Assignment_09

Thursday, October 28, 2021 11:17 PM

Homework Assignment 9 [30 points]

STAT430 Unsupervised Learning – Fall 2021

Due: Friday, October 29 on Compass at 11:59pm CST.

- **Answer questions 1-5 in case study 1** (assignment_09_fashion_mnist.ipynb)
- **Answer questions 1-4 in case study 2** (assignment_09_shill_bidding.ipynb)
- **Answer questions 1-2 in case study 3** (in this pdf below)

Problems	Points
Case Study 1	
1.1.1	0.25
1.1.2	0.5
1.2	0.5
2.1	0.5
2.2	0.25
3.1	2
3.2	0.75
4.1	1.25
4.2.1	1.25
4.2.2	1.25
4.2.3	1
4.3.1	0.5
4.3.2	0.5
4.3.3	0.75
4.4	1.25
4.5	1.25
4.6	1
4.7.1	1
4.7.2	1
5.1	0.75
5.2	0.25
5.3	0.5
Case Study 2	
1	0.25
2	0.25
3	0.5
4.1.	1
4.2.1	0.5
4.2.3	0.25
3.1	0.5
3.2	0.5
3.3	0.25
3.3	0.5
4.1	0.5
4.2	0.25
4.3	0.5
5.1	0.5
5.2	0.25
5.3	0.5
6.1	0.5
6.2	0.75
Case Study 3	
1	3
2	0.5

Case Study 3: Artificial Dataset

Question 1

A partially completed, mean-centered, $X_{5\times4}$ matrix is given below. It's covariance matrix $C_{4\times4}$ is also given below. Finally, the four eigenvectors of $C_{4\times4}$ and their corresponding eigenvalues are given below. Note that these eigenvalues are not in order.

We decide to use PCA to project $X_{5\times4}$ onto the matrix $Y_{5\times2}$ (ie. two principal components) also shown below and partially completed. Use the information given to fill in the blanks for the 7 green boxes shown below. Explain your answers or show your work.

C: Covariance Matrix of X

	Attribute 1	Attribute 2	Attribute 3	Attribute 4
Attribute 1	2.7	-0.4	0.45	-1.25
Attribute 2	-0.4	0.8	0.6	-1
Attribute 3	0.45	0.6	1.7	-2.25
Attribute 4	-1.25	-1	-2.25	3.5

Eigenvalues of C

Eigenvalue 1	Eigenvalue 2	Eigenvalue 3	Eigenvalue 4
0.037	2.518	5.769	0.376

Eigenvectors of C

Eigenvector 1	Eigenvector 2	Eigenvector 3	Eigenvector 4
0.259	-0.875	0.362	-0.190
0.475	0.384	0.185	-0.770
0.581	0.245	0.493	0.600
0.608	-0.162	-0.769	0.109

$= (5.769 + 2.518 + 0.376 + 0.037) - (2.7 + 0.8 + 3.5)$

* Variance is preserved
* Eigen Values correspond to the variances

Original Mean-Scaled Dataset X

	Attribute 1	Attribute 2	Attribute 3	Attribute 4
Object 1	1.2	-0.4	-0.8	1
Object 2	?	?	?	?
Object 3	?	?	?	?
Object 4	?	?	?	?
Object 5	?	?	?	?

Principal Components Y

	Principal Component 1	Principal Component 2
Object 1	-0.804008	-1.562107
Object 2	?	?
Object 3	?	?
Object 4	?	?
Object 5	?	?

Covariance Matrix of Y

	Principal Component 1	Principal Component 2
Principal Component 1	5.769	0
Principal Component 2	0	2.518

Case Study 3 Computational verification for reference

```
[191]: cov_mat = np.matrix([[2.7, -0.4, 0.45, -1.25],
                            [-0.4, 0.8, 0.6, -1],
                            [0.45, 0.6, 1.7, -2.25],
                            [-1.25, -1, -2.25, 3.5]])
       cov_mat

[191]: matrix([[ 2.7 , -0.4 ,  0.45, -1.25],
               [-0.4 ,  0.8 ,  0.6 , -1.  ],
               [ 0.45,  0.6 ,  1.7 , -2.25],
               [-1.25, -1.  , -2.25,  3.5 ]])

[200]: eigen_values, eigen_vectors = np.linalg.eig(cov_mat)
       eigen_values, eigen_vectors.T

[200]: (array([5.76899073, 2.51809038, 0.03658336, 0.37633552]),
        matrix([[ 0.36152747,  0.18526195,  0.49280215, -0.7694946 ],
                [-0.87547055,  0.38369222,  0.24517909, -0.16192225],
                [ 0.25855183,  0.4752297 ,  0.58105816,  0.60801242],
                [-0.1897371 , -0.76981493,  0.59950369,  0.10945381]]))

[193]: original_object1 = np.matrix([[1.2, -0.4, -0.8, 1]])

[202]: np.dot(original_object1, eigen_vectors.T[0].T)

[202]: matrix([[-0.80400814]])

[203]: np.dot(original_object1, eigen_vectors.T[1].T)

[203]: matrix([[-1.56210706]])
```

Question 2

What percent of total original attribute variability is preserved in the two principal components above?

$= \dfrac{5.769 + 2.518}{5.769 + 2.518 + 0.376 + 0.037}$

$= 0.9525$ or 95.25%