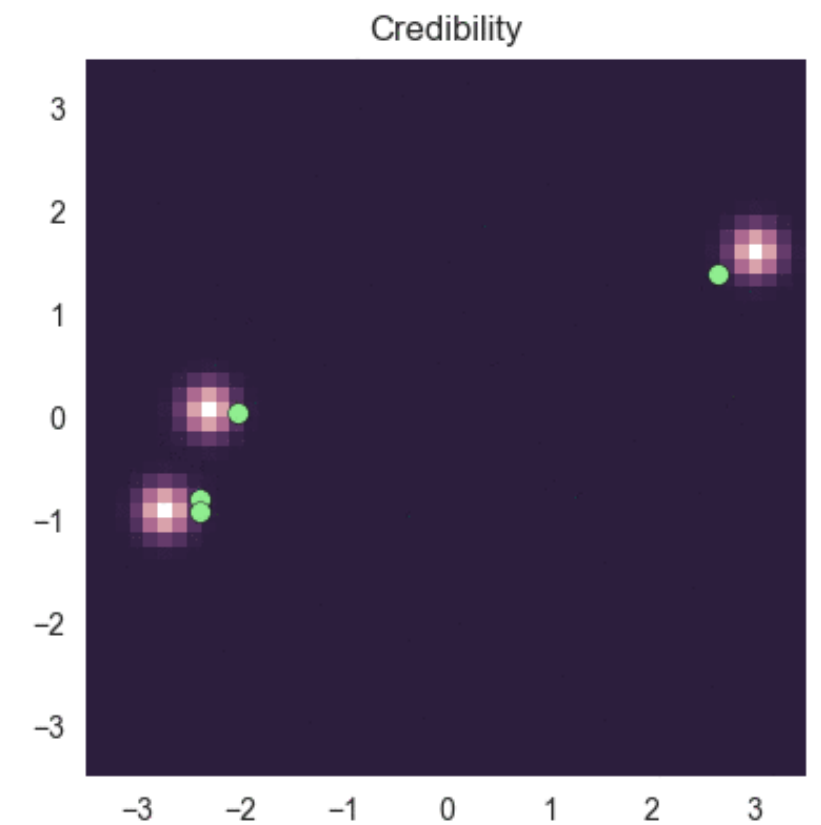
# テーマ4：ベイズ最適化



探索空間



ガウス過程回帰で得られた  
推定モデル



ガウス過程回帰で得られた  
信用区間

## 講義のポイント

- ベイズ最適化で利用される「ガウス過程回帰」「獲得関数」を理解する
- ベイズ最適化でよく語られる「活用」と「探索」を理解する
- 1次元最適化と2次元最適化を試してみる