

Python を用いたケモメトリックス演習講習会 II

○羽子岡 仁志

産業科学研究所 技術室 工作班

1. 派遣先の機関名および派遣期間

派遣先: 自然科学研究機構 分子科学研究所 明大寺キャンパス(愛知県岡崎市明大寺町字西郷中 38 番地)

派遣期間: 令和6年8月2日 10:00~16:00

2. 派遣用務

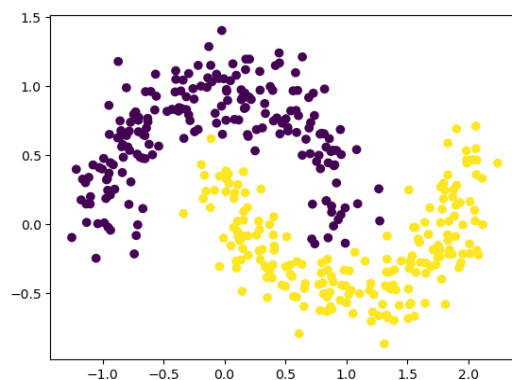
分子科学研究所にて開催された「Python を用いたケモメトリックス演習講習会II」に参加し、Python を活用したデータ解析手法、特にケモメトリックス分野における実践的な技術を習得する。(本講習は 2024 年 3 月に行われた同講習Iの継続内容となる。)

3. 出張の概要および成果

概要

本講習では、Python の主要なライブラリ(numpy, pandas, matplotlib, sklearn)を用い、以下のケモメトリックスに関するデータ解析手法について、講義と演習が行われた。

- **データハンドリング:** データベースの読み込み(pandas)
- **データ可視化:** 各種プロット作成、箱ひげ図作成(matplotlib)
- **データ前処理:** スケーリング、スペクトルのベースライン補正、Savitzky-Golay フィルター処理
- **次元削減・特徴抽出:** 特異値分解 (SVD)
- **クラスタリング:** 階層的クラスタリングの実施とデンドログラムによる可視化
- **回帰分析:**
 - 線形単回帰
 - 主成分回帰 (PCR) におけるクロスバリデーションを用いた最適な成分数(ハイパーパラメーター)の決定
 - PLS 回帰
- **分類分析:** 線形判別分析 (LDA)、混同行列による評価
- **モデリング:** モデルデータを用いた非線形モデリングの概念理解



Python によるデータの可視化

成果

- 上記一連のケモメトリックス解析手法について、Python を用いた具体的な実装コードと手順を学んだ。
- データ読み込みから前処理、モデリング、評価に至るデータ解析のワークフローを実践的に体験し、基本的な操作スキルを習得した。
- 特に、numpy での数値計算、pandas でのデータ操作、matplotlib での可視化、sklearn を用いた機械学習モデリング(回帰、分類、前処理、モデル選択)の基礎的な活用方法について理解を深めた。

4.出張の感想および希望など

感想

演習形式であったため、Python のコーディング自体は手順に沿って進めることができ、内容の 8 割程度は理解できたと感じている。numpy や pandas、matplotlib、sklearn といったライブラリの具体的な使い方を学ぶ良い機会となった。一方で、ケモトリックスという学問分野自体の基礎知識が不足していたため、特に講習の後半で扱われた応用的な分析手法 (PLS 回帰、線形判別分析など) については、手法の理論的な背景や結果の解釈について十分に理解が及ばない点が見られた。

希望・今後の課題

今回の講習で学んだ Python によるデータ解析技術を今後の業務で有効活用するためにも、ケモトリックスの基礎理論について、別途学習を進める必要性を感じた。関連書籍を読む、あるいは基礎的な内容を扱ったセミナー等があれば参加を検討したい。