

Homework 1

Zihan Zhao

1001103708

1

(a)

First determine $E[Z]$

$$\begin{aligned} E[Z] &= E[(X - Y)^2] \\ &= E[X^2 - 2XY + Y^2] \\ &= E[X^2] - E[2XY] + E[Y^2] \\ &= E[X^2] - 2E[X]E[Y] + E[Y^2] \end{aligned}$$

Since X and Y both are variables from uniform distribution in $[0, 1]$, $E[x] = E[Y]$ and $E[X^2] = E[Y^2]$. Therefore,

continue :

$$\begin{aligned} E[Z] &= 2E[X^2] - 2E[X]^2 \\ &= 2 \int_0^1 x^2 f(x) dx - 2 \left(\int_0^1 x f(x) dx \right)^2 \end{aligned}$$

Note that $f(x)$ is the pdf of X. So $f(x) = \frac{1}{1-0} = 1$. Therefore,

continue :

$$\begin{aligned} E[Z] &= 2 \int_0^1 x^2 dx - 2 \left(\int_0^1 x dx \right)^2 \\ &= 2 * \frac{1}{3} - 2 * \left(\frac{1}{2} \right)^2 \\ &= \frac{1}{6} = 0.16667 \end{aligned}$$

Second determine $Var[Z]$

$$\begin{aligned} Var[Z] &= E[Z^2] - E[Z]^2 \\ &= E[(X - Y)^4] - \left(\frac{1}{6} \right)^2 \\ &= E[X^4 - 4X^3Y + 6X^2Y^2 - 4XY^3 + Y^4] - \left(\frac{1}{6} \right)^2 \\ &= E[X^4] - 4E[X^3]E[Y] + 6E[X^2]E[Y^2] - 4E[X]E[Y^3] + E[Y^4] - \left(\frac{1}{6} \right)^2 \end{aligned}$$

Since X and Y both are variables from uniform distribution in $[0, 1]$, $E[x] = E[Y]$, $E[X^2] = E[Y^2]$, $E[X^3] = E[Y^3]$, and $E[X^4] = E[Y^4]$ Therefore,

$$\begin{aligned} Var[Z] &= 2E[X^4] - 8E[X^3]E[X] + 6E[X^2]^2 - \left(\frac{1}{6}\right)^2 \\ &= 2 \int_0^1 x^4 f(x) dx - 8 \left(\int_0^1 x^3 f(x) dx \right) \left(\int_0^1 x f(x) dx \right) + 6 \left(\int_0^1 x^2 f(x) dx \right)^2 - \left(\frac{1}{6}\right)^2 \\ &= 2 * \frac{1}{5} - 8 * \frac{1}{4} * \frac{1}{2} + 6 \left(\frac{1}{3}\right)^2 - \left(\frac{1}{6}\right)^2 = \frac{7}{180} = 0.38889 \end{aligned}$$

(b)

First determine $E[R]$. Since X_1, \dots, Y_d are i.i.d, $(X_1 - Y_1)^2, \dots, (X_d - Y_d)^2$ are independent. Therefore,

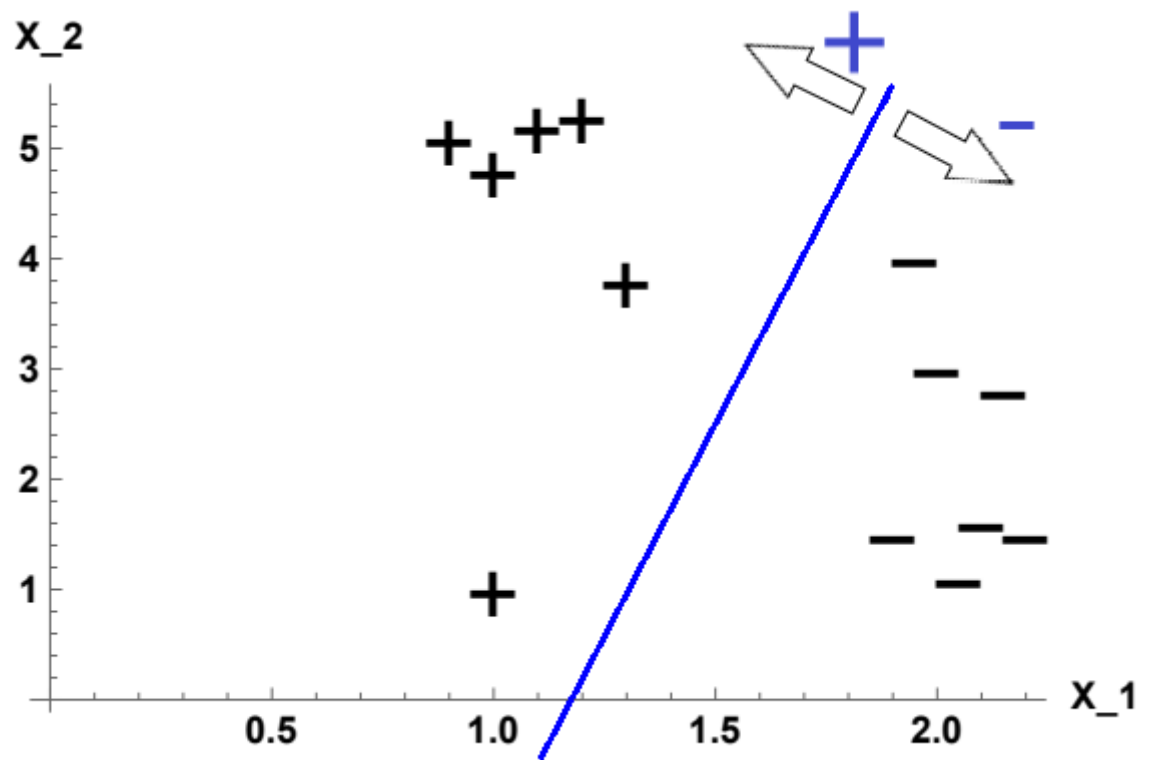
$$\begin{aligned} E[R] &= E\left[\sum_{i=1}^d Z_i\right] \\ &= \sum_{i=1}^d E[Z_i] \\ &= dE[Z] = \frac{d}{6} \end{aligned}$$

Second determine $Var[R]$. Since Z_1, \dots, Z_d are independent, $Cov(Z_i, Z_j) = 0$ where i and j are arbitrary picks from 1 to d. Therefore,

$$\begin{aligned} Var[R] &= Var\left[\sum_{i=1}^d Z_i\right] \\ &= \sum_{i=1}^d Var[Z_i] + 2 \sum_{i < j} Cov[Z_i, Z_j] \\ &= dVar[Z] = \frac{7d}{180} \end{aligned}$$

2

(a)



(b)

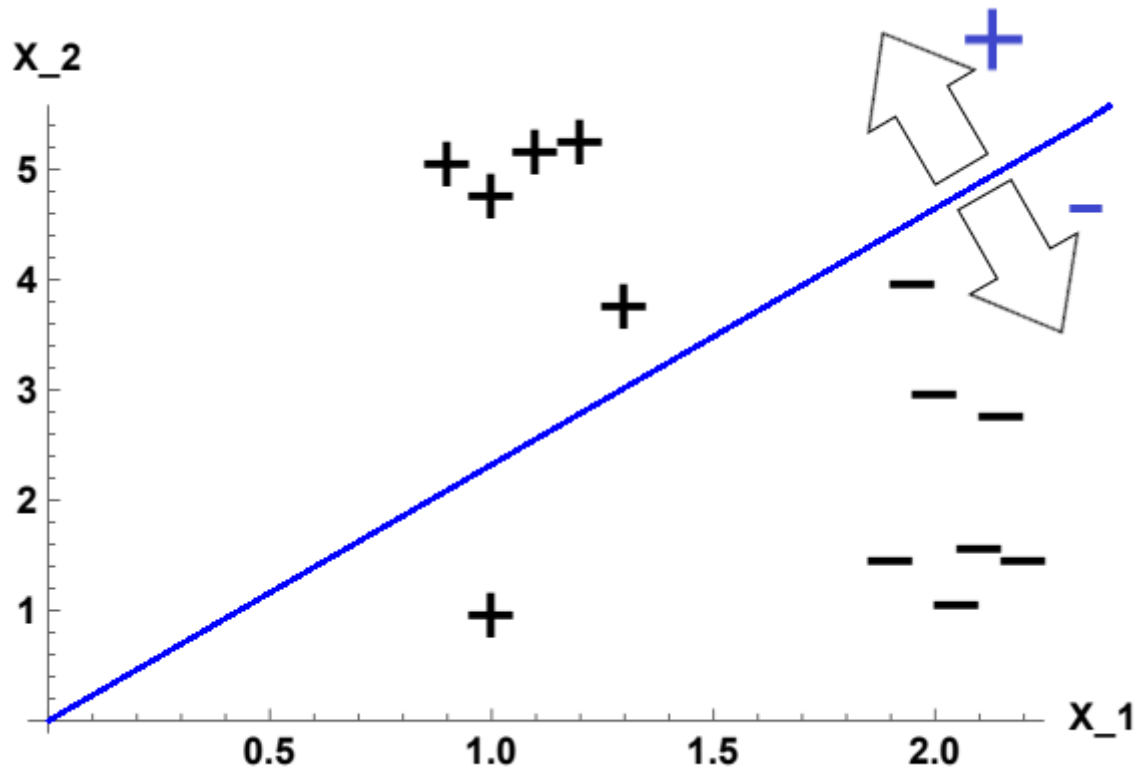
No, it is not unique.

(c)

0. The left side of the boundary marks "+" and the right side marks "-". All "+" points are in left side and all "-" points are in right side.

(d)

As λ approaches ∞ , $J(w)$ heavily depends on $\lambda\omega_0^2$. In order to minimize $\lambda\omega_0^2$, ω_0 should be 0. So the boundary should pass through (0,0). The sketch is shown below.

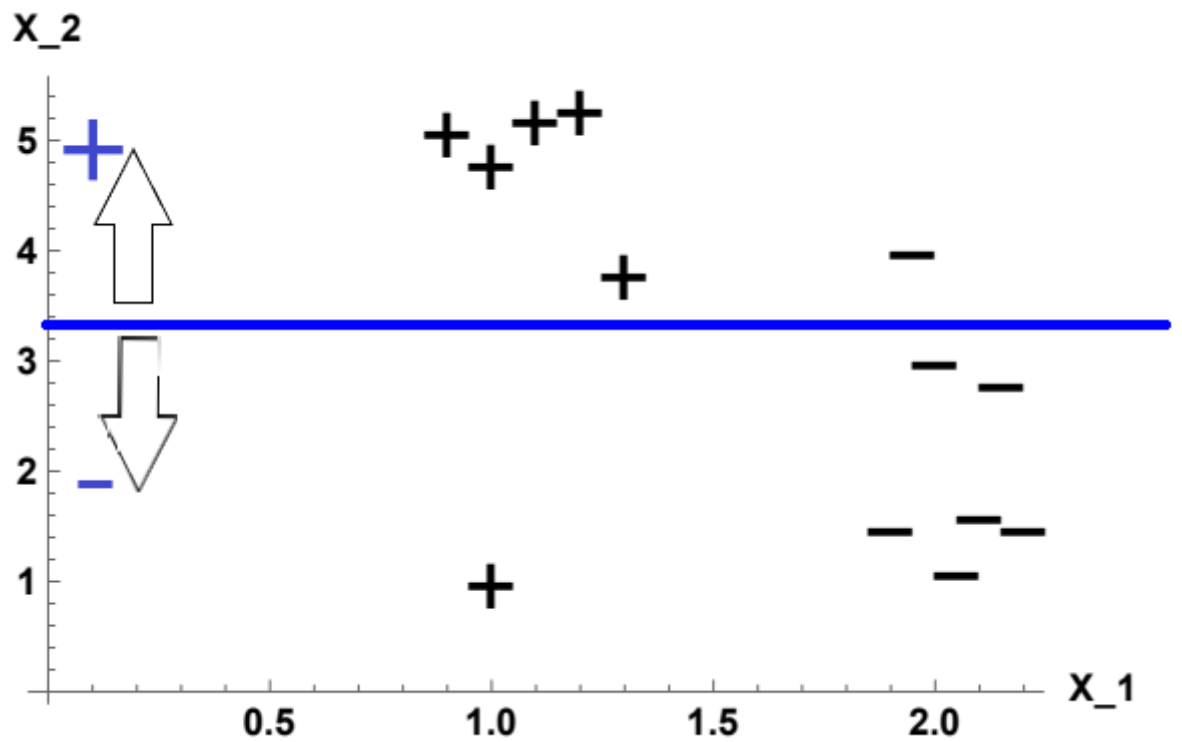


(e)

1. The upper side of the boundary marks "+" and the down side marks "-". All "+" points are in left side except one point is at down side. All "-" points are in right side. So the error is 1.

(f)

Similar with (d), in order to minimize $\lambda\omega_1^2$, ω_1 should be 0. So the boundary line is $\omega_0 + \omega_2x_2 = 0$. It should be a horizontal line. The sketch is shown below.

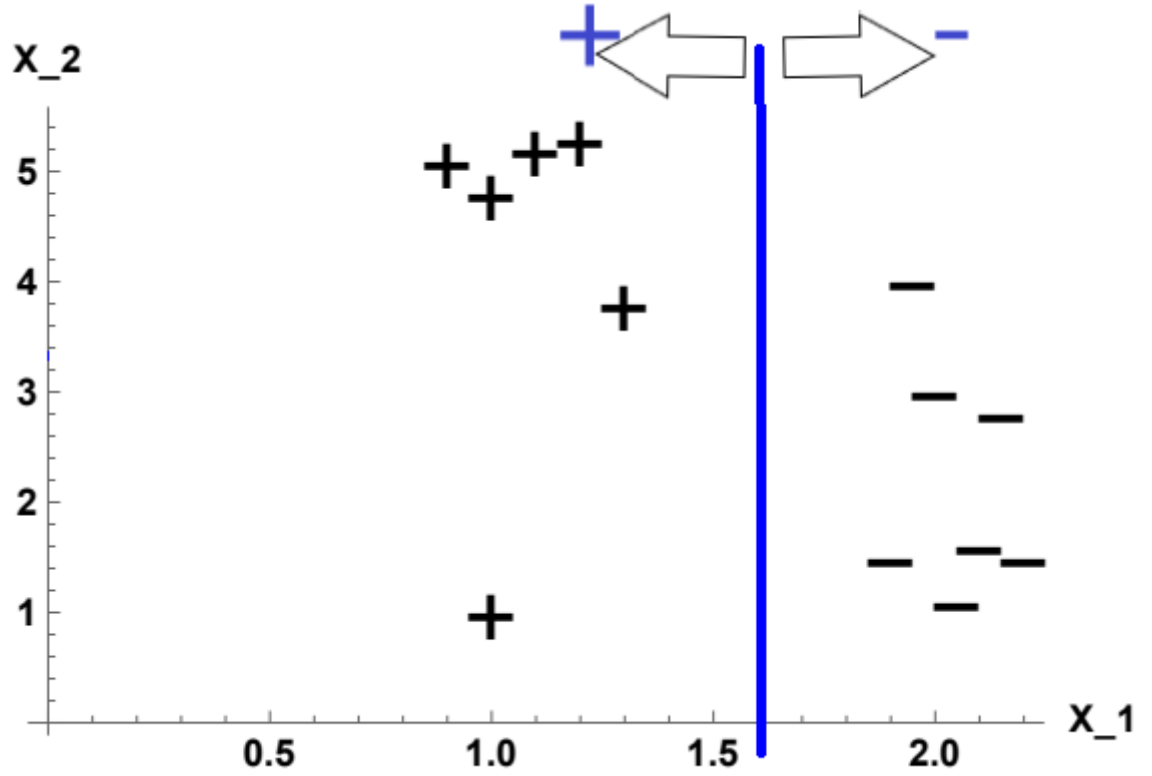


(g)

2. The upper side of the boundary marks "+" and the down side marks "-". All "+" points are in left side except one point is at down side. All "-" points are in right side except one point is at upper side. So the error is 2.

(h)

Similar with (d), in order to minimize $\lambda\omega_2^2$, ω_2 should be 0. So the boundary line is $\omega_0 + \omega_1x_1 = 0$. It should be a vertical line. The sketch is shown below.



(i)

0. The left side of the boundary marks "+" and the right side marks "-". All "+" points are in left side and all "-" points are in right side.

3

(a)

Since $p(x)$ is possibility mass, $0 \leq p(x) \leq 1$. Therefore,

$$\begin{aligned} p(x) &\leq 1 \\ \frac{1}{p(x)} &\leq 1 \\ \log\left(\frac{1}{p(x)}\right) &\geq 0 \end{aligned}$$

since $p(x) \geq 0$,

$$\begin{aligned} p(x) \log\left(\frac{1}{p(x)}\right) &\geq 0 \\ H(X) = \sum_x p(x) \log\left(\frac{1}{p(x)}\right) &\geq 0 \end{aligned}$$

(b)

In order to prove $KL(p||q) \geq 0$, it is equivalent to prove $\sum_x^X p(x) \log(\frac{p(x)}{q(x)}) \geq 0$. Let's define new variable Y which is $y = f(x) = \frac{q(x)}{p(x)}$. Then let's define a new function $\phi(y) = -\log y$, which is convex function. According to Jensen's Inequality,

$$\begin{aligned} E[\phi(Y)] &\geq \phi(E[Y]) \\ \sum_x^X p(x) \phi(f(x)) &\geq \phi(E[Y]) \\ -\sum_x^X p(x) \log\left(\frac{q(x)}{p(x)}\right) &\geq \phi(E[Y]) \\ \sum_x^X p(x) \log\left(\frac{p(x)}{q(x)}\right) &\geq \phi(E[Y]) \\ KL(p||q) &\geq \phi(E[Y]) \\ KL(p||q) &\geq \phi(E_x[f(x)]) \\ KL(p||q) &\geq \phi\left(\sum_x^X p(x) \frac{q(x)}{p(x)}\right) \\ KL(p||q) &\geq \phi\left(\sum_x^X q(x)\right) \\ KL(p||q) &\geq \phi(1) \\ KL(p||q) &\geq -\log(1) \\ KL(p||q) &\geq 0 \end{aligned}$$

(c)

In order to show that

$$I[Y; X] = KL(p(x, y)||p(x)p(y))$$

On left side,

$$\begin{aligned} I[Y; X] &= H(Y) - H(Y|X) \\ &= \sum_y p(y) \log\left(\frac{1}{p(y)}\right) - \left(-\sum_y \sum_x p(x, y) \log(p(y|x))\right) \\ &= \sum_y p(y) \log\left(\frac{1}{p(y)}\right) + \sum_y \sum_x p(x, y) \log(p(y|x)) \end{aligned}$$

On right side,

$$KL(p(x, y)||p(x)p(y)) = \sum_y \sum_x p(x, y) \log\left(\frac{p(x, y)}{p(x)p(y)}\right)$$

note that $p(y|x) = \frac{p(x,y)}{p(x)}$

$$\begin{aligned}
KL(p(x,y)||p(x)p(y)) &= \sum_y \sum_x p(x,y) \log\left(\frac{p(y|x)}{p(y)}\right) \\
&= \sum_y \sum_x (p(x,y)(\log(p(y|x)) + \log\left(\frac{1}{p(y)}\right))) \\
&= \sum_y \sum_x (p(x,y)(\log(p(y|x))) + \sum_y \sum_x (p(x,y) \log\left(\frac{1}{p(y)}\right))) \\
&= \sum_y \sum_x (p(x,y)(\log(p(y|x))) + \sum_y (p(y) \log\left(\frac{1}{p(y)}\right))) \\
&= I[Y; X]
\end{aligned}$$

Therefore right side is equal to left side.

Done.