Homework 2

Question 1

Que (a)

$$\because A^T A = I \tag{1}$$

So, A is a reversible matrix.

$$\Sigma_x = \frac{1}{n}\sum_{i=1}^{6} x_i x_i^T \tag{2}$$

Among this equation,

$$\begin{aligned} x_i x_i^T &= Az_i(Az_i)^T \\ &= Az_i z_i^T A^T \\ &= A(z_i z_i^T)A^T \\ &= AA^T \end{aligned} \tag{3}$$

So that, the equation can be simplified to

$$\begin{aligned} \Sigma_x &= \frac{1}{n}\sum_{i=1}^{6} AA^T \\ &= A\frac{1}{n}\sum_{i=1}^{6} A^T \end{aligned} \tag{4}$$

Then, compute the eigenvalue(s) of Σ_z

$$\begin{aligned} |\lambda E - \Sigma_z| &= \begin{vmatrix} \lambda-3 & 0 & 0 & 0 \\ 0 & \lambda-21 & 0 & 0 \\ 0 & 0 & \lambda-13 & -8 \\ 0 & 0 & -8 & \lambda-13 \end{vmatrix} \\ &= (\lambda-3)(\lambda-21)[(\lambda-13)^2-(-8)^2] \\ &= (\lambda-3)(\lambda-21)(\lambda-21)(\lambda-5) \\ &= 0 \end{aligned} \tag{5}$$

Solve it, we can get that

$$\lambda_1 = 3, \lambda_2 = 21, \lambda_3 = 21, \lambda_4 = 5 \tag{6}$$

However, $A \in \mathbb{R}^{6\times 4}$, so we need to transform Σ_x and λ_x

$$\begin{aligned} \Sigma_x &= AV\Lambda V^T A^T \\ &= (AV)\Lambda(AV)^T \end{aligned} \tag{7}$$

Let's expand (AV) to $\mathbb{R}^6$, so that

$$\lambda_1 = 3, \lambda_2 = 21, \lambda_3 = 21, \lambda_4 = 5, \lambda_5 = 0, \lambda_6 = 0 \tag{8}$$

Then, compute the sum of eigenvalues

$$\begin{aligned} S_\lambda &= Tr(\Sigma_x) \\ &= Tr(A\Sigma_z A^T) \\ &= Tr((AA^T)\Sigma_z) \\ &= Tr(\Sigma_z) \\ &= 3 + 21 + 21 + 5 + 0 + 0 \\ &= 50 \end{aligned} \tag{9}$$

Que (b)

First, compute the estimated value of mean reconstruction error

$$\begin{aligned} error(d) &= \frac{1}{N}\sum_{i=1}^{N} ||x_i - \tilde{x}_i||_2^2 \\ &= \sum_{i=d+1}^{D} \lambda_i \\ &< \frac{50}{4} = 17.5 \end{aligned} \tag{10}$$

It means that

$$\lambda_1 + \lambda_2 + \cdots + \lambda_d \geq 37.5 \tag{11}$$

Meanwhile, sum of the top d values shows that

$$\begin{cases} \max \lambda_1 = 21 < 37.5 \\ \max \lambda_1 + \max \lambda_2 = 21 + 21 = 42 > 37.5 \end{cases} \tag{12}$$

In conclusion, the value of PCA direction must be **not less than** 2

Que (c)

As the solving process in **Q-a** and **Q-b** above, compute the new y

$$x \sim \mathcal{N}(0, \Sigma_x), v \sim \mathcal{N}(0, \Sigma_v) \tag{13}$$

And x is independent of v, so that

$$y \triangleq x + v \in \mathcal{N}(0 + 0, \Sigma_x + \Sigma_v) = \mathcal{N}(0, \Sigma_x + \Sigma_v) \tag{14}$$

Then compute Σ_y

$$\begin{aligned}
\Sigma_y &= \Sigma'_x + \Sigma_z \qquad (15)\\
&= \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 21 & 0 & 0 \\ 0 & 0 & 13 & 8 \\ 0 & 0 & 8 & 13 \end{bmatrix}' + \begin{bmatrix} 5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 5 \end{bmatrix} \\
&= \begin{bmatrix} 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 21 & 0 & 0 & 0 & 0 \\ 0 & 0 & 13 & 8 & 0 & 0 \\ 0 & 0 & 8 & 13 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 5 \end{bmatrix} \\
&= \begin{bmatrix} 8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 26 & 0 & 0 & 0 & 0 \\ 0 & 0 & 18 & 8 & 0 & 0 \\ 0 & 0 & 8 & 18 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 5 \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
|\lambda E - \Sigma_y| &= \begin{vmatrix} \lambda' - 8 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda' - 26 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda' - 18 & -8 & 0 & 0 \\ 0 & 0 & -8 & \lambda' - 18 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda' - 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda' - 5 \end{vmatrix} \\
&= (\lambda' - 8)(\lambda' - 26)[(\lambda' - 18) - (-8)^2](\lambda' - 5)(\lambda' - 5) \\
&= (\lambda' - 8)(\lambda' - 26)(\lambda' - 26) - (\lambda' - 10)(\lambda' - 5)(\lambda' - 5) \\
&= 0
\end{aligned}$$

Solve it, and we get that

$$\lambda'_1 = 8, \lambda'_2 = 26, \lambda'_3 = 26, \lambda'_4 = 10, \lambda'_5 = 5, \lambda'_6 = 5 \qquad (16)$$

Then compute the sum of eigenvalues

$$\begin{aligned}
S_{\lambda'} &= \sum_{i=1}^{6} \lambda'_i \qquad (17)\\
&= 8 + 26 + 26 + 10 + 5 + 5 \\
&= 80
\end{aligned}$$

$$\begin{aligned} error'(d) &= \sum_{i=d+1}^{D} \lambda'_i \\ &< \frac{80}{4} = 20 \end{aligned} \tag{18}$$

$$\lambda'_1 + \lambda'_2 + \cdots + \lambda'_d \geq 60 \tag{19}$$

Meanwhile, sum of the top d values shows that

$$\Big\{ \max \lambda'_1 + \max \lambda'_2 = 26 + 26 = 52 < 60 \max \lambda'_1 + \max \lambda'_2 + \max \lambda'_3 = 26 + 26 + 10 = 62 > 60 \tag{20}$$

In conclusion, the new value of PCA direction must be **not less than** 3

Question 2

We consider two situations of the questions.

1. a>1

First iteration

We choose x_1 and x_4 as fixed points. Then calculate the distance between each pair of fixed point and the unfixed point.

Assume that the fixed point is $p_i = (x_i, y_i)$, unfixed point is $p_j = (x_j, y_j)$ The distance(d_{ij}) between them can be calculated by equation21:

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{21}$$

Group Result: The iteration result is shown in table 1. The group result is follows, G_{ij} is respect for the j group in i iteration.

Table 1: First iteration

	x_1	x_4
x_2	1	$\sqrt{a^2+1}$
x_3	$\sqrt{a^2+1}$	1

$$Result_1 : x_1, x_2 \in G_{11}, x_3, x_4 \in G_{12} \tag{22}$$

Then generate another group of fixed points of a new group, $p_i = (x_i, y_i) \in P$. All fixed points of a iteration belong to the set P. x_i and y_i is calculated by equation 23.X, Y are the coordinates of the points belong to the same group.

$$x_i = \bar{X}, y_i = \bar{Y} \tag{23}$$

The result is $p_1(0, \frac{1}{2})$, $p_2(a, \frac{1}{2})$

Second iteration

Then we choose the p_1 and p_2 as fixed points. Then calculate the distance between fixed point and sample point.

Table 2: Second iteration

	p_1	p_2
x_1	$\frac{1}{2}$	$\sqrt{a^2+\frac{1}{4}}$
x_2	$\frac{1}{2}$	$\sqrt{a^2+\frac{1}{4}}$
x_3	$\sqrt{a^2+\frac{1}{4}}$	$\frac{1}{2}$
x_4	$\sqrt{a^2+\frac{1}{4}}$	$\frac{1}{2}$

Result: According to K-Means, the result is:

$$Result_2 : x_1, x_2 \in G_{21}, x_3, x_4 \in G_{22} \tag{24}$$

Therefore, we can get $Result_1 = Result_2$, in others words, the K-Means algorithm has converged.

2. a<1

We choose x_1 and x_2 as the initial fixed points.

First iteration

First iteration result is shown in table 3

Table 3: First iteration

	x_1	x_2
x_3	$\sqrt{a^2+1}$	a
x_4	a	$\sqrt{a^2+1}$

Group Result

The group result is

$$Result_1 : x_1, x_4 \in G_{11}, x_2, x_3 \in G_{12} \tag{25}$$

Similar to the $a > 1$ condition, we choose $p_1 = (\frac{a}{2}, 1), p_2 = (\frac{a}{2}, 0)$ as fixed points

Second iteration

Apply K-Means to the points, we get the iteration result:

Table 4: Secondly iteration

	p_1	p_2
x_1	$\frac{a}{2}$	$\sqrt{\frac{a^2}{4}+1}$
x_2	$\sqrt{\frac{a^2}{4}+1}$	$\frac{a}{2}$
x_3	$\sqrt{\frac{a^2}{4}+1}$	$\frac{a}{2}$
x_4	$\frac{a}{2}$	$\sqrt{\frac{a^2}{4}+1}$

Group Result

The group result is

$$Result_2 : x_1, x_4 \in G_{21}, x_2, x_3 \in G_{22} \tag{26}$$

Therefore, $Result_1 = Reuslt_2$, in others words, the K-Means algorithm has converged.

3. a=1

If $a = 1$ we can divide the four points into x_1, x_2 and x_3, x_4 or we can also divide them into x_1, x_3 and x_2, x_4

Question 3

It has several options.

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i \triangleq \bar{x}, \Sigma_x = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)(x_i - \mu_x)^T \tag{27}$$

Option 1

$$\mu_1(1, 0), \mu_2(10^5, 0) \tag{28}$$

So that μ_1 has all the points while μ_2 has none.

Option 2

$$\mu_1(-1,0), \mu_2(2,0) \tag{29}$$

So that mu_1 has the left point and the middle circle while mu_2 has the right circle.

Conclusion

Because of this, the **Option 1** attains a lower value of the objective function.

Question 4

There are two rules of PLA:
First rule: The best W is exist.

$$\exists w_i \in W_f, \forall x_i \in X_n, s.t. y_i = sign(W_f^T x_i) \tag{30}$$

Second rule: Only if the clustering result is not correct, the W will be updated.

$$y_{n(t)} W_t^T X_{n(t)} \leqslant 0 \tag{31}$$

In every iteration, $y_{n(t)} W_t^T X_{n(t)} \geqslant \min\limits_n y_n W_f^T X_n > 0$. Then we use $W_f^T \cdot W_t$ to measure the similarity between W_f and W_t

$$\begin{aligned} W_f^T \cdot W_t &= W_f^T \\ &= W_f(W_(t-1) + \lambda(d_t - y_t)X_t \\ &\geqslant W_f^T \cdot W_0 + t\min\limits_n y_n W_f^T X_n \\ &= \min\limits_n y_n W_f^T X_n \end{aligned} \tag{32}$$

Then we should distinguish the update direction:

$$\begin{aligned} \|W_{t+1}\|^2 &= \|W_t + y_n X_n\|^2 \\ &= \|W_t\|^2 + 2y_n W_t^T X_n + \|y_n X_n\|^2 \\ &\leqslant \|W_t\|^2 + \|y_n X_n\|^2 \\ &\leqslant \|W_t\|^2 + \max\limits_n \|y_n X_n\|^2 \\ &\leqslant \|W_0\|^2 + T \cdot \max\limits_n \|y_n X_n\|^2 \end{aligned} \tag{33}$$

Finally, we assume that T is the time of iterations.

$$\begin{aligned}
\frac{W_f^T}{\|W_f^T\|}\frac{W_T}{\|W_T\|} &= \frac{T\cdot \min\limits_{n} y_n W_f^T X_n}{\|W_f^T\|\|W_T\|} \\
&\geqslant \frac{T\cdot \min\limits_{n} y_n W_f^T X_n}{\|W_f^T\|\sqrt{T}\cdot \max\limits_{n}\|y_n X_n\|^2} \\
&\geqslant \frac{\sqrt{T}\cdot \min\limits_{n} y_n W_f^T X_n}{\|W_f^T\|\cdot \max\limits_{n}\|y_n X_n\|^2} \\
&\because \frac{W_f^T}{\|W_f^T\|}\frac{W_T}{\|W_T\|} \leqslant \cos\left(\langle W_f, W_T\rangle\right) \leqslant 1 \\
&\therefore \sqrt{T} \leqslant \frac{\|W_f^T\|\cdot \max\limits_{n}\|y_n X_n\|^2}{\min\limits_{n} y_n W_f^T X_n} \\
&\therefore T \leqslant \left(\frac{\|W_f^T\|\cdot \max\limits_{n}\|y_n X_n\|^2}{\min\limits_{n} y_n W_f^T X_n}\right)^2
\end{aligned} \tag{34}$$