HW 4

Bayard Walsh

April 2023

1 Introduction

see code at the end

2 The SVD

2.1

First, we find a linearly independent basis that span the three points. We can remove x_2 from the basis because of the following calculation:
As we have

$$x_1 = \begin{bmatrix} \sqrt{2} \\ 0 \end{bmatrix}, x_3 = \begin{bmatrix} 0 \\ \sqrt{7} \end{bmatrix}$$

Consider

$$\frac{3}{\sqrt{2}}\sqrt{2} \cdot x_1 + \frac{2}{\sqrt{7}}\sqrt{7} \cdot x_3 = \begin{bmatrix} 3 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix} = x_2$$

Therefore we can represent x_2 as a non-zero linear combination of x_1 and x_3.

Next we consider if x_1 and x_3 are linearly independent. Consider

$$a \cdot \sqrt{7} = 0, b \cdot \sqrt{2} = 0$$

As the only solution for these linear equations is $a = b = 0$, there is no non-zero solution, implying linear independence. Therefore x_1, x_3 form a linearly independent, and therefore orthogonal basis.

From here, we will normalize the vectors via Gram-Schmidt process. First we normalize x_1 to u_1.

$$x_1 = \begin{bmatrix} \sqrt{2} \\ 0 \end{bmatrix}, u_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} \sqrt{2} \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Next, we normalize x_3 with respect to u_1

$$x_3' = x_3 - (u_1 \cdot u_1^T) \cdot x_3 = \begin{bmatrix} 0 \\ \sqrt{7} \end{bmatrix} - (\begin{bmatrix} 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \end{bmatrix}) \cdot \begin{bmatrix} 0 \\ \sqrt{7} \end{bmatrix}$$

$$x_3' = \begin{bmatrix} 0 \\ \sqrt{7} \end{bmatrix} - (\begin{bmatrix} 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ \sqrt{7} \end{bmatrix})$$

$$x_3' = \begin{bmatrix} 0 \\ \sqrt{7} \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \sqrt{7} \end{bmatrix}$$

Now we normalize x_3' to get u_2:

$$x_3' = \begin{bmatrix} 0 \\ \sqrt{7} \end{bmatrix}, u_2 = \frac{1}{\sqrt{7}} \begin{bmatrix} 0 \\ \sqrt{7} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Finally we have our orthonormal basis as:

$$U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

2.2

2.2.1

To consider the projection matrix P in terms of v, we have the following equation

$$P = v(v^T v)^{-1} v^T$$

2.2.2

We will generalize the squared distance formula for any x_i in the following equation:

$$D = \|(P \cdot x_i - x_i)\|_2^2$$
$$(P \cdot x_i - x_i)^T (P \cdot x_i - x_i)$$
$$(x_i^T \cdot P^T - x_i^T)(P \cdot x_i - x_i)$$
$$(x_i^T \cdot P^T \cdot P \cdot x_i) - (x_i^T \cdot P^T \cdot x_i) - (x_i^T \cdot P \cdot x_i) + (x_i^T \cdot x_i)$$
$$(x_i^T \cdot P^2 \cdot x_i) - (x_i^T \cdot P^T \cdot x_i) - (x_i^T \cdot P \cdot x_i) + (x_i^T \cdot x_i)$$
$$(x_i^T \cdot P \cdot x_i) - (x_i^T \cdot P^T \cdot x_i) - (x_i^T \cdot P \cdot x_i) + (x_i^T \cdot x_i)$$
$$-(x_i^T \cdot P^T \cdot x_i) + (x_i^T \cdot x_i)$$
$$-(x_i^T \cdot P^T \cdot x_i) + \|(x_i)\|_2^2$$

Therefore distance, $D = \|(x_i)\|_2^2 - (x_i^T \cdot P \cdot x_i)$

2.2.3

To minimize the sum of squared matrices, we can maximize $v^T XX^T v$. First we calculate:

$$XX^T = \begin{bmatrix} \sqrt{2} & 3 & 0 \\ 0 & 2 & \sqrt{7} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 \\ 3 & 2 \\ 0 & \sqrt{7} \end{bmatrix}$$

$$XX^T = \begin{bmatrix} 11 & 6 \\ 6 & 11 \end{bmatrix}$$

Next we find $v^T XX^T v$

$$v^T XX^T v = \begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} 11 & 6 \\ 6 & 11 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$$v^T XX^T v = \begin{bmatrix} 11v_1 + 6v_2 & 6v_1 + 11v_2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$$v^T XX^T v = 11v_1^2 + 6v_2v_1 + 6v_1v_2 + 11v_2^2$$

$$v^T XX^T v = 11v_1^2 + 12v_2v_1 + 11v_2^2$$

Next we substitute for v_1. Because we want a **normalized** vector in v, we have the following:

$$1 = \|(v)\|_2^2 = \sqrt{(v_1^2 + v_2^2)}$$

$$1 = \sqrt{(v_1^2 + v_2^2)}$$

$$1 = (v_1^2 + v_2^2)$$

quick substitution

$$v^T XX^T v = 11(v_1^2 + v_2^2) + 12v_2v_1$$

$$v^T XX^T v = 11 + 12v_2v_1$$

back to v_2

$$v_2^2 = 1 - v_1^2$$

$$v_2 = \sqrt{1 - v_1^2}$$

$$v^T XX^T v = 11 + 12v_1\sqrt{1 - v_1^2}$$

$$v^T XX^T v = 11 + 12\sqrt{v_1^2 - v_1^4}$$

To maximize we first take the derivative:

$$\frac{df(v_1)}{v_1}(11 + 12\sqrt{v_1^2 - v_1^4})$$

$$\frac{6\left(2v_1 - 4v_1^3\right)}{\sqrt{v_1^2 - v_1^4}}$$

Now we set to 0 and find the zeroes

$$0 = \frac{6\left(2v_1 - 4v_1^3\right)}{\sqrt{v_1^2 - v_1^4}}$$

$$0 = 2v_1 - 4v_1^3$$

$$4v_1^3 = 2v_1$$

$$2v_1^2 = 1$$

$$v_1^2 = 1/2$$

$$v_1 = \sqrt{1/2}$$

Notice that $v_1 = \sqrt{1/2}$ also holds for $\sqrt{v_1^2 - v_1^4}$. As we have $v_1 = \sqrt{1/2}$, it follows that $v_2 = \sqrt{1/2}$ by normality of v. Therefore

$$v = \begin{bmatrix} \sqrt{1/2} \\ \sqrt{1/2} \end{bmatrix}$$

Note that v is not unique as this is another optional to achieve the same outcome:

$$v' = \begin{bmatrix} -\sqrt{1/2} \\ -\sqrt{1/2} \end{bmatrix}$$

However, regardless of the sign for our v, because the projection matrix $P = vv^T$, it will be unique.

$$P = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

2.3

Note that finding v_1 and v_2 that minimizes the sum of the squared distance for all data points also gives us u_1, because by the definition of the SVD, u_1 is the vectors that minimize the sum of the squared distance for all data points. Therefore we have

$$u_1 = v = \begin{bmatrix} \sqrt{1/2} \\ \sqrt{1/2} \end{bmatrix}$$

Note that we need u_2 to be orthonormal in relation to u_1, so we have u_2 as follows:

$$u_2 = \begin{bmatrix} -\sqrt{1/2} \\ \sqrt{1/2} \end{bmatrix}$$

Next we compute σ_1 We can find this through the following formula

$$\sigma_1 = ||X^T u_1||_2 = \begin{bmatrix} \sqrt{2} & 0 \\ 3 & 2 \\ 0 & \sqrt{7} \end{bmatrix} \cdot \begin{bmatrix} \sqrt{1/2} \\ \sqrt{1/2} \end{bmatrix}$$

$$||X^T u_1||_2 = \begin{bmatrix} 1 \\ 5\sqrt{1/2} \\ \sqrt{7/2} \end{bmatrix}$$

$$\sigma_1 = ||X^T u_1||_2 = \sqrt{17}$$

We compute σ_2 We can find this through the following formula

$$\sigma_2 = ||X^T u_2||_2 = \begin{bmatrix} \sqrt{2} & 0 \\ 3 & 2 \\ 0 & \sqrt{7} \end{bmatrix} \cdot \begin{bmatrix} -\sqrt{1/2} \\ \sqrt{1/2} \end{bmatrix}$$

$$||X^T u_2||_2 = \begin{bmatrix} -1 \\ -\sqrt{1/2} \\ \sqrt{7/2} \end{bmatrix}$$

$$\sigma_1 = ||X^T u_1||_2 = \sqrt{5}$$

Now that we have the components of U and considering that our sigmas will be diagonal, we can construct U and Σ below

$$U = \begin{bmatrix} \sqrt{1/2} & -\sqrt{1/2} \\ \sqrt{1/2} & \sqrt{1/2} \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \sqrt{17} & 0 & 0 \\ 0 & \sqrt{5} & 0 \end{bmatrix}$$

3

3.1

As we have $X^{n \cdot p} = U \Sigma V^T$, let $r = min(n, p)$. Then we apply our sum across rank one matrices, which sums across the product of *ith* column of v, v_i, the *ith* column of u, u_i, and σ_i for every given i. Therefore we have

$$X = \Sigma_{i=1}^{r} \sigma_i \cdot u_i \cdot v_i$$

3.2

Similar to above, as we have $X^{n \cdot p} = U \Sigma V^T$, let $r = min(n, p)$. However, here we want to calculate X_k for $k < r$, meaning we will now sum across k instead of r. Then we apply our sum across rank one matrices, which sums across the product of *ith* column of v, v_i, the *ith* column of u, u_i, and σ_i for every given i. Therefore we have

$$X_k = \Sigma_{i=1}^{k} \sigma_i \cdot u_i \cdot v_i$$

4

Finding SVD of following matrix

$$X = \begin{bmatrix} -3 & 0 \\ 0 & -1 \end{bmatrix}$$

In order to calculate u_1, we will start by maximizing a for $a^T XX^T a$, where

$$a = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

First we have:

$$XX^T = \begin{bmatrix} 9 & 0 \\ 0 & 1 \end{bmatrix}$$

Next:

$$a^T XX^T a = \begin{bmatrix} a_1 & a_2 \end{bmatrix} \cdot \begin{bmatrix} 9 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

$$a^T XX^T a = 9a_1^2 + a_2^2$$

Next we substitute for a_2. Because we want a **normalized** vector in a, we have the following:

$$1 = \|(a)\|_2^2 = \sqrt{(a_1^2 + a_2^2)}$$

$$1 = \sqrt{(a_1^2 + a_2^2)}$$

$$1 = (a_1^2 + a_2^2)$$

$$a_2^2 = 1 - a_1^2$$

$$a^T XX^T a = 9a_1^2 + 1 - a_1^2$$

$$a^T XX^T a = 8a_1^2 + 1$$

To maximize we first take the derivative:

$$\frac{df(a_1)}{da_1}(8a_1^2 + 1)$$

$$16a_1$$

Therefore, a_1 is strictly increasing, so in order to maximize a_1 with respect to our equation, we pick the largest value in our range. As we want an orthnormal vector, the largest value is 1.
Given $a_1 = 1$, and $1 = \sqrt{(1^2 + a_2^2)}$, $a_2 = 0$. Therefore we have

$$u_1 = \begin{bmatrix} a_1 = 1 \\ a_2 = 0 \end{bmatrix}$$

In order to calculate U_2, we will start by calculating $x_1^{\sim}$ as follows:

$$x_1^{\sim} = x_i - P_{u_1} x_i$$

$$\tilde{x_1} = \begin{bmatrix} -3 & 0 \\ 0 & -1 \end{bmatrix} - u_1 u_1^T \cdot \begin{bmatrix} -3 & 0 \\ 0 & -1 \end{bmatrix}$$

$$\tilde{x_1} = \begin{bmatrix} -3 & 0 \\ 0 & -1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} -3 & 0 \\ 0 & -1 \end{bmatrix}$$

$$\tilde{x_1} = \begin{bmatrix} -3 & 0 \\ 0 & -1 \end{bmatrix} - \begin{bmatrix} -3 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\tilde{x_1} = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$$

Now again, we maximize a for $a^T X X^T a$ (with respect to $\tilde{x_1}$ as X), where

$$a = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

First we have:

$$XX^T = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Next:

$$a^T X X^T a = \begin{bmatrix} a_1 & a_2 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

$$a^T X X^T a = a_2^2$$

To maximize we first take the derivative:

$$\frac{df(a_2)}{da_2}(a_2^2)$$

$$2a_2$$

Therefore, a_2 is strictly increasing, so in order to maximize a_2 with respect to our equation, we pick the largest value in our range. As we want an orthnormal vector, this value is 1.
Given $a_2 = 1$, and $1 = \sqrt{(a_1^2 + 1^2)}$, $a_1 = 0$. Therefore we have

$$u_2 = \begin{bmatrix} a_1 = 0 \\ a_2 = 1 \end{bmatrix}$$

Which also gives us

$$U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Next we compute σ_1,σ_2 We can find this through the following formula

$$\sigma_1 = ||X^T u_1||_2 = \begin{bmatrix} -3 & 0 \\ 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\sigma_1 = 3$$

$$\sigma_2 = ||X^T u_1||_2 = \begin{bmatrix} -3 & 0 \\ 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\sigma_2 = 1$$

As we take the diagonals for the eigenvalues, we have the following

$$\Sigma = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$$

Next we compute v_1,v_2 We can find this through the following formula

$$v_1 = \frac{1}{\sigma_1} \cdot \begin{bmatrix} -3 & 0 \\ 0 & -1 \end{bmatrix}^T \cdot u_1$$

$$v_1 = \frac{1}{3} \cdot \begin{bmatrix} -3 & 0 \\ 0 & -1 \end{bmatrix}^T \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$v_1 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

$$v_2 = \frac{1}{1} \cdot \begin{bmatrix} -3 & 0 \\ 0 & -1 \end{bmatrix}^T \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$v_2 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

Therefore

$$V = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

With this we have our solution,

$$U\Sigma V^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

5

Finding SVD of following matrix

$$X = \begin{bmatrix} -5 & 0 \\ 0 & 1 \end{bmatrix}$$

In order to calculate u_1, we will start by maximizing a for $a^T X X^T a$, where

$$a = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

First we have:

$$XX^T = \begin{bmatrix} 25 & 0 \\ 0 & 1 \end{bmatrix}$$

Next:

$$a^T X X^T a = \begin{bmatrix} a_1 & a_2 \end{bmatrix} \cdot \begin{bmatrix} 25 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

$$a^T X X^T a = 25a_1^2 + a_2^2$$

Next we substitute for a_2. Because we want a **normalized** vector in a, we have the following:

$$1 = \|(a)\|_2^2 = \sqrt{(a_1^2 + a_2^2)}$$

$$1 = \sqrt{(a_1^2 + a_2^2)}$$

$$1 = (a_1^2 + a_2^2)$$

$$a_2^2 = 1 - a_1^2$$

$$a^T X X^T a = 25a_1^2 + 1 - a_1^2$$

$$a^T X X^T a = 24a_1^2 + 1$$

To maximize we first take the derivative:

$$\frac{df(a_1)}{da_1}(24a_1^2 + 1)$$

$$48a_1$$

Therefore, a_1 is strictly increasing, so in order to maximize a_1 with respect to our equation, we pick the largest value in our range. As we want an orthnormal vector, this value is 1.
Given $a_1 = 1$, and $1 = \sqrt{(1^2 + a_2^2)}$, $a_2 = 0$. Therefore we have

$$u_1 = \begin{bmatrix} a_1 = 1 \\ a_2 = 0 \end{bmatrix}$$

In order to calculate U_2, we will start by calculating $x_1^{\sim}$ as follows:

$$x_1^{\sim} = x_i - P_{u_1} x_i$$

$$x_1^{\sim} = \begin{bmatrix} -5 & 0 \\ 0 & 1 \end{bmatrix} - u_1 u_1^T \cdot \begin{bmatrix} -5 & 0 \\ 0 & 1 \end{bmatrix}$$

$$x_1^{\sim} = \begin{bmatrix} -5 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} -5 & 0 \\ 0 & 1 \end{bmatrix}$$

$$x_1^{\sim} = \begin{bmatrix} -5 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} -5 & 0 \\ 0 & 0 \end{bmatrix}$$

$$x_1^{\sim} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Now again, we maximize a for $a^T X X^T a$ (with respect to $\tilde{x_1}$ as X), where

$$a = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

First we have:

$$XX^T = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Next:

$$a^T X X^T a = \begin{bmatrix} a_1 & a_2 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

$$a^T X X^T a = a_2^2$$

To maximize we first take the derivative:

$$\frac{df(a_2)}{da_2}(a_2^2)$$

$$2a_2$$

Therefore, a_2 is strictly increasing, so in order to maximize a_2 with respect to our equation, we pick the largest value in our range. As we want an orthnormal vector, this value is 1.
Given $a_2 = 1$, and $1 = \sqrt{(a_1^2 + 1^2)}$, $a_1 = 0$. Therefore we have

$$u_2 = \begin{bmatrix} a_1 = 0 \\ a_2 = 1 \end{bmatrix}$$

Which also gives us

$$U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Next we compute σ_1,σ_2 We can find this through the following formula

$$\sigma_1 = ||X^T u_1||_2 = \begin{bmatrix} -5 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\sigma_1 = 5$$

$$\sigma_2 = ||X^T u_1||_2 = \begin{bmatrix} -5 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\sigma_2 = 1$$

As we take the diagonals for the eigenvalues, we have the following

$$\Sigma = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}$$

Next we compute v_1,v_2 We can find this through the following formula

$$v_1 = \frac{1}{\sigma_1} \cdot \begin{bmatrix} -5 & 0 \\ 0 & 1 \end{bmatrix}^T \cdot u_1$$

$$v_1 = \frac{1}{5} \cdot \begin{bmatrix} -5 & 0 \\ 0 & 1 \end{bmatrix}^T \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$v_1 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

$$v_2 = \frac{1}{1} \cdot \begin{bmatrix} -5 & 0 \\ 0 & -1 \end{bmatrix}^T \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$v_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Therefore

$$V = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

With this we have our solution,

$$U\Sigma V^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

6

as question is not graded, no answer given

7

7.1

Error rate: 0.11160714285714286

7.2

To find $\hat{w}\lambda$ we have the following computation taken from the λ LSE solution given in class

$$w = (X^T X + \lambda I)^{-1} X^T y$$

$$w = (V\Sigma^T U^T U \Sigma V^T + \lambda I)^{-1} V \Sigma^T U^T y$$

$$w = (V\Sigma^T \Sigma V^T + \lambda I)^{-1} V \Sigma^T U^T y$$

$$w = (V\Sigma^T \Sigma V^T + \lambda V V^T)^{-1} V \Sigma^T U^T y$$

$$w = (V(\Sigma^T \Sigma + \lambda) V^T)^{-1} V \Sigma^T U^T y$$

$$w = V(\Sigma^T \Sigma + \lambda)^{-1} \Sigma^T U^T y$$

The final equation is applied in the code provided
Error rate: 0.04799107142857143

7.3

While these features may result in a slightly smaller error rate **for some, but not all** random inputs, this process is **not helpful** as it adds no new information to the data set, because it is projecting the data in X to more information, meaning that any information in the new columns is represented by the basis of X. Therefore, all three of the new columns are within the basis of X, and therefore linearly dependent, so their information can be represented by just operating on X. The overall impact on the error rate is a slight variant from the initial error rate (which can be positive or negative depending on the random combination)