

## 主成分分析

$X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  を  $\mathbb{R}^m$  の部分集合とする。 $\mathbf{x}_1, \dots, \mathbf{x}_n$  を縦に並べることで、 $X$  を  $n \times m$  行列と見なす。 $\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$  とおき、 $\bar{X} = \{\mathbf{x}_1 - \bar{\mathbf{x}}, \dots, \mathbf{x}_n - \bar{\mathbf{x}}\}$  とおく。 $\mathbf{x}_1 - \bar{\mathbf{x}}, \dots, \mathbf{x}_n - \bar{\mathbf{x}}$  を縦に並べることで、 $\bar{X}$  を  $n \times m$  行列と見なす。 $d < m$  とする。ある  $m \times d$  行列  $W$  を構成し、 $\bar{X}W$  を計算することによって、 $\bar{X}$  の元を  $\mathbb{R}^d$  に射影することを考える。このとき、 $\bar{X}$  の持つ情報が  $\bar{X}W$  を計算することによってなるべく失われないように  $W$  を構成する必要がある。 $\text{Var}(\bar{X}) = \frac{1}{n} \bar{X}^T \bar{X}$  とおく。 $\text{Var}(\bar{X})$  を  $\bar{X}$  の共分散行列という。 $\text{Var}(\bar{X})$  は  $m$  次正方行列であり、固有値  $\lambda_1, \dots, \lambda_r$  ( $r \leq m, \lambda_1 \geq \dots \geq \lambda_r$ ) を持つ。 $\mathbf{v}_1, \dots, \mathbf{v}_r$  を  $\lambda_1, \dots, \lambda_r$  の固有ベクトルとする。 $\mathbf{v}_i$  を  $\bar{X}$  の第  $i$  主成分という。 $\text{Var}(\bar{X})$  は対称行列であるから、これらは互いに直行する。そこで、ある  $d \leq r$  を選び、 $W$  を  $\mathbf{v}_1^T, \dots, \mathbf{v}_d^T$  を横に並べることで得られる  $m \times d$  行列とする。

例 0.1. (コードは主成分分析.ipynb)

#データを生成

```
import numpy as np
X=np.random.rand(10, 4)
print(X)
>[[0.57474458 0.16751176 0.35264397 0.4309027 ]
 [0.18800478 0.60807862 0.91871515 0.0232249 ]
 [0.6039462  0.9651043  0.57734071 0.51948878]
 [0.64712874 0.17814374 0.14483057 0.03151927]
 [0.62575414 0.87310484 0.64126612 0.54786158]
 [0.49916027 0.259644   0.97128355 0.46407301]
 [0.22934534 0.38941088 0.42845559 0.1687856 ]
 [0.46389123 0.60854082 0.06718675 0.73438477]
 [0.80296074 0.39762404 0.02174463 0.33213051]
 [0.243695   0.38014572 0.63414172 0.57011153]]
```

```
from sklearn.decomposition import PCA
pca = PCA(n_components=2) # 2次元空間へ射影する
pca.fit(X) #Xの平均を0にしてから、Wを構成する

pca.components_ #主成分を表示
>array([[ -0.34658604,  0.29375681,  0.89071272, -0.01469369],
```

```

[ 0.23731118,  0.75216648, -0.14587216,  0.59720207]])

X2d = pca.fit_transform(X) #X を射影する
print(X2d)
>[[-0.23308638 -0.16946345]
 [ 0.54056884 -0.25590155]
 [ 0.18992958  0.45751559]
 [-0.43428412 -0.35248728]
 [ 0.21186806  0.40111113 ]
 [ 0.37071732 -0.18853457]
 [ 0.0231864  -0.25212114]
 [-0.32383051  0.35883747]
 [-0.53787078  0.04705959]
 [ 0.1928016  -0.04601596]]

```

## 参考文献

- [1] Andriy Burkov. (2019). The hundred-page machine learning book.
- [2] Marc Peter Deisenroth., A. Aldo Faisal., Cheng Soon Ong. (2020). Mathematics for machine learning. Cambridge University Press.
- [3] Aurélien Géron. (2019). Hands-on machine learning with Scikit-Learn, Keras & TensorFlow. 2nd Edition. Oreilly.
- [4] 小縣信也., 斎藤翔汰., 溝口聡., 若杉一幸. (2021). ディープラーニング E 資格エンジニア問題集 インプレス.
- [5] Sebastian Raschka., Vahid Mirjalili. (2019). Python machine learning. Third Edition. Packt.