

機械学習応用 実践演習 テキスト

荒木雅弘

2018 年 7 月 4 日

第 1 章

Python 入門 (1)

1.1 プログラミング言語 Python

プログラミング言語 Python^{*1}はオブジェクト指向スクリプト言語です。コンパイルが不要なので、短いコードを書いてその実行結果を確認しながらプログラムを組んでゆくことができます。

Python ではひとつのファイルからなるスクリプトをモジュールとよび、そのファイル名から拡張子 (.py) を取り除いたものがモジュール名となります。モジュールは、他のスクリプトで `import` 文を使って読み込むことができ、読み込んだモジュールのクラス・関数・変数は「モジュール名. 識別子」で参照することができます。

また、複数のモジュールをまとめたものをパッケージとよびます。Python は数値計算・グラフ表示などのパッケージが充実しており、科学技術計算のプログラム作成に適しています。特に機械学習に関しては、最新の手法のほとんどが Python で実装されており、深層学習も含め、パッケージ化が進んでいます。

1.2 環境構築

機械学習に関連する Python のパッケージには以下のものがあります。

- `numpy` : 多次元配列を効率よく扱うライブラリ
- `scipy` : 高度な数値計算ライブラリ
- `matplotlib` : グラフ描画のためのライブラリ
- `pandas` : データ解析を支援するライブラリ
- `scikit-learn` : 機械学習のためのライブラリ

また、Python をブラウザ上で対話的に実行でき、Markdown 記法と LaTeX の数式記法で同時にメモを残せる Jupyter notebook を利用することで、小規模なプログラムを対話的に作成できます。これらの環境をすべて事前にインストールした、Anaconda^{*2}をインストールすることをお勧めします。

^{*1} 本章ではバージョン 3.6.5 に基づいて解説してゆきます。

^{*2} <https://www.anaconda.com/download/>

1.3 Jupyter Notebook の使い方

Anaconda をインストール後、メニューから Jupyter notebook を選択すると、ブラウザが起動して、notebook が作成可能になります。適当なフォルダに移動して、New→Python 3 を選択すると、新しい notebook が作成されます。メニュー・アイコンで提供されている機能や、notebook の名前の変更法を確認しておきましょう。

1.3.1 Markdown によるメモ

入力可能なボックスにカーソルを置き、メニューから Markdown を選ぶと、Markdown 記法でメモを挟むことができます。

よく使う Markdown 記法

- 見出し：“#”（レベルは“##”，“###”などで調整）
- 改行：行末で 2 つ以上の空白
- 箇条書き：“*” と空白を先頭に。入れ子は先頭に空白を入れて、記号を“-”に変更。
- 数式：“\$” で囲み、LaTeX の数式記法で書く

1.3.2 Jupyter notebook のキーボードショートカット

以下のキーボードショートカットは便利なのでぜひ憶えておきましょう。

- esc を押してコマンドモード
 - Enter: セルの編集
 - m: マークダウンモード
 - y: コードモード
 - c: セルのコピー
 - v: コピーしたセルのペースト
 - dd: セルの削除
 - Space: スクロールダウン
 - Shift + Space: スクロールアップ
 - h: ショートカット一覧の表示
- セルの編集モードの時
 - Shift+Enter: セルの実行、次のセルへ移動
 - ctrl +Enter: セルの実行のみ

1.4 Python の文法

公式チュートリアル^{*3}を使って、変数・データ構造・制御文・関数の定義・モジュールの扱いについて学んでおきましょう。チュートリアルの要点は Tutorial1.ipynb を参照してください。

1.5 データの準備とグラフ表示

機械学習用データの扱いに慣れておくために、行列の扱い・グラフ表示・CSV ファイルの読み込みについて学んでおきましょう。手順は Tutorial2.ipynb を参照してください。

また、matplotlib パッケージを用いたグラフ表示については、以下のブログが参考になります。

<http://bicycle1885.hatenablog.com/entry/2014/02/14/023734>

実践演習 1-1

パターン認識実践で作成した数字画像の特徴ベクトルを保存した CSV ファイルを読み込み、分散 x に関するヒストグラムおよび分散 x, y の散布図を作成せよ。

^{*3} <http://docs.python.jp/3/tutorial/>