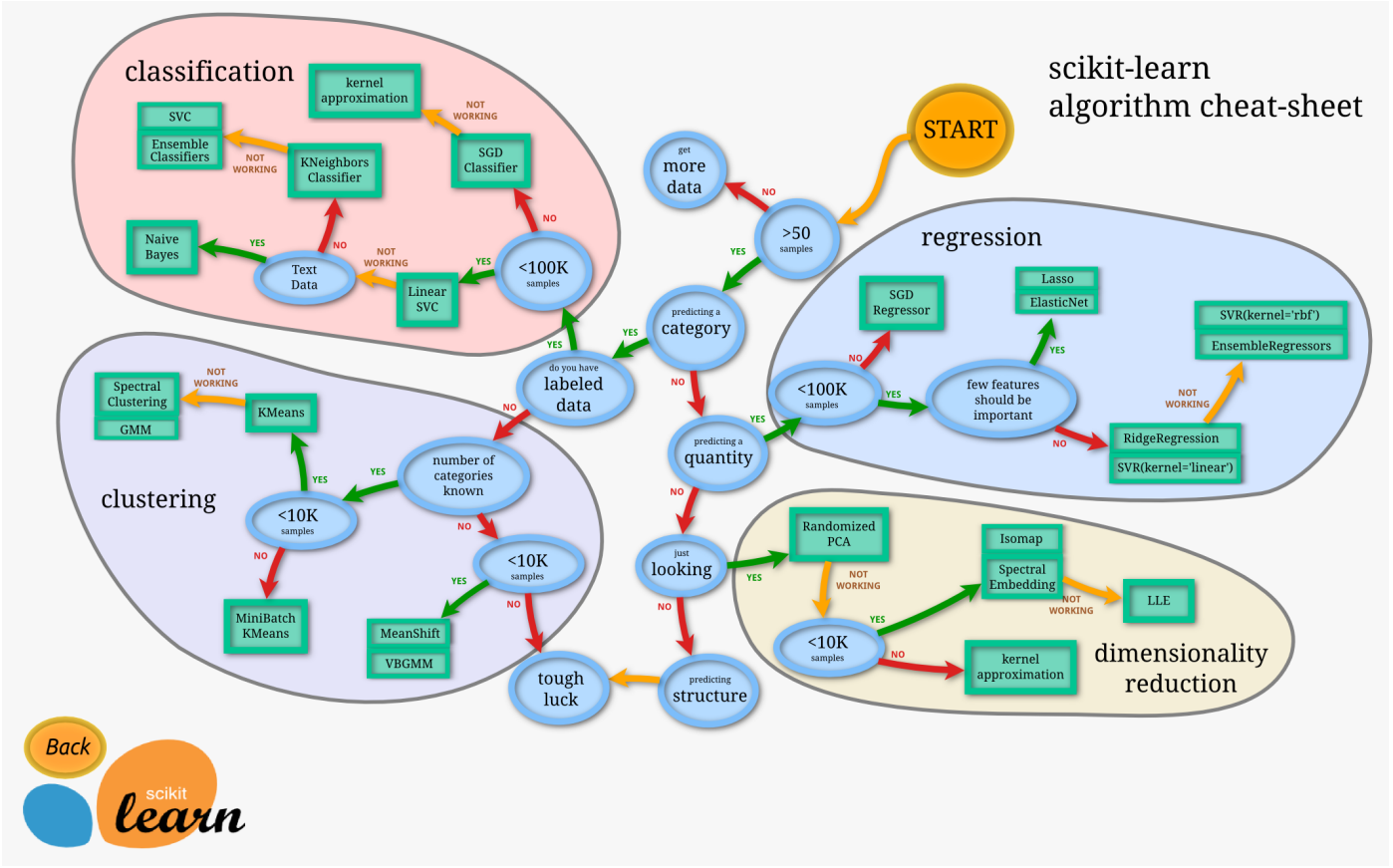
Section 5

- scikit-learnによるコーディング

機械学習ライブラリの紹介 Scikit-learn

- Python の機械学習ライブラリ
- 最新のアルゴリズムが実装されている
- Jyputer Notebook を使ったインタラクティブな開発が可能



- Jupyter Notebook
  - ブラウザで実行可能な Python 環境

jupyter ex02-2

Logout

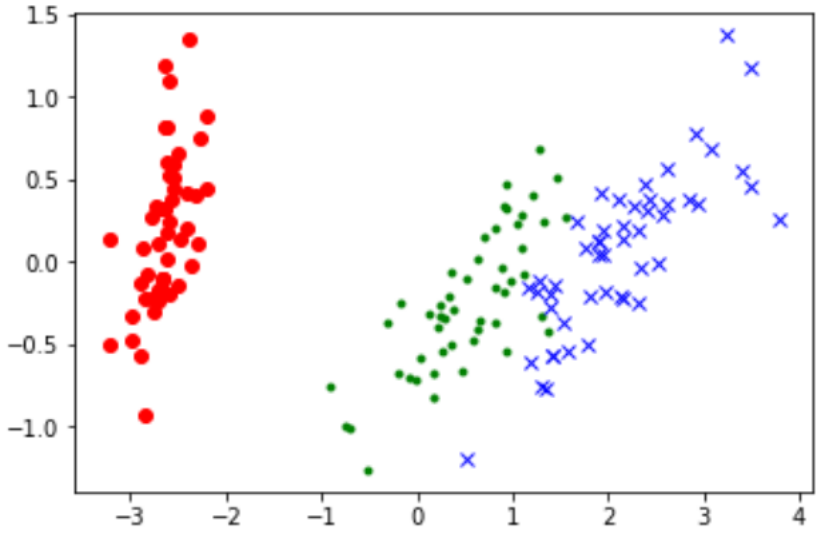
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Markdown

前処理 (主成分分析で2次元に変換し、結果を散布図として表示)

In [3]:
```
pca = PCA(n_components=2)
X2 = pca.fit_transform(X)
plt.plot(X2[y==0,0], X2[y==0,1],"ro")
plt.plot(X2[y==1,0], X2[y==1,1],"g.")
plt.plot(X2[y==2,0], X2[y==2,1],"bx")
```

Out[3]:
```
[<matplotlib.lines.Line2D at 0x1e6c6f50a90>]
```



Python の特長

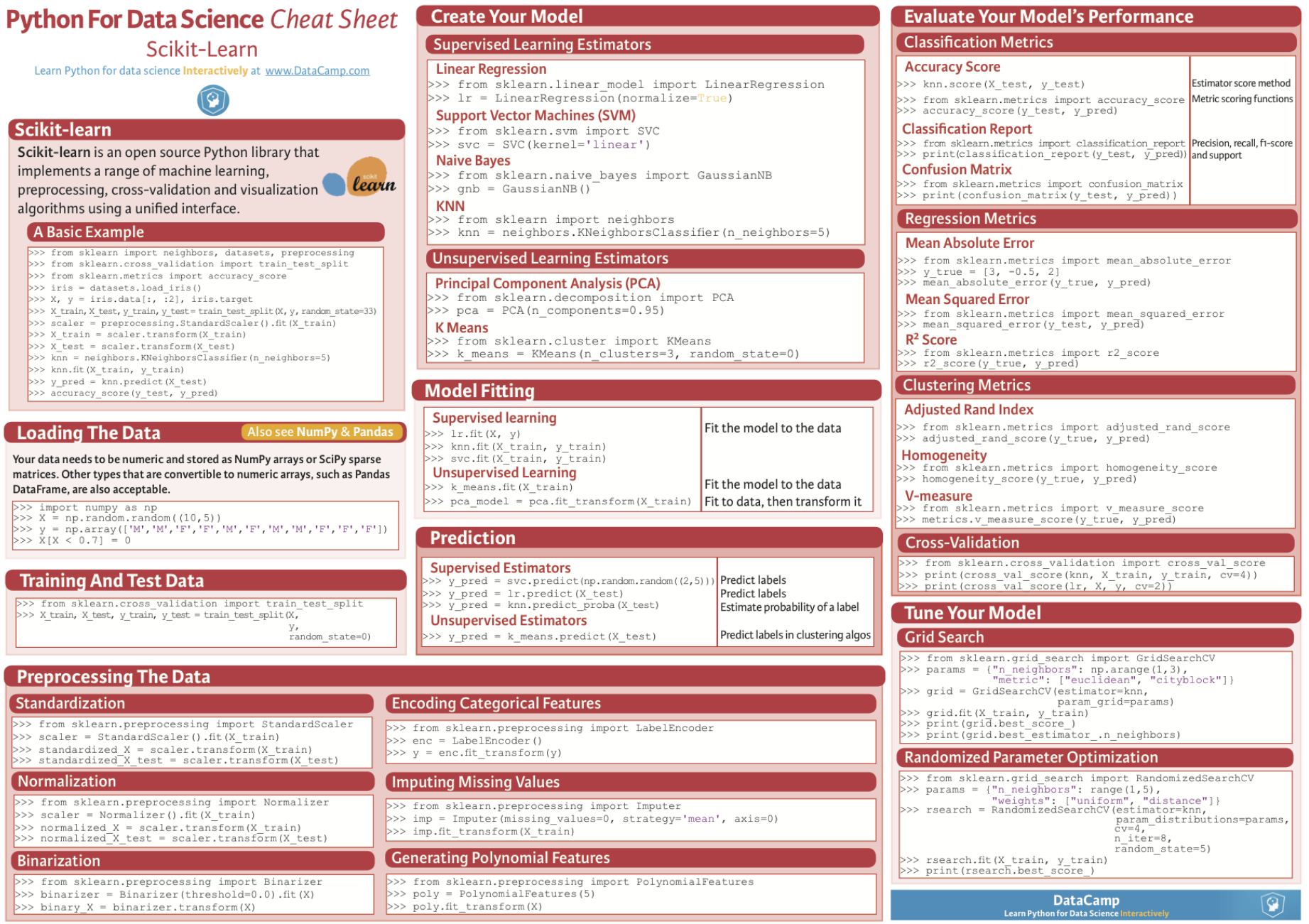
- インタプリタ型言語でコンパイルが不要
- 可変長配列や辞書などの高水準の型を持つ
- 実行文のグループ化をインデントで行う
- 変数の型は代入時に決まるので，宣言不要（動的型付け）

Scikit-learn

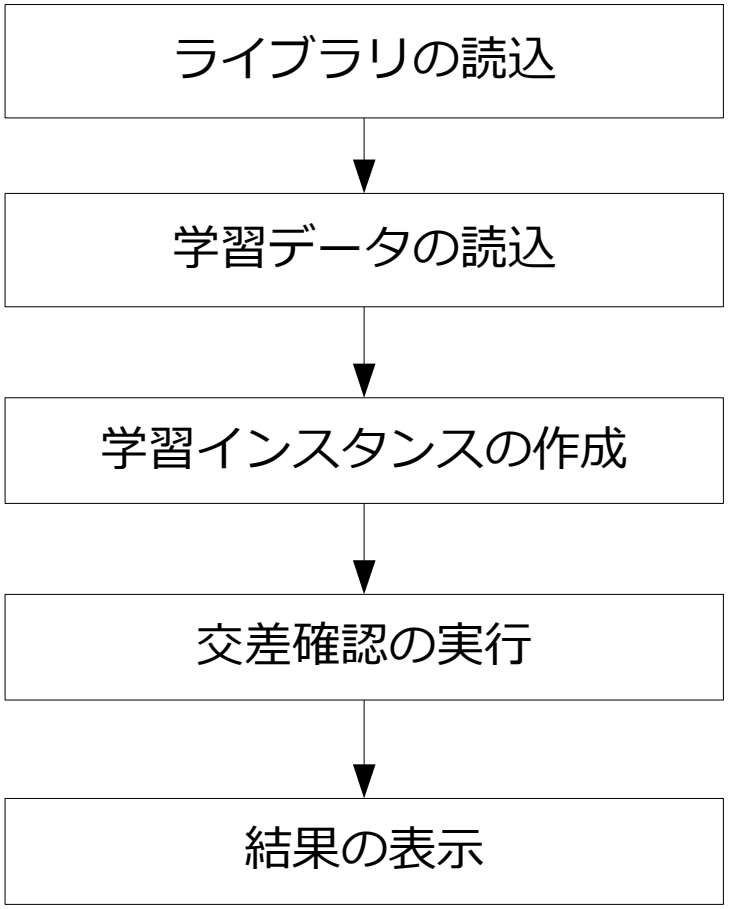
- 識別・回帰・クラスタリング・次元削減などのツールが実装されたパッケージ
- 各アルゴリズムはクラスとして設計されていて，以下の共通した基本仕様からなる
  - コンストラクタ：クラスの初期化
    - 引数はアルゴリズムのパラメータ.
  - fit() メソッド：学習
    - 引数は学習データと正解ラベル．必要に応じてデータに依存したパラメータ.
  - predict() メソッド：予測
    - 学習済みのインスタンスに対して，予測対象のデータを引数として与えると，結果を返す.

- Scikit-learn cheat sheet

https://cdn-images-1.medium.com/max/1500/1*k6_-XfgogqSqjgMJP6_8Iw.png



Scikit-learn のコードパターン



Section5 のまとめ

- 機械学習のライブラリ
  - Scikit-learn: 最も広く使われている Python のライブラリ
- 実行環境
  - Jupyter Notebook によるインタラクティブな環境
- 環境構築
  - Anaconda がお勧め