Chapter 5 - Dimensionality Reduction Methods

Segment 2 - Principal component analysis (PCA)

```python
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 datasets


from sklearn import decomposition
from sklearn.decomposition import PCA


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

▾ PCA on the iris dataset

```python
iris = datasets.load_iris()
X = iris.data
variable_names = iris.feature_names

X[0:10,]
```

```
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]])
```

```python
pca = decomposition.PCA()
iris_pca = pca.fit_transform(X)
```

```python
pca.explained_variance_ratio_
```

```
array([0.92461872, 0.05306648, 0.01710261, 0.00521218])
```

```python
pca.explained_variance_ratio_.sum()
```

```
1.0
```

```python
comps = pd.DataFrame(pca.components_, columns=variable_names)
comps
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	0.361387	-0.084523	0.856671	0.358289
1	0.656589	0.730161	-0.173373	-0.075481
2	-0.582030	0.597911	0.076236	0.545831
3	-0.315487	0.319723	0.479839	-0.753657

```python
sb.heatmap(comps, cmap="Blues", annot=True)
```

`<matplotlib.axes._subplots.AxesSubplot at 0x26077d4bdd8>`