



chall3.md

📁 課題 3: カスタム機械学習の概要

背景

課題 2 では、学習のためのデータを準備しました。課題 3 では、カスタムの機械学習を開始します。

AdventureWorks は、自社のデータサイエンスチームとあなたのチームが協力して、カスタムの機械学習モデルを構築したいと考えています。機械学習に精通するには時間がかかるので、安定して良質なドキュメントが揃っている人気の既存のフレームワークから始めることをお勧めします。カスタムの機械学習モデルを構築する理由には、以下のようなことがあります。

- さまざまなアルゴリズムを試して、データに最適なものを選ぶ
- 自分の要件に最適なアーキテクチャや構成にモデルをデプロイする
- 既存のサービスの要件（許容データサイズ等）と合わない

この課題では、従来の機械学習の手法を使用します。少なくともプロジェクトの初期段階では、従来の機械学習で十分な効果が得られます。

- 高価な GPU は不要
- 軽量なためエッジでの機械学習の実行に使用可能
- シンプルな API でプロトタイプを高速実行
- 商用システムで使用されている
- アルゴリズムの種類が豊富

前提条件

- チームが Jupyter でコードを共有し共同編集できる環境をセットアップしていること
- 課題 2 で事前処理した登山用品カタログ画像
- Python パッケージ `scikit-learn` の最新版がインストールされていること ([ヒント](#)のセクションを参照)

課題

`scikit-learn` とは、広く普及し定評のある Python の機械学習パッケージです。初心者向けとしては十分な機能を備えており、商用システムで使用されることもあります。シンプルで直観的な API (他のパッケージのベースとしても使用される) を備え、従来の機械学習を学ぶのに最適です。

チーム向けのセットアップ環境で、チームの代表者が次の手順を実施します。

ノンパラメトリックな分類アルゴリズム ([参考資料](#)を参照) で、課題 2 で用意した処理済みの 128x128x3 の画像データで学習を実行して、新しい画像のクラスを予測します。アルゴリズムは、`scikit-learn` ライブラリからノンパラメトリックな手法を選択します。

事前に処理されたデータを学習用とテスト用に分けます。これらはそれぞれ、機械学習モデルの学習とモデルの精度の計算に使用します。通常は学習用データセットの方が大幅に多くなりますが、モデルの品質を最適化するため、十分なテストを実施できるだけのデータセット量を確保しておいてください。

詳細は[参考資料](#)と[ヒント](#)を参照してください。

以下の操作を実行します。

- 課題 2 で処理済みの画像行列データを学習用とテスト用に分けます。
- `scikit-learn` のドキュメントからアルゴリズムを選択します。
- 学習用データでモデルの学習を実行します。
- [こちら](#)の登山用品のクラスをモデルで予測します。
- テスト用データでそれぞれのクラスの分類を確認し、`Confusion Matrix`を使用してモデルを評価します。
- テスト用データで全体の精度を計算します。

完了条件

- Jupyter Notebook でコード セルを 1 つ実行して、上記 URL の登山用品のクラスを正確に予測できる。
- Jupyter Notebook でコード セルを 1 つ実行して、テスト用データで精度のスコアを計算できる。また、このスコアが 80% を超えていること。

参考資料

はじめに

- `scikit-learn` アルゴリズムのドキュメントは[こちらを参照 \(英語\)](#)
- Jupyter の詳細は[こちらを参照 \(英語\)](#)
- パラメトリックアルゴリズムとノンパラメトリックアルゴリズムの違いは[こちらを参照 \(英語\)](#)

機械学習と `scikit-learn`

- `scikit-learn` による機械学習のガイドと用語は[こちらを参照 \(英語\)](#)
- `scikit-learn` による教師あり学習の詳細は[こちらを参照 \(英語\)](#)
- `scikit-learn` の全般的なユーザー ガイドは[こちらを参照 \(英語\)](#)

ヒント

- `pip install <パッケージ名>` を実行してパッケージをインストールします。最新バージョンに更新する場合は、`pip install --upgrade <パッケージ名>` を実行します。
- これは「画像読み込み」機能を作成する際に役立ちます。