CS484 - Introduction to Machine Learning

Assignment - 4

Problem 1 :-

1) Given,

we have 2D data points

The first principal component is $(0.694, 0.720)$

Here, we need to find the representation for data #1 and data #8 in the first principal space

The formula would be dot product with principal component.

$$z_n = x_n \cdot v_1 + y_n \cdot v_2$$

data #1:

$$z_1 = (5.51 \times 0.694) + (5.35 \times 0.720)$$
$$= 3.82394 + 3.852$$
$$= 7.67594$$

data #8:

$$z_8 = (-6.08 \times 0.694) + (-5.22 \times 0.720)$$
$$= (-4.21952) + (-3.7584)$$
$$= -7.97792$$

2) To reconstruct the data in the raw space using the first principal component we need to multiply representation from the first part with the first principal component vector.

data #1

$$x = 7.67594 \times 0.694 = 5.32710236$$
$$y = 7.67594 \times 0.720 = 5.5266768$$

data #8

$$x = -7.97792 \times 0.694 = -5.5366 7648$$
$$y = -7.97792 \times 0.720 = -5.7441024$$

3) The second principal component is perpendicular to the first principal component. To find the second principal component we use the fact that its dot product with the first principal component must be zero:

data #1:
$$= (5.51 \times -0.720) + (5.35 \times 0.694)$$
$$= -3.9672 + 3.7129$$
$$= -0.2543$$

data #8:
$$= (-6.08 \times -0.720) + (-5.22 \times 0.694)$$
$$= 4.3776 + (-3.62268)$$
$$= 0.75492$$

4) when using both principal components in 2D space, there is no reconstruction error. This is because we are using all the available dimensions to represent the data.

So, Reconstruction error = 0.

Problem 2 to 5:-

These problems are attached in separate pdf.