

ソフトウェア演習Ⅲ〔課題 6: 描画処理〕 青野雅樹

第 6 回の資料にある「あやめ」のデータの散布図にならい、「酵母」(Yeast)データの散布図を作成せよ。データは <https://www.kde.cs.tut.ac.jp/~aono/data/yeast.csv> にある。散布図をプロットするプログラム `kadai6.py`, 及び(2 パターンの)散布図の実行結果(それぞれ、JPEG 画像あるいは PDF)を ZIP にまとめ Moodle にアップロードせよ。〆切は 11 月 22 日(火)(当日、入試で休講ですが)の夜までとする。

[1] Yeast クラスを作成し、`yeasy.csv` データを読み込み、各行のデータをクラスで保存せよ。ここで、Yeast データには、以下の属性データが含まれる。

- 1 コラム: (酵母) シーケンス名 (文字列)
- 2 コラム: MCG (McGeogh 法: 信号列認識法) (実数)
- 3 コラム: GVH (von Heijne 法: 信号列認識法) (実数)
- 4 コラム: ALM (ALOM 膜のスコア) (実数)
- 5 コラム: MIT (真・非ミトコンドリアタンパク質に関するアミノ酸解析スコア) (実数)
- 6 コラム: ERL (“HDEL”部分列の存在) (実数)
- 7 コラム: POX (ペルオキシソームに関する信号) (実数)
- 8 コラム: VAC (液胞または細胞内タンパク質のアミノ酸解析スコア) (実数)
- 9 コラム: NUC (核または非核タンパク質の核局所化信号スコア) (実数)
- 10 コラム: 10 種類の酵母クラス (CYT, NUC, MIT, ME3, ME2, ME1, EXC, VAC, POX, ERL) (3 文字の文字列)

[2] [1]で述べた 10 種類の酵母クラスに異なる色を適当に割り当て、横軸と縦軸を 2 から 9 コラムに該当する実数属性のうち、MCG, GVH, ALM, MIT, NUC の 5 種類のコラム属性のうち、任意の 2 つを選んで、それぞれ、縦軸と横軸にどの属性が選ばれたかわかるように表示し、タイトルに各自の名前と実行した日時(秒まで)を表示し、すべてのデータを用いて散布図を描画せよ。 10 種類の酵母クラスに割りあてた色に関しては凡例 (legend) に表示すること (後述の例参照)。実行は、コマンドラインから、`$ python kadai6.py XXX YYY` のように与えること。ただし、XXX, YYY は、それぞれ X 軸, Y 軸を表す上記の 5 種類の文字列から、異なる 2 つの属性の名前とする。散布図は JPEG または PDF で出力したものとする。また、実行にあたっては、2 つの異なる XXX, YYY のパターンで描画実行した結果を含めること。

【コメントとヒント】

コマンドライン入力 of XXX, YYY が MCG, GVH, ALM, MIT, NUC のいずれかであり、XXX と YYY は異なることのチェックは必ず行い、条件にあてはまらない文字列の場合、あるいは XXX も YYY も同じ文字列の場合は、警告を出して終了するようにしてください。色の選択

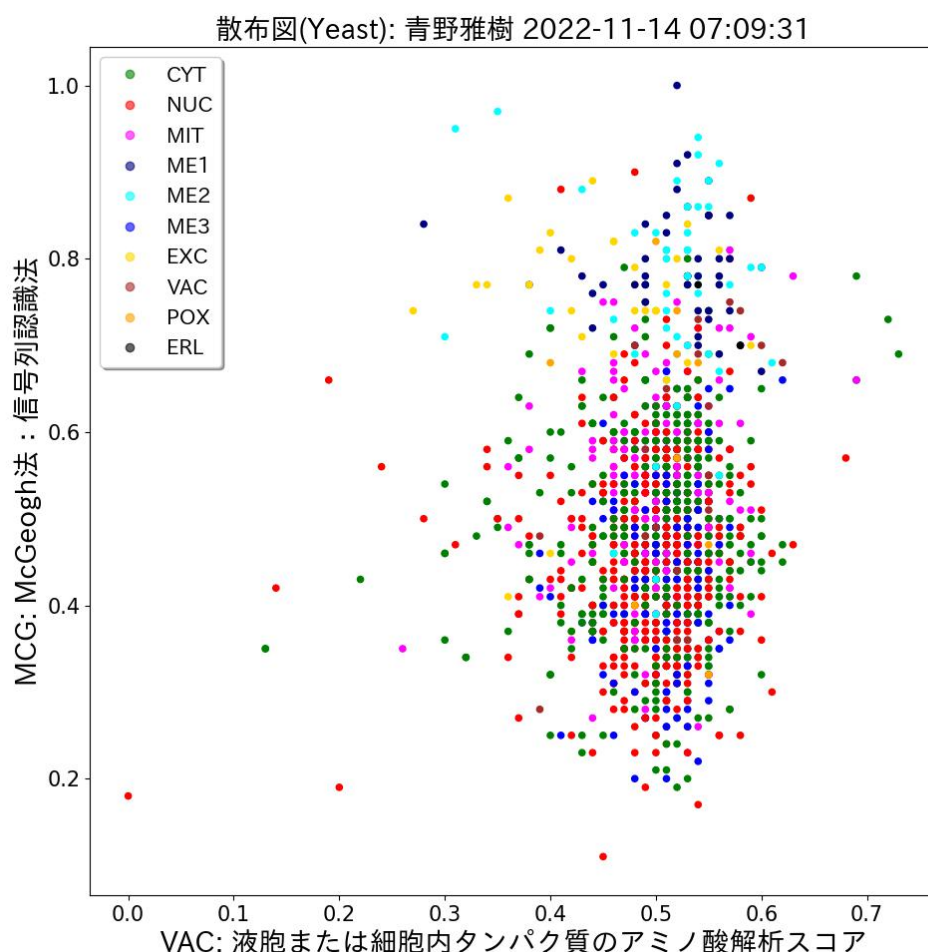
は、たとえば、<https://matplotlib.org/stable/tutorials/colors/colors.html>や
https://matplotlib.org/stable/gallery/color/named_colors.html

を参考に、クラスごとに異なる色を指定してください。

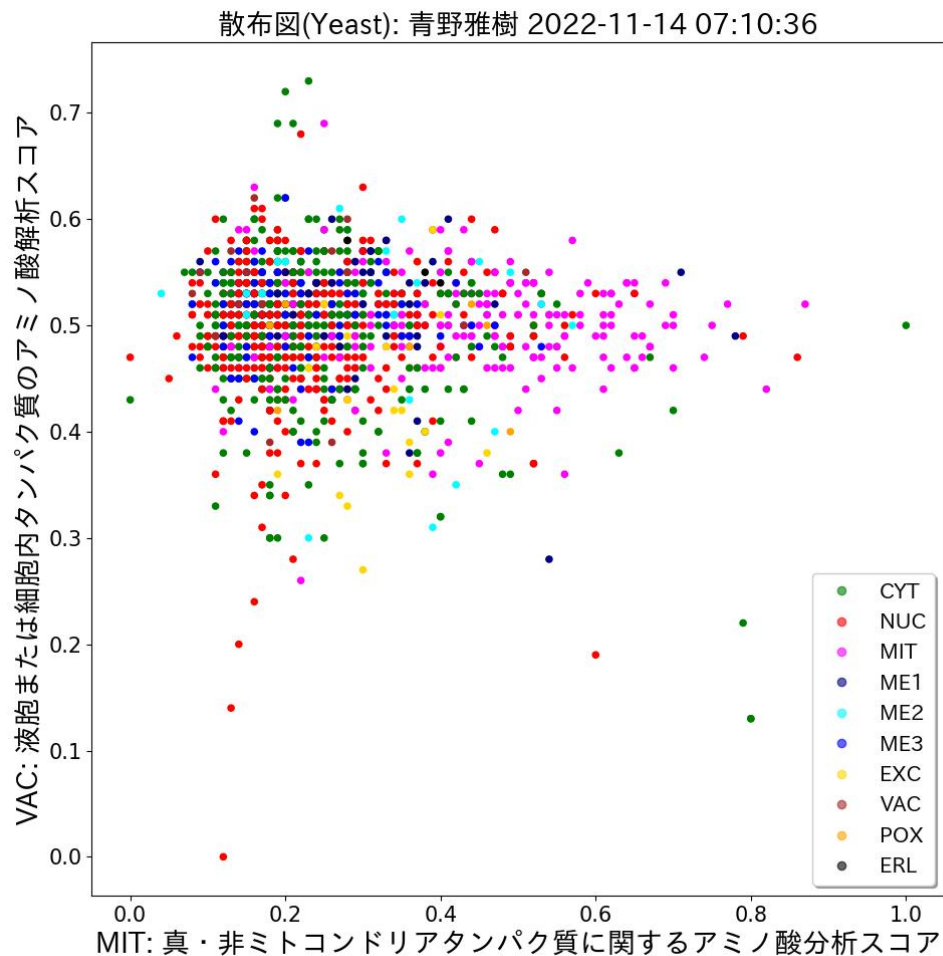
Yeast データの元ネタは、<https://archive.ics.uci.edu/ml/datasets/Yeast> にありますので、個々のデータ属性の意味は、そこにある **Data Set Description** からダウンロードできる `yeast.names` ファイルを参照してください。

実行結果例は以下のようです。ただし、以下は、課題では含んでいない **VAC** 属性を横軸、あるいは縦軸に選んでいます。タイトルに名前と実行時間が含まれていることに注意してください。

```
$ python kadai6.py VAC MCG
```



```
$ python kadai6.py MIT VAC
```



【注】matplotlibのタイトルや軸のラベルに日本語で表示したい場合の例として、IPAex フォントをシステム(Fonts内)に実装するのが1つの方法です。ダウンロードやインストールの詳細は、<https://moji.or.jp/ipafont/ipafontdownload/>を参照ください。

Matplotlib側では、たとえば、以下のように設定しておきます。(Windowsの例)

```
font_path='C:/Windows/Fonts/ipaexg.ttf'  
# プログラム6-1 (資料) の32行目を以下のように変更します。  
font_prop = fm.FontProperties(fname=font_path)
```

なお、日本語フォントは、C:/Windows/Fonts以下でなく、ユーザのローカルな環境にインストールされることがありますので、その場合は、管理者権限で、C:/Windows/Fonts下にコピーしておいてください。

(注) Yeastの元データは<https://archive.ics.uci.edu/ml/datasets/Yeast>です。