

統計数学と主成分分析による ビッグ・データの可視化とパターン認識

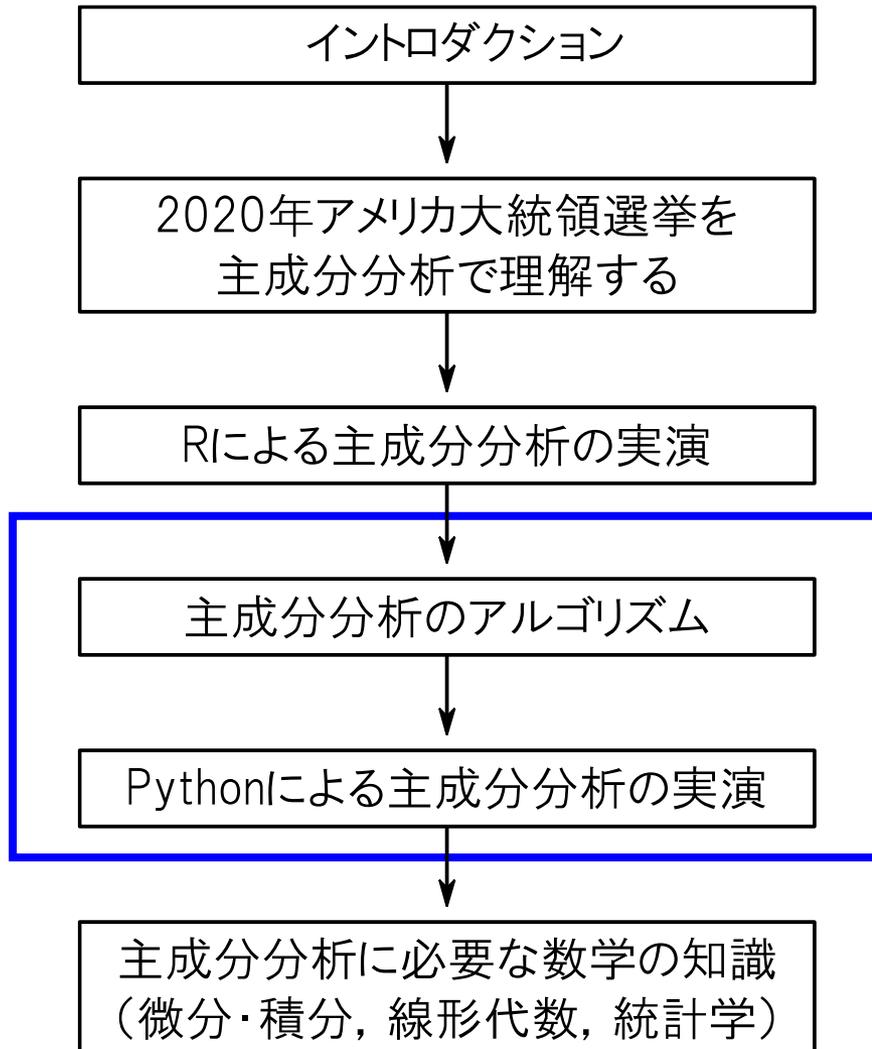
Part 2: 応用編

リニア・テック 別府 伸耕

linear tec : Nobuyasu Beppu

セミナー全体の流れ

Sample



「Part 2 応用編」では
この部分を扱います。

平均, 分散, 標準偏差

Sample

入手したデータの傾向を1つの値で表したものを「代表値」(representative value)という。
代表値の中でよく使われるものとして「平均」, 「分散」, 「標準偏差」がある。

● 平均 (mean)

全データの合計値をデータの個数で割り算したもの。いわゆる「真ん中の値」。

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{k=1}^n x_k$$

● 分散 (variance)

全データの「平均との差の2乗」を合計し, 個数で割り算したもの。いわゆる「データのばらつき」。

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n} = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2$$

● 標準偏差 (standard deviation)

分散の平方根をとったもの。

$$s = \sqrt{\frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2}$$

“numpy”を使って固有値を求める

Sample

```
values, vectors = numpy.linalg.eig( matrix )
```

固有値の配列

固有ベクトルの配列

固有値・固有ベクトルを求める行列

```
vectors = 固有ベクトルは縦方向に並んでいる (列ベクトル)
[ 4.684e-01 -8.259e-02  5.488e-02 -4.005e-01 -6.895e-02 -7.071e-01 -2.457e-01 -9.882e-02  1.872e-01 -2.275e-02 ]
[ -4.684e-01  8.259e-02 -5.488e-02  4.005e-01  6.895e-02 -7.071e-01  2.457e-01  9.882e-02 -1.872e-01  2.275e-02 ]
[ 1.351e-01  5.082e-01 -1.497e-01  1.855e-02  9.427e-02  1.204e-16  4.255e-01 -6.852e-01  1.939e-01 -5.961e-02 ]
[ 3.462e-01 -8.420e-02 -4.596e-01  2.090e-01  6.825e-01  1.050e-15 -1.963e-01 -8.517e-03 -3.359e-01 -3.047e-02 ]
[ 3.501e-01  1.379e-01  8.363e-03  7.483e-01 -2.807e-01  1.036e-15 -2.694e-01  9.239e-02  3.692e-01 -4.696e-02 ]
[ -1.820e-01  4.469e-01 -1.657e-01 -2.093e-01  3.648e-01 -7.989e-16 -7.579e-02  4.735e-01  5.566e-01  1.519e-01 ]
[ -4.890e-02 -3.426e-01 -6.955e-01 -3.372e-02 -3.224e-01 -3.712e-16  1.360e-01 -6.102e-02  1.892e-01  4.832e-01 ]
[ 1.162e-01  4.745e-01  1.452e-01 -2.102e-02 -1.434e-01  9.996e-16 -2.255e-01 -1.856e-04 -4.334e-01  6.933e-01 ]
[ 4.447e-01  1.831e-01 -1.393e-01 -7.830e-02 -2.283e-01 -4.648e-15  5.709e-01  5.231e-01 -2.397e-01 -1.843e-01 ]
[ 2.319e-01 -3.554e-01  4.567e-01  1.544e-01  3.567e-01 -2.374e-15  4.314e-01  2.728e-02  2.285e-01  4.703e-01 ]
```

numpyの行列(array)は“array [行の範囲, 列の範囲]”でアクセスできる。

(例) 2次元配列“vectors”の「1列目」(0列目)を取得するには、“vectors[:, 0]”と記述する。

```
vectors[:,0] =
[ 0.46835794 -0.46835794  0.13514328  0.34616598  0.35013893 -0.18200759
 -0.04889967  0.11615519  0.44474375  0.23191129]
```