The Engineering World #DataScience 18 & 19

May 31, 2018

AKKAL BAHADUR BIST
DATA SCIENTIST AT
KATHMANDU INSTITUTE OF APPLIED SCIENCES (KIAS)
Center for Conservation Biology (CCB)

1 FACTOR ANALYSIS

1.0.1 Explanatory Factor Analysis

```
In [1]: import numpy as np
        import pandas as pd
        import sklearn
        from sklearn.decomposition import FactorAnalysis
        from sklearn import datasets
```

1.0.2 Factor analysis on iris data sets

```
In [2]: iris = datasets.load_iris()
        X = iris.data
        variable_names =iris.feature_names
        X[0:10,]
```

```
Out[2]: array([[5.1, 3.5, 1.4, 0.2],
               [4.9, 3. , 1.4, 0.2],
               [4.7, 3.2, 1.3, 0.2],
               [4.6, 3.1, 1.5, 0.2],
               [5. , 3.6, 1.4, 0.2],
               [5.4, 3.9, 1.7, 0.4],
               [4.6, 3.4, 1.4, 0.3],
               [5. , 3.4, 1.5, 0.2],
               [4.4, 2.9, 1.4, 0.2],
               [4.9, 3.1, 1.5, 0.1]])
```

```
In [3]: factor = FactorAnalysis().fit(X)
```

```
In [4]: pd.DataFrame(factor.components_,columns = variable_names)
```

Out[4]:

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	0.707227	-0.153147	1.653151	0.701569
1	0.114676	0.159763	-0.045604	-0.014052
2	-0.000000	0.000000	0.000000	0.000000
3	-0.000000	0.000000	0.000000	-0.000000

2 PRINCIPAL COMPONENT ANALYSIS AND SINGULAR VALUE DECOMPOSITION

2.0.1 Principal Component Analysis(PCA)

In [5]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import pylab as plt
import seaborn as sb
from IPython.display import Image
from IPython.core.display import HTML
from pylab import rcParams
import sklearn
from sklearn import decomposition
from sklearn.decomposition import PCA
from sklearn import datasets
```

In [6]:
```
%matplotlib inline
rcParams['figure.figsize'] = 5, 4
sb.set_style('whitegrid')
```

2.0.2 PCA on the iris dataset

In [7]:
```
iris = datasets.load_iris()
X = iris.data
variable_names = iris.feature_names
X[0:10,]
```

Out[7]:
```
array([[5.1, 3.5, 1.4, 0.2],
       [4.9, 3. , 1.4, 0.2],
       [4.7, 3.2, 1.3, 0.2],
       [4.6, 3.1, 1.5, 0.2],
       [5. , 3.6, 1.4, 0.2],
       [5.4, 3.9, 1.7, 0.4],
       [4.6, 3.4, 1.4, 0.3],
       [5. , 3.4, 1.5, 0.2],
       [4.4, 2.9, 1.4, 0.2],
       [4.9, 3.1, 1.5, 0.1]])
```

In [8]:
```
pca = decomposition.PCA()
iris_pca = pca.fit_transform(X)
pca.explained_variance_ratio_
```

```
Out[8]: array([0.92461621, 0.05301557, 0.01718514, 0.00518309])

In [9]: pca.explained_variance_ratio_.sum()

Out[9]: 1.0

In [10]: comps = pd.DataFrame (pca.components_, columns = variable_names)
         comps
```

Out[10]:

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	0.361590	-0.082269	0.856572	0.358844
1	0.656540	0.729712	-0.175767	-0.074706
2	-0.580997	0.596418	0.072524	0.549061
3	0.317255	-0.324094	-0.479719	0.751121

```
In [11]: sb.heatmap(comps)

Out[11]: <matplotlib.axes._subplots.AxesSubplot at 0x7f2fddf3ad30>
```