pca

July 2, 2024

1 Principal Component Analysis (PCA)

1.0.1 Imports

[15]:
```
import numpy as np
from sklearn.decomposition import PCA
```

1.0.2 Implementation

[16]:
```
def covariance(x, y):
    x_mean = x.mean()
    y_mean = y.mean()
    n = len(x)

    return sum([(x[i] - x_mean) * (y[i] - y_mean) for i in range(len(x))]) / (n␣
↪- 1)
```

[58]:
```
def pca(data, k):
    # Center data
    data_centered = data - np.mean(data, axis = 0)

    # Compute covariance matrix
    n = data.shape[1]
    covariance_matrix = np.empty([n, n])
    for i in range(n):
        for j in range(n):
            covariance_matrix[i, j] = covariance(data_centered[:, i],␣
↪data_centered[:, j])

    # Perform eigen decomposition of covariance matrix and sort in descending␣
↪order to choose top k principal components
    eigen_vals, eigen_vecs = np.linalg.eig(covariance_matrix)
    sorted_indices = np.argsort(eigen_vals)[::-1]
    eigen_vals = eigen_vals[sorted_indices][:k]
    eigen_vecs = eigen_vecs[:, sorted_indices][:, :k]

    # Project onto new feature space using top k principal components
```

```python
    return np.dot(data_centered, eigen_vecs)
```

1.0.3 Create Test Dataset

[59]:
```python
data = np.random.randn(30, 5)
```

1.0.4 Compare Implementation VS. Library Function

[61]:
```python
# Perform PCA using sklearn.decomposition.PCA
pca_sklearn = PCA(1)
transformed_data_sklearn = pca_sklearn.fit_transform(data)

# Perform PCA using your custom function
transformed_data_custom = pca(data, 1)

# Compare the results
print("Sklearn PCA result:")
print(transformed_data_sklearn)
print("\nCustom PCA result:")
print(transformed_data_custom)
```

```
Sklearn PCA result:
[[ 1.38549345]
 [ 0.59693986]
 [ 0.47308285]
 [-0.37803657]
 [ 0.7151092 ]
 [ 1.82261136]
 [ 1.66109513]
 [-0.07122455]
 [-1.58217617]
 [ 0.14388147]
 [ 0.84908191]
 [-1.77249329]
 [-0.71744546]
 [ 0.01458373]
 [-1.10829776]
 [-0.14099715]
 [-0.87321643]
 [-0.38054763]
 [ 1.24638152]
 [-0.838617  ]
 [ 1.54380084]
 [-1.66461489]
 [-1.08581505]
 [ 0.06909277]
 [ 1.75389033]
```

```
 [ 1.49958998]
 [-0.56610264]
 [-1.08607467]
 [-1.10891903]
 [-0.4000561 ]]

Custom PCA result:
[[ 1.38549345]
 [ 0.59693986]
 [ 0.47308285]
 [-0.37803657]
 [ 0.7151092 ]
 [ 1.82261136]
 [ 1.66109513]
 [-0.07122455]
 [-1.58217617]
 [ 0.14388147]
 [ 0.84908191]
 [-1.77249329]
 [-0.71744546]
 [ 0.01458373]
 [-1.10829776]
 [-0.14099715]
 [-0.87321643]
 [-0.38054763]
 [ 1.24638152]
 [-0.838617  ]
 [ 1.54380084]
 [-1.66461489]
 [-1.08581505]
 [ 0.06909277]
 [ 1.75389033]
 [ 1.49958998]
 [-0.56610264]
 [-1.08607467]
 [-1.10891903]
 [-0.4000561 ]]
```