Quiz 8 Solutions

CSE 4/574

Fall, 2024

Question 1

In this problem you will use Principal Components Analysis (PCA) to analyze a simple data set.
Consider the following data set, X:

Name	Feature 1	Feature 2	Feature 3
A	7	4	3
B	4	1	8
C	6	3	5
D	8	6	1
E	8	5	7
F	7	2	9
G	5	3	3
H	9	5	8
I	7	4	5
J	8	2	2

The mean of the data is [6.9, 3.5, 5.1]

The covariance matrix of centered X is given by:

2.32	1.61	−0.43
1.61	2.50	−1.28
−0.43	−1.28	7.88

The eigen vectors (each column is a eigen vector) and the corresponding eigen values of the covariance matrix are given by:

−0.70	0.70	−0.14
0.71	0.66	−0.25
0.08	0.27	0.96

and

0.75
3.68
8.27

respectively.

Apply PCA to the centered version of the data set showed above, using only the top principal component to embed data into a 1-D latent space. Answer the following questions about the location of the data points in the latent space (rounded the second decimal):

Correct Choice

E: 1.29, F: 4.11
Choice Explanation:
The eigenvector corresponding to the largest eigenvalue $(\lambda = 8.27) is$:
$[-0.14$
$, [-0.25], [0.96]]$

Project Data onto the Top Principal Component:
For E:
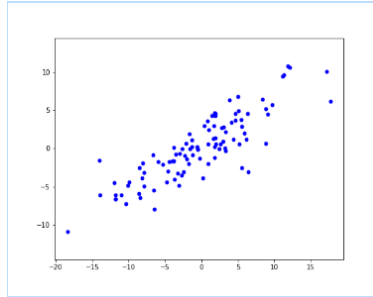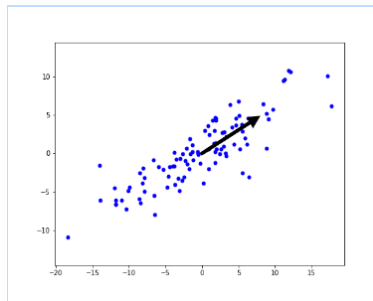$z_E = 1.1(-0.14) + 1.5(-0.25) + 1.9(0.96) = 1.29$
For F:
$z_F = 1.1(-0.14) + (-1.5)(-0.25) + 3.9(0.96) = 4.11$

Question 2

In the following plots, a set of data points are given. Please identify the first principal component of the set of points.

Correct Choice

Explanation: The first principal component is the direction in space that has the most variance in a set of data points.