

データサイエンス特論 授業課題 第一回分

機械学習でよく用いられる [IRIS \(あやめ\) のデータ](#) に関して、Python 言語（または Jupyter notebook）を用いて 2 次元の散布図（3 種類のあやめで、異なるカラーとすること）を [Matplotlib](#) でプロットするプログラムを作成せよ。締切は、一週間後【6 月 20 日(月)】の深夜(23:59)までで、[Moodle LMS](#) にアップロードすること。アップロードは、作成した Python プログラム（拡張子.py）または Jupyter notebook（拡張子.ipynb）ならびに、散布図(JPG や PNG のような画像でなくスケーラブルな PDF の図)をまとめて ZIP して提出のこと。Python は 3.6 以上とし、こちらで動作確認ができることを条件とする（一般的でないパッケージ(conda install や pip install で利用できないパッケージ)は使用しないこと）。

【ヒント】IRIS データは、いろいろな Python パッケージに組み込まれているデータです。たとえば、[scikit-learn パッケージ](#)では、load_iris()関数で IRIS データをロードすることができます。その際 from sklearn.datasets import load_iris を宣言しておくといいでしょう。IRIS データ自体は、全体で 150 個のデータで 3 種類のあやめが 50 個ずつ含まれており、属性は 4 種類あります（がくの幅、がくの長さ、花弁の幅、花弁の長さ）。2D プロットでは、たとえば、このうちの適当な 2 つの軸を X, Y にとって描画して結構です。勿論、主成分分析などの次元削減手法で 2 軸（第一主軸と第二主軸）を設定しても結構です。なお、scikit-learn の load_iris は独自のデータ構造のため、[Pandas パッケージ](#)の DataFrame で「データフレーム」に変換してからプロットしても結構です。また、先頭行のハイパーリンクにあるオリジナル IRIS データを読み込んでプロットしても OK です。以下、可視化例です。PDF 生成は Matplotlib の [savefig 関数](#)で PDF 名（名前は自由）を使用ください。

