Pattern Recognition Theory 18794 Homework 1

Mengwen He (Alex)

October 6, 2016

Problem 1

1. A vector can be broken up into its direction and magnitude. The equation $A\vec{v} = \vec{w}$ can seen as applying a linear transformation to vector $\vec{v}$ to get a different vector $\vec{w}$. If $\vec{v}$ happens to be an eigenvector of A, describe what happens to the direction of $\vec{v}$ after transformation?
   Answer:
   $\because$ $\vec{v}$ is an eigenvector of A
   $\therefore$ $A\vec{v} = \lambda\vec{v}$
   If the eigenvalue $\lambda > 0$, then the direction of $\vec{v}$ will not change.
   If the eigenvalue $\lambda < 0$, then the direction of $\vec{v}$ will be inversed.
   If the eigenvalue $\lambda = 0$, then the direction of $\vec{v}$ will to be an arbitrary direction (definition of the zero vector).

2. Given

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 3 & 3 & 3 \end{bmatrix},\ B = \begin{bmatrix} -1 & 2 \\ 2 & 2 \end{bmatrix},\ C = \begin{bmatrix} 2 & 8 & 1 & 0 \\ 3 & 12 & 2 & 0 \\ 2 & 8 & 3 & -1 \end{bmatrix},\ D = \begin{bmatrix} 2 & 8 & 1 & 1 \\ 3 & 12 & 2 & 1 \\ 2 & 8 & 3 & -1 \end{bmatrix}$$

   (a) What is the rank of each matrix?
   Answer:
   - For the square matrix A:

$$|A - \lambda I| = 0 \Rightarrow \lambda^3 - 8\lambda^2 + 6\lambda = 0 \Rightarrow \begin{cases} \lambda_1 &= 4 + \sqrt{10} \\ \lambda_2 &= 4 - \sqrt{10} \\ \lambda_3 &= 0 \end{cases} \Rightarrow rank(A) = 2$$

   - For the square matrix B:

$$|B - \lambda I| = 0 \Rightarrow \lambda^2 - \lambda - 6 = 0 \Rightarrow \begin{cases} \lambda_1 &= 3 \\ \lambda_2 &= -2 \end{cases} \Rightarrow rank(B) = 2$$

   - For the "fat" matrix C:

$$\begin{bmatrix} 2 & 8 & 1 & 0 \\ 3 & 12 & 2 & 0 \\ 2 & 8 & 3 & -1 \end{bmatrix} \xrightarrow[R_3 = R_3 - R_1]{R_2 = R_2 - \frac{3}{2}R_1} \begin{bmatrix} 2 & 8 & 1 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 2 & -1 \end{bmatrix}$$

$$\xrightarrow{R_3 = R_3 - 4R_2} \begin{bmatrix} 2 & 8 & 1 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \Rightarrow rank(C) = 3$$

- For the "fat" matrix D:

$$\begin{bmatrix} 2 & 8 & 1 & 1 \\ 3 & 12 & 2 & 1 \\ 2 & 8 & 3 & -1 \end{bmatrix} \xrightarrow[]{\substack{R_2 = R_2 - \frac{3}{2}R_1 \\ R_3 = R_3 - R_1}} \begin{bmatrix} 2 & 8 & 1 & 1 \\ 0 & 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 2 & -2 \end{bmatrix}$$

$$\xrightarrow{R_3 = R_3 - 4R_2} \begin{bmatrix} 2 & 8 & 1 & 1 \\ 0 & 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 0 & 0 \end{bmatrix} \Rightarrow rank(D) = 2$$

(b) What is the trace of each matrix?
Answer:

- For square matrix A: $trace(A) = 1 + 4 + 3 = (4 + \sqrt{10}) + (4 - \sqrt{10}) + 0 = 8$
- For square matrix B: $trace(B) = -1 + 2 = 3 + (-2) = 1$
- For "fat" matrices C and D: trace is only defined for square matrix.

(c) What is the determinant of matrices A, B?
Answer:

- For matrix A: $det(A) = (4 + \sqrt{10}) * (4 - \sqrt{10}) * 0 = 0$
- For matrix B: $det(B) = 3 * (-2) = -6$

(d) Find the eigenvalues and eigenvectors of B.
Answer:
From 2.(a), B's eigenvalues are $\lambda_1 = 3,\ \lambda_2 = -2$
Assume B's eigenvectors are $\vec{v}_1 = [v_{11}, v_{12}]^T (v_{11} \geq 0),\ \vec{v}_2 = [v_{21}, v_{22}]^T (v_{21} \geq 0)$

$$\begin{bmatrix} -1 & 2 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} v_{11} & v_{21} \\ v_{12} & v_{22} \end{bmatrix} = \begin{bmatrix} v_{11} & v_{21} \\ v_{12} & v_{22} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \Rightarrow \begin{cases} 2v_{11} - v_{12} = 0 \\ v_{21} + 2v_{22} = 0 \\ v_{11}^2 + v_{12}^2 = 1 \\ v_{21}^2 + v_{22}^2 = 1 \end{cases} \Rightarrow \begin{cases} v_{11} = \frac{\sqrt{5}}{5} \\ v_{12} = \frac{2\sqrt{5}}{5} \\ v_{21} = \frac{\sqrt{5}}{5} \\ v_{22} = -\frac{2\sqrt{5}}{5} \end{cases}$$

Therefore B's eigenvector are $\vec{v}_1 = \left[\frac{\sqrt(5)}{5}, \frac{2\sqrt{5}}{5}\right]^T$, $\vec{v}_2 = \left[\frac{\sqrt(5)}{5}, -\frac{2\sqrt{5}}{5}\right]^T$

(e) What is the dimension of the null space of each of the matrices?
Answer:

- For the matrix A: $dim(Null(A)) = dim(A) - rank(A) = 3 - 2 = 1$
- For the matrix B: $dim(Null(B)) = dim(B) - rank(B) = 2 - 2 = 0$
- For the matrix C:
  - Right null space: $dim(Null(C)) = col\_dim(C) - rank(C) = 4 - 3 = 1$
  - Left null space: $dim(Null(C^T)) = col\_dim(C^T) - rank(C^T) = 3 - 3 = 0$
- For the matrix D:
  - Right null space: $dim(Null(D)) = col\_dim(D) - rank(D) = 4 - 2 = 2$
  - Left null space: $dim(Null(D^T)) = col\_dim(D^T) - rank(D^T) = 3 - 2 = 1$

Problem 2

1. Prove that if P is an orthogonal matrix, then PP^T is a projection matrix.
   Answer:

   (a) Method 1:
   $\because P$ is an orthogonal matrix
   $\therefore PP^T = I$
   $\therefore PP^T$ is a projection matrix.

   (b) Method 2:
   $\because P$ is an orthogonal matrix
   $\therefore (PP^T)(PP^T) = P(P^TP)P^T = PP^T$
   $\therefore PP^T$ is a projection matrix.

2. Under what condition(s), for arbitrary non-zero matrices $A, B \in \mathbb{R}^{m\times n}$, do we have $rank(A+B) = rank(A) + rank(B)$? (Hint: Think singular values/vectors)
   Answer:
   $\because dim(colsp([A,B])) = dim(colsp(A)) + dim(colsp(B)) - dim(colsp(A) \bigcap colsp(B))$
   $\therefore dim(colsp([A,B])) = rank(A) + rank(B) - dim(colsp(A) \bigcap colsp(B))$
   $\because colsp(A+B)$ is a subspace of $colsp([A,B])$
   $\therefore dim(colsp([A,B])) \geq dim(colsp(A+B)) = rank(A+B)$
   $\therefore rank(A) + rank(B) \geq rank(A+B) + dim(colsp(A) \bigcap colsp(B))$
   If we have $rank(A+B) = rank(A) + rank(B)$,
   then $dim(colsp(A) \bigcap colsp(B)) = 0 \Rightarrow colsp(A) \bigcap colsp(B) = \{\vec{0}\} \Rightarrow A^TB = 0$
   Therefore, we get a necessary condition as below:
   For all matrices $A, B \in \mathbb{R}^{m\times n}$, $rank(A+B) = rank(A) + rank(B) \Rightarrow A^TB = 0$.
   So, if under $colsp(A) \bigcap colsp(B) = \{\vec{0}\}$ (strong) or $A^TB = 0$ (weak), for all matrices $A, B \in \mathbb{R}^{m\times n}$, we can have $rank(A+B) = rank(A) + rank(B)$.

3. In general, does a relation exist between the rank of a sum of matrices and the sum of the ranks of the individual matrices?
   Answer:
   To find the relationship between $rank(\sum_{i=1}^N A_i)$ and $\sum_{i=1}^N rank(A_i)$
   From 2, we know that $rank(A+B) \leq rank(A) + rank(B) - dim(colsp(A) \bigcap colsp(B))$
   Therefore, we gradually derive that relationship as below:
   $rank(\sum_{i=1}^N A_i)$
   $\leq rank(A_1) + rank(\sum_{i=2}^N A_i) - dim(colsp(A_1) \bigcap colsp(\sum_{i=2}^N A_i))$
   $\leq rank(A_1) + rank(A_2) + rank(\sum_{i=3}^N A_i) - dim(colsp(A_2) \bigcap colsp(\sum_{i=3}^N A_i)))$
   ...
   $\leq \sum_{i=1}^N rank(A_i)$

4. Under the premise of problem 1.2, you arrived at a condition. Is the condition sufficient to ensure that $\sigma_{max}(A+B) = max(\sigma_{max}(A), \sigma_{max}(B))$? If not, what else do you need to constrain ? Here, $\sigma_{max}(A)$ is the maximum singular value of A
   Answer:
   $\because$ From the definition of matrix LSV-norm:

   $$||A||_2 = \max_{\vec{x}:||\vec{x}||_2=1} ||A\vec{x}||_2 = \sigma_{max}(A)$$

   $\therefore$

   $$\sigma_{max}(A+B) = \max_{\vec{x}:||\vec{x}||_2=1} ||(A+B)\vec{x}||_2$$

$\therefore$

$$\sigma^2_{max}(A+B) = \max_{\vec{x}:||\vec{x}||_2=1}((A+B)\vec{x})^T((A+B)\vec{x})$$

$$\sigma^2_{max}(A+B) = \max_{\vec{x}:||\vec{x}||_2=1}((A\vec{x})^T A\vec{x} + (B\vec{x})^T B\vec{x} + 2\vec{x}^T(A^T B)\vec{x})$$

$\because$ we are under the condition $rank(A+B) = rank(A) + rank(B) \Rightarrow A^T B = 0$ given by 2
$\therefore$

$$\sigma^2_{max}(A+B) = \max_{\vec{x}:||\vec{x}||_2=1}((A\vec{x})^T A\vec{x} + (B\vec{x})^T B\vec{x})$$

$$\sigma^2_{max}(A+B) = \max_{\vec{x}:||\vec{x}||_2=1}(||A\vec{x}||_2^2 + ||B\vec{x}||_2^2)$$

Assume:

$$\vec{x}^*_A = \arg\max_{\vec{x}:||\vec{x}||_2=1}||A\vec{x}||_2$$

$$\vec{x}^*_B = \arg\max_{\vec{x}:||\vec{x}||_2=1}||B\vec{x}||_2$$

$$\vec{x}^* = \arg\max_{\vec{x}:||\vec{x}||_2=1}||(A+B)\vec{x}||_2$$

- If $\vec{x}^* = \vec{x}^*_A$, then $\sigma_{max}(A+B) = \sqrt{\sigma^2_{max}(A) + ||B\vec{x}^*_A||_2^2} \geq \sigma_{max}(A)$
- If $\vec{x}^* = \vec{x}^*_B$, then $\sigma_{max}(A+B) = \sqrt{||A\vec{x}^*_B||_2^2 + \sigma^2_{max}(B)} \geq \sigma_{max}(B)$

$\therefore$ with the condition $rank(A+B) = rank(A) + rank(B)$, we derived that $\sigma_{max}(A+B) \geq max(\sigma_{max}(A), \sigma_{max}(B))$
$\therefore$ $rank(A+B) = rank(A) + rank(B) \not\Rightarrow \sigma_{max}(A+B) = max(\sigma_{max}(A), \sigma_{max}(B))$ (it is not a sufficient condition)
However, we can add another condition that $rowsp(A) \bigcap rowsp(B) = \{\vec{0}\} \Rightarrow AB^T = 0$
First we can guarantee that:

- If $\vec{x}^* = \vec{x}^*_A \in rowsp(A)$, then $\sigma_{max}(A+B) = \sqrt{\sigma^2_{max}(A) + ||B\vec{x}^*_A||_2^2} = \sigma_{max}(A)$
- If $\vec{x}^* = \vec{x}^*_B \in rowsp(B)$, then $\sigma_{max}(A+B) = \sqrt{||A\vec{x}^*_B||_2^2 + \sigma^2_{max}(B)} = \sigma_{max}(B)$

Then, we need to demonstrate that:

$$\forall \text{ unit } \vec{x} \in \mathbb{R}^m, \sqrt{||A\vec{x}||_2^2 + ||B\vec{x}||_2^2} \leq max(\sigma_{max}(A), \sigma_{max}(B))$$

Do SVD decomposition on any matrix $M \in \mathbb{R}^{n\times m}$: $M = U_M S_M V_M^T$
$\because$ For $M\vec{x} = U_M S_M V_M^T \vec{x}$, the unit $\vec{x}$ is first projected into $rowsp(V_M)$ by V_M, then scaled by S_M along the bases in $rowsp(V_M)$, and finally projected into $colsp(U_M)$ (but actually into $colsp(M)$) by U_M.
So we define a $\hat{S}_M$ from S_M as below:

$$\hat{S}^{ij}_M = \begin{cases} 1, & S^{ij}_M \neq 0 \\ 0, & S^{ij}_M = 0 \end{cases}$$

And we define the energy projected into $colsp(A)$ from unit $\vec{x}$ by using $U_M \hat{S}_M V_M^T$ as

$$E_M(\vec{x}) = ||U_M \hat{S}_M V_M^T \vec{x}||_2^2$$

By definition, we know that (assume $rowsp(A) \bigcap rowsp(B) = \{\vec{0}\}$):

- If unit $\vec{x} \in rowsp(A)$, then $E_A(\vec{x}) = 1$
- If unit $\vec{x} \notin rowsp(A)$, then $E_A(\vec{x}) < 1$
- If unit $\vec{x} \in rowsp(B)$, then $E_A(\vec{x}) = 0$
- If unit $\vec{x} \in rowsp([A, B])$, then $E_A(\vec{x}) + E_B(\vec{x}) = 1$
- If unit $\vec{x} \notin rowsp([A, B])$, then $E_A(\vec{x}) + E_B(\vec{x}) < 1$
- If unit $\vec{x} = \vec{x}^*_A$, then $||A\vec{x}||_2^2 = E_A(\vec{x})\sigma^2_{max}(A)$; otherwise, $||A\vec{x}||_2^2 \leq E_A(\vec{x})\sigma^2_{max}(A)$

$\therefore$

$$\forall \text{ unit } \vec{x} \in \mathbb{R}^m,\ ||A\vec{x}||_2^2 + ||B\vec{x}||_2^2 \leq E_A(\vec{x})\sigma^2_{max}(A) + E_B(\vec{x})\sigma^2_{max}(B)$$
$$\leq (E_A(\vec{x}) + E_B(\vec{x}))max(\sigma^2_{max}(A), \sigma^2_{max}(B)) \leq max(\sigma^2_{max}(A), \sigma^2_{max}(B))$$

$\therefore$ If $rank(A + B) = rank(A) + rank(B)$ and $rowsp(A) \bigcap rowsp(B) = \{\vec{0}\}$, then $\sigma_{max}(A + B) = max(\sigma_{max}(A), \sigma_{max}(B))$.

5. Let $A \in \mathbb{R}^{N\times N}$ be a full-rank matrix, with columns $\vec{a}_1, ..., \vec{a}_N$. Prove that for every $\vec{\beta} \in \mathbb{R}^N$ with $\beta_i \geq 0$, $\sum_{i=1}^N \beta_i = 1$, we have $rank(A - Q) = N - 1$. Where $Q = [A\vec{\beta}, A\vec{\beta}, ..., A\vec{\beta}]$ is formed by repeating the column vector $(A\vec{\beta})$, N times, i.e. $Q \in \mathbb{R}^{N\times N}$. (Hint: Try to scale the problem down and approach it from a geometric paradigm.)

Answer:
$\because \forall\vec{\beta} \in \mathbb{R}^N:\ \beta_i \geq 0,\ \sum_{i=1}^N \beta_i = 1$
$\therefore P = \{\vec{\beta}:\ \vec{\beta} \in \mathbb{R}^N,\ \beta_i \geq 0,\ \sum_{i=1}^N \beta_i = 1\}$ forms a $(N-1)$-Dim hyperplane in N-Dim space.
$\therefore \exists\vec{\gamma}$, $\forall\vec{\beta}_i$ and $\vec{\beta}_j \in P$, $(\vec{\beta}_i - \vec{\beta}_j)\cdot\vec{\gamma} = 0$
$\because A$ is a full-rank matrix
$\therefore$ after the affine transformation of P by A, a transformed normal vector $\vec{\gamma}_A = (A^T)^{-1}\vec{\gamma}$:

$$A(\vec{\beta}_i - \vec{\beta}_j)\cdot\vec{\gamma}_A == (\vec{\beta}_i - \vec{\beta}_j)^T(A^T(A^T)^{-1})\vec{\gamma} = (\vec{\beta}_i - \vec{\beta}_j)\cdot\vec{\gamma} = 0$$

Define $\vec{\beta}^i \in \mathbb{R}^N$ as below:

$$\vec{\beta}^i_j = \begin{cases} 1 & j = i \\ 0 & j \neq i \end{cases}$$

$\therefore \vec{\beta}^i \in P$.
$\therefore \forall\vec{\beta} \in P$, and $\forall i \in \{1, 2, \ldots, N\}$

$$\begin{aligned} A(\vec{\beta} - \vec{\beta}^i)\cdot\vec{\gamma}_A &= 0 \\ (A\vec{\beta} - A\vec{\beta}^i)\cdot\vec{\gamma}_A &= 0 \\ (A\vec{\beta} - \vec{\alpha}_i)\cdot\vec{\gamma}_A &= 0 \end{aligned}$$

$\therefore\ \exists$ a non-zero $\vec{\gamma}_A \in \mathbb{R}^N,\ (A - Q)^T\vec{\gamma}_A = 0$
$\because P$ is $(N-1)$-Dim hyperplane,
$\therefore\ rank(A - Q) = N - 1$

Problem 3

1. Find the maximum and minimum values of $f(x,y) = 81x^2 + y^2$ subject to the constraint $4x^2 + y^2 = 9$
   Answer:

$$L(x,y,\lambda) = 81x^2 + y^2 - \lambda(9 - 4x^2 - y^2)$$

$$\left\{\begin{array}{lcl} \frac{\partial L}{\partial \lambda} & = & 4x^2 + y^2 - 9 = 0 \\ \frac{\partial L}{\partial x} & = & (162 + 8\lambda)x = 0 \\ \frac{\partial L}{\partial y} & = & (2 + 2\lambda)y = 0 \end{array}\right. \Rightarrow \left\{\begin{array}{lcl} \lambda & = & -1 \\ x & = & 0 \\ y & = & 3 \end{array}\right. \text{ or } \left\{\begin{array}{lcl} \lambda & = & -\frac{81}{4} \\ x & = & \frac{3}{2} \\ y & = & 0 \end{array}\right.$$

$\therefore$

$$max(f(x,y)) = f(\frac{3}{2}, 0) = \frac{729}{4}$$

$$min(f(x,y)) = f(0,3) = 9$$

2. Find the maximum and minimum values of $f(x,y,z) = x + y + 2z$ on the surface $x^2 + y^2 + z^2 = 3$
   Answer:

$$L(x,y,z,\lambda) = x + y + 2z - \lambda(3 - x^2 - y^2 - z^2)$$

$$\left\{\begin{array}{lcl} \frac{\partial L}{\partial \lambda} & = & x^2 + y^2 + z^2 - 3 = 0 \\ \frac{\partial L}{\partial x} & = & 2\lambda x + 1 = 0 \\ \frac{\partial L}{\partial y} & = & 2\lambda y + 1 = 0 \\ \frac{\partial L}{\partial z} & = & 2\lambda z + 2 = 0 \end{array}\right. \Rightarrow \left\{\begin{array}{lcl} \lambda & = & \frac{\sqrt{2}}{2} \\ x & = & -\frac{\sqrt{2}}{2} \\ y & = & -\frac{\sqrt{2}}{2} \\ z & = & -\sqrt{2} \end{array}\right. \text{ or } \left\{\begin{array}{lcl} \lambda & = & -\frac{\sqrt{2}}{2} \\ x & = & \frac{\sqrt{2}}{2} \\ y & = & \frac{\sqrt{2}}{2} \\ z & = & \sqrt{2} \end{array}\right.$$

$\therefore$

$$max(f(x,y,z)) = f(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, \sqrt{2}) = 3\sqrt{2}$$

$$min(f(x,y,z)) = f(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, -\sqrt{2}) = -3\sqrt{2}$$

Problem 6

The two random variables X and Y are jointly Gaussian. Let

$$X \sim \mathcal{N}(\mu_x, \sigma_x^2),$$
$$Y \sim \mathcal{N}(\mu_y, \sigma_y^2).$$

The correlation coefficient ϕ between X and Y is 0.5. Two new PDFs are generated as follows,

$$Z = X + Y,$$
$$W = 3X + 2Y.$$

Determine the decision boundaries of the two PDFs which give the minimum errors using $\mu_x = 0$, $\mu_y = 4$, $\sigma_x = 1$, $\sigma_y = 2$ for the cases below.

Preparation:

(1) Given:

$$X \sim \mathcal{N}(\mu_x, \sigma_x^2)$$
$$Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$$

Then:

$$\mu_x = E(X)$$
$$\sigma_x^2 = E((X - E(X))^2) = E(X^2) - E(X)^2$$
$$\sigma_{xy}^2 = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y)$$
$$\rho_{xy} = \frac{\sigma_{xy}^2}{\sigma_x \sigma_y}$$

(2) For $Z = X + Y$

$$\mu_z = E(X + Y) = E(X) + E(Y) = \mu_x + \mu_y = 4$$
$$\sigma_z^2 = E((X + Y)^2) - E(X + Y)^2 = (E(X^2) - E(X)^2) + (E(Y^2) - E(Y)^2) + 2(E(XY) - E(X)E(Y))$$
$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2 + 2\rho_{xy}\sigma_x\sigma_y = 7$$

For $W = 3X + 2Y$

$$\mu_w = E(3X + 2Y) = 3E(X) + 2E(Y) = 3\mu_x + 2\mu_y = 8$$
$$\sigma_w^2 = E((3X+2Y)^2) - E(3X+2Y)^2 = 9(E(X^2) - E(X)^2) + 4(E(Y^2) - E(Y)^2) + 12(E(XY) - E(X)E(Y))$$
$$\sigma_w^2 = 9\sigma_x^2 + 4\sigma_y^2 + 12\rho_{xy}\sigma_x\sigma_y = 37$$

(3) The decision boundary is derived by the following equation:

$$P(\omega_1) f(x|\omega_1) = P(\omega_2) f(x|\omega_2)$$
$$ln\left(\frac{f(x|\omega_1)}{f(x|\omega_2)}\right) = ln\left(\frac{P(\omega_2)}{P(\omega_1)}\right)$$
$$ln\left(\frac{\frac{1}{\sqrt{2\pi\sigma_z^2}} exp\left(-\frac{(x-\mu_z)^2}{2\sigma_z^2}\right)}{\frac{1}{\sqrt{2\pi\sigma_w^2}} exp\left(-\frac{(x-\mu_w)^2}{2\sigma_w^2}\right)}\right) = ln\left(\frac{P(\omega_2)}{P(\omega_1)}\right)$$
$$ln\left(\frac{\sigma_w}{\sigma_z}\right) + \frac{(x - \mu_w)^2}{2\sigma_w^2} - \frac{(x - \mu_z)^2}{2\sigma_z^2} = ln\left(\frac{P(\omega_2)}{P(\omega_1)}\right)$$

1. $P(\omega_1) = P(\omega_2)$
   Answer:

$$ln\left(\frac{\sigma_w}{\sigma_z}\right) + \frac{(x-\mu_w)^2}{2\sigma_w^2} - \frac{(x-\mu_z)^2}{2\sigma_z^2} = ln\left(\frac{P(\omega_2)}{P(\omega_1)}\right)$$

$$ln\left(\frac{\sqrt{37}}{\sqrt{7}}\right) + \frac{(x-8)^2}{2*37} - \frac{(x-4)^2}{2*7} = 0$$

$$x1 = -1.2898240621304666,\ x2 = 7.4231573954638$$

$$\begin{cases} x > x_1 \ and \ x < x_2 & \Rightarrow & \omega_1 \\ x < x_1 \ or \ x > x_2 & \Rightarrow & \omega_2 \end{cases}$$

2. $P(\omega_1) = 3P(\omega_2)$
   Answer:

$$ln\left(\frac{\sigma_w}{\sigma_z}\right) + \frac{(x-\mu_w)^2}{2\sigma_w^2} - \frac{(x-\mu_z)^2}{2\sigma_z^2} = ln\left(\frac{P(\omega_2)}{P(\omega_1)}\right)$$

$$ln\left(\frac{\sqrt{37}}{\sqrt{7}}\right) + \frac{(x-8)^2}{2*37} - \frac{(x-4)^2}{2*7} = ln\left(\frac{1}{3}\right)$$

$$x1 = -3.0935592751662626,\ x2 = 9.226892608499595$$

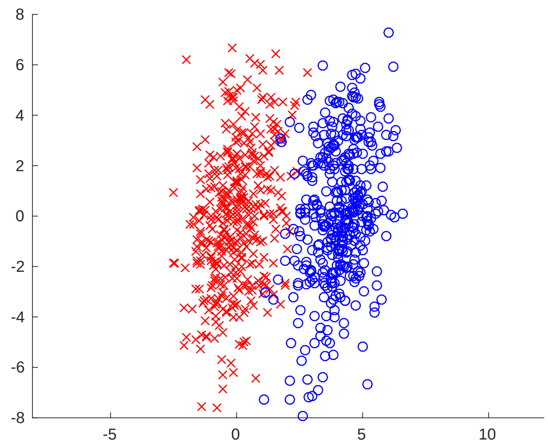
$$\begin{cases} x > x_1 & and & x < x_2 & \Rightarrow & \omega_1 \\ x < x_1 & or & x > x_2 & \Rightarrow & \omega_2 \end{cases}$$

Problem 7

MATLAB. For each distribution pair below: Generate 400 samples from each of the distributions. (Hint: Look at the randn documentation). Print a plot with your samples. Use x for points from one distribution, and o for points from the other distribution. Make the axis limits axis([-6 10 -8 8]). On your plot printouts, draw the general shape (it does not need to be exact) of the decision boundary between the distributions (assume equal priors). Do not compute the decision boundary mathematically. Attach the plots and your .m file. Describe how the decision boundary changes when the distributions do not have the same covariance.

1.

$$x_1 \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 8 \end{bmatrix}\right),\ x_2 \sim \mathcal{N}\left(\begin{bmatrix} 4 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 8 \end{bmatrix}\right)$$

- 


- ```
figure
hold on;

mu1=[0;0];
sigma1=[1 1 ; 1 8];
r1=chol(sigma1);
x1=repmat(mu1,1,400)+r1'*randn(2,400);

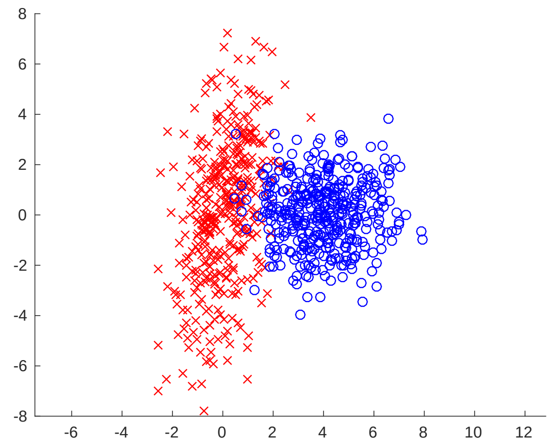
mu2=[4;0];
sigma2=[1 1 ; 1 8];
r2=chol(sigma2);
x2=repmat(mu2,1,400)+r2'*randn(2,400);

plot(x1(1,:),x1(2,:),'rx',x2(1,:),x2(2,:),'bo');
axis([-6 10 -8 8]);
axis equal;

hold off;
```
```

2.

$$x_1 \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 8 \end{bmatrix}\right),\ x_2 \sim \mathcal{N}\left(\begin{bmatrix} 4 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\right)$$

- 


- ```
figure
hold on;

mu1=[0;0];
sigma1=[1 1 ; 1 8];
r1=chol(sigma1);
x1=repmat(mu1,1,400)+r1'*randn(2,400);

mu2=[4;0];
sigma2=[2 0 ; 0 2];
r2=chol(sigma2);
x2=repmat(mu2,1,400)+r2'*randn(2,400);

plot(x1(1,:),x1(2,:),'rx',x2(1,:),x2(2,:),'bo');
axis([-6 10 -8 8]);
axis equal;

hold off;
```
```

Problem 8

A two-dimensional feature vector is use to select one of two classes shown in Figure 1. The joint PDFs are uniform over a square for class ω_1 and a triangle for class ω_2. For each case, determine the probability of assigning x to the wrong class ω (compute ϵ_1 and ϵ_2), and the total probability of error (compute P_e).

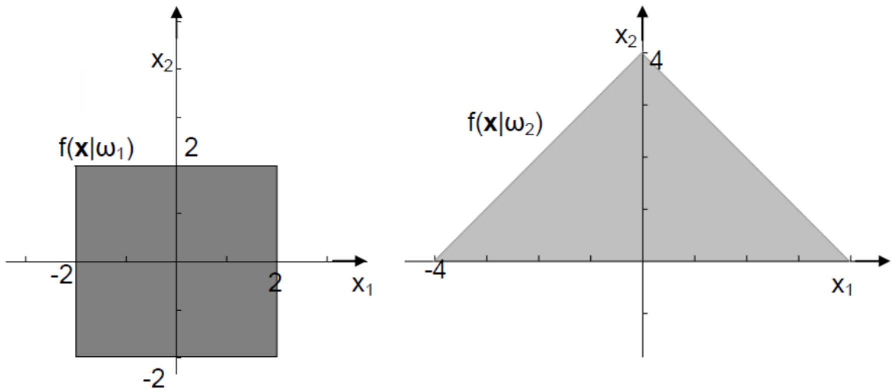


Figure 1: Corresponds to Problem 8

1. Case 1: $P(\omega_1) = P(\omega_2)$
   Answer:
   Define the square as A, and the triangle as B
   Then the area of A, $S(A) = 16$, and the area of B, $S(B) = 16$.

$$\begin{cases} f(\vec{x}|\omega_1) &= \begin{cases} \frac{1}{16} & \vec{x} \in A \\ 0 & \vec{x} \notin A \end{cases} \\ f(\vec{x}|\omega_2) &= \begin{cases} \frac{1}{16} & \vec{x} \in B \\ 0 & \vec{x} \notin B \end{cases} \end{cases}$$

$$\begin{cases} \epsilon_1 &= f(\vec{x}|\omega_1)S(B-A) = \frac{1}{16} * 8 = 1/2 \\ \epsilon_2 &= f(\vec{x}|\omega_2)S(A-B) = \frac{1}{16} * 8 = 1/2 \end{cases}$$

$$P_e = \epsilon_1 P(\omega_1) + \epsilon_2 P(\omega_2) = P(\omega_2)$$

2. Case 2: $P(\omega_1) = 2P(\omega_2)$
   Answer:

$$\begin{cases} \epsilon_1 &= f(\vec{x}|\omega_1)S(B-A) = \frac{1}{16} * 8 = 1/2 \\ \epsilon_2 &= f(\vec{x}|\omega_2)S(A-B) = \frac{1}{16} * 8 = 1/2 \end{cases}$$

$$P_e = \epsilon_1 P(\omega_1) + \epsilon_2 P(\omega_2) = \frac{3}{2}P(\omega_2)$$