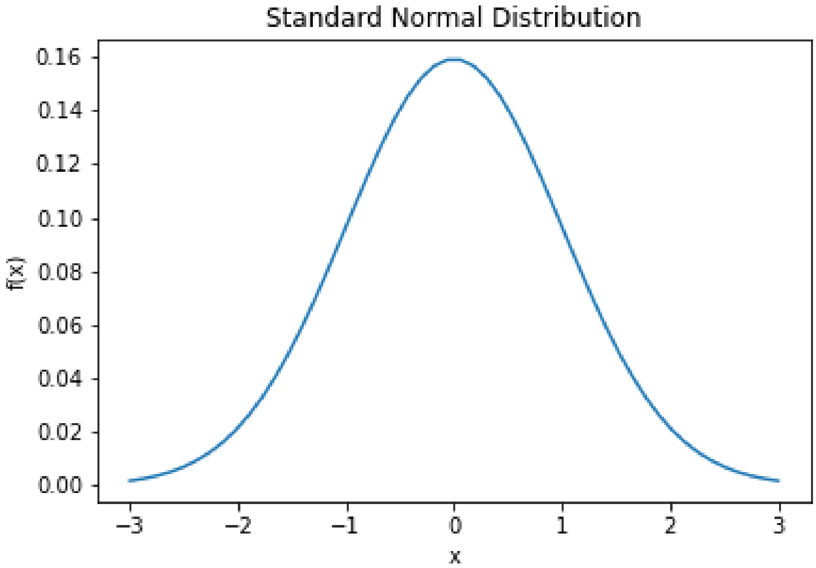
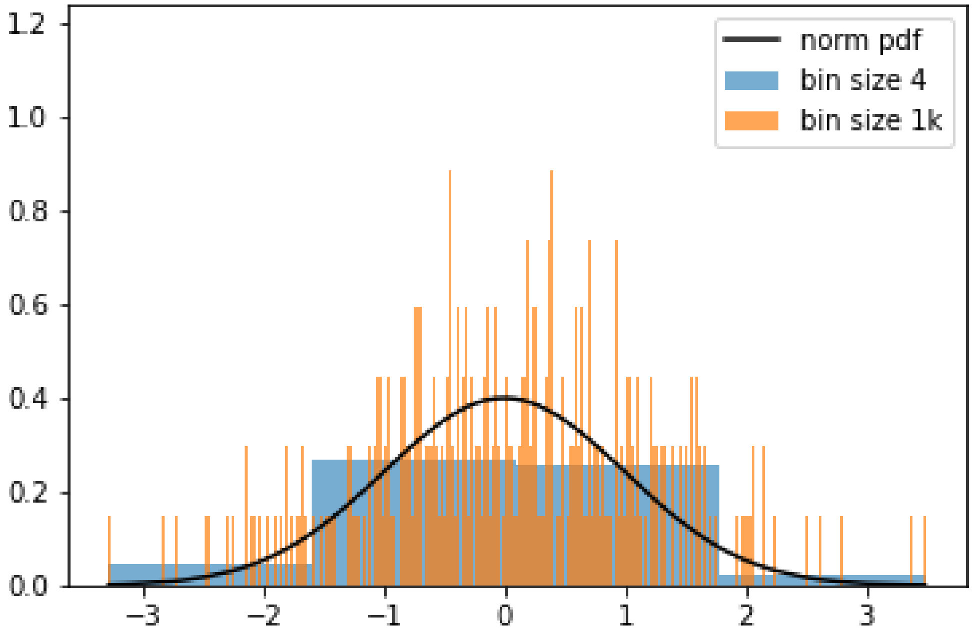
Ex. 1

a)



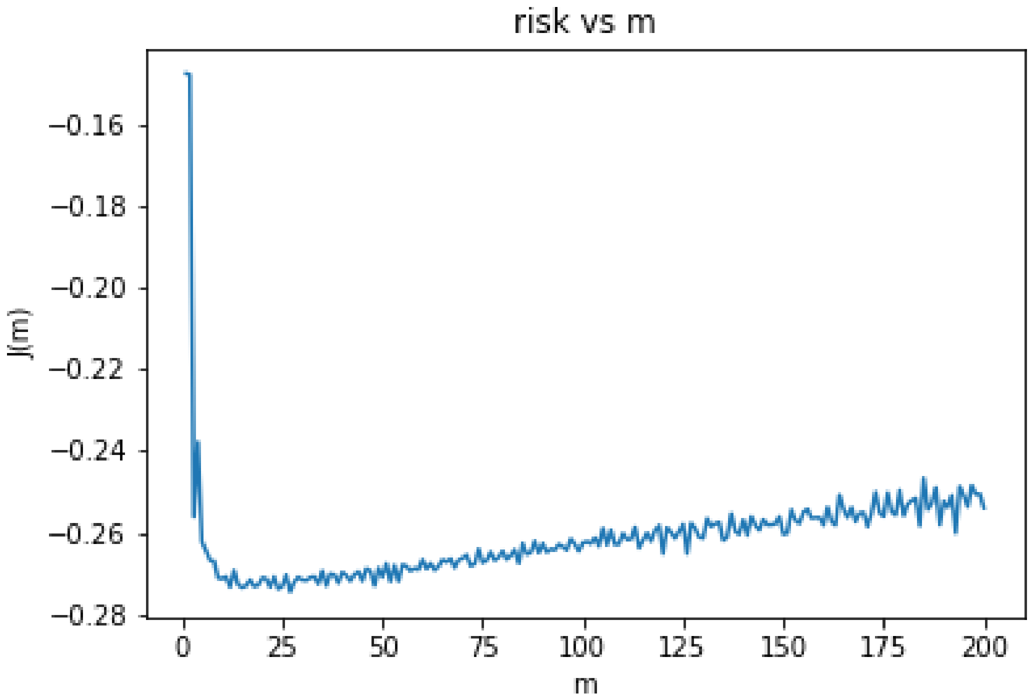
Standard Normal Distribution
f(x)
x

b)

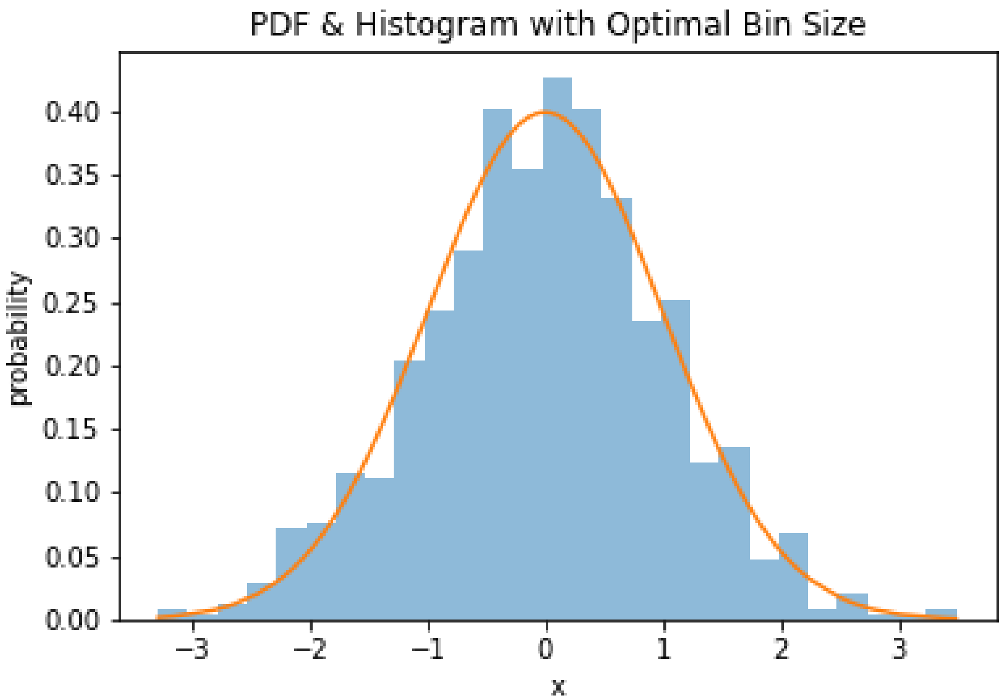


norm pdf
bin size 4
bin size 1k

c) i)


risk vs m
J(m)
−0.16
−0.18
−0.20
−0.22
−0.24
−0.26
−0.28
0
25
50
75
100
125
150
175
200
m

c) ii)


PDF & Histogram with Optimal Bin Size
probability
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−3
−2
−1
0
1
2
3
x

Ex 2.

a) $$f_X(x) = \frac{1}{\sqrt{(2\pi)^2 |\Sigma|}} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right\}.$$

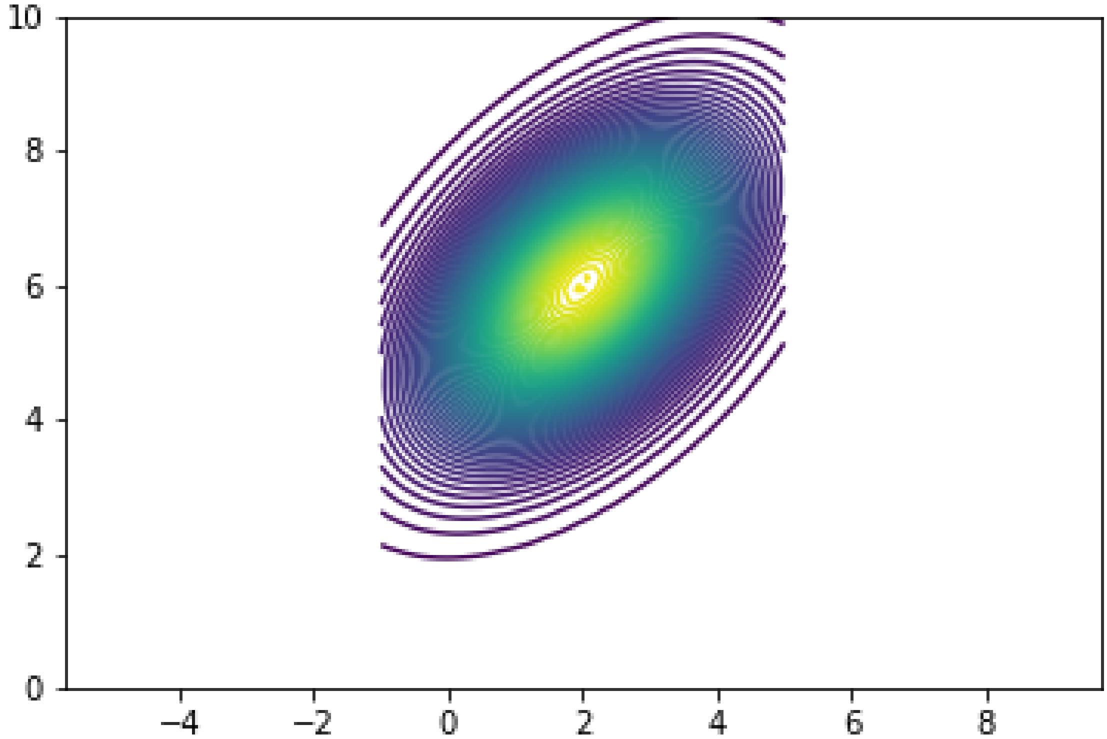
$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mu = \begin{bmatrix} 2 \\ 6 \end{bmatrix}, \quad \text{and} \quad \Sigma = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

(i) $|\Sigma| = 4 - 1 = 3$. $\quad \Sigma^{-1} = \frac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$

$$-\frac{1}{2}[x_1-2, x_2-6] \cdot \frac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \cdot \begin{bmatrix} x_1-2 \\ x_2-6 \end{bmatrix}$$

$$= -\frac{1}{6}\left(2x_1^2 + 4x_1 - 2x_1x_2 + 2x_2^2 - 20x_2 + 56\right)$$

$$\boxed{f_X(x) = \frac{1}{2\pi\sqrt{3}} \cdot \exp\left\{-\frac{1}{6}\left(2x_1^2 + 4x_1 - 2x_1x_2 + 2x_2^2 - 20x_2 + 56\right)\right\}}$$



2. b)

(ii) $A \in \mathbb{R}^{d \times d}$ $b \in \mathbb{R}^d$, $Y = AX + b$.

$\mu_Y \stackrel{def}{=} \mathbb{E}[Y]$, $\Sigma_Y \stackrel{def}{=} \mathbb{E}[(Y-\mu_Y)(Y-\mu_Y)^T]$

show $\mu_Y = b$, $\Sigma_Y = AA^T$.

$$\mathbb{E}(Y) = \mathbb{E}(Ax+b) = A\,\mathbb{E}(\cancel{X}) + b = b = \mu_Y$$

$$\begin{aligned}
\mathbb{E}[(Y-\mu_Y)\cdot(Y-\mu_Y)^T] &= \mathbb{E}((Y-b)\cdot(Y-b)^T) \\
&= \mathbb{E}[AX\cdot(AX)^T] \\
&= \mathbb{E}(AX\cdot X^T A^T) \\
&= A\cdot\mathbb{E}[X\cdot X^T]\cdot A^T \\
&= A\cdot\Sigma_X\cdot A^T \\
&= AA^T \\
&= \Sigma_Y
\end{aligned}$$

ii) $\Sigma_Y^T = (AA^T)^T = (A^T)^T \cdot A^T = AA^T = \Sigma_Y$. (symmetric)

Let q be $A^T x$, $q^T = x^T A$

$$q^T \cdot q = x^T A \cdot A^T x$$

$$= x^T \cdot \Sigma_Y x$$

Since inner product of a vector is always ≥ 0,

$x^T \Sigma_Y x \geq 0$. & Σ_Y is positive semi-definite

iii) For $x^T \Sigma_Y x > 0$.

Σ_Y must be invertible.

$\Rightarrow AA^T$ is invertible

A has independent columns.

iv) $Y \sim N(\mu_Y, \Sigma_Y)$.

$\mu_Y = \begin{bmatrix} 2 \\ 6 \end{bmatrix}$. $\Sigma_Y = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$.

$\boxed{b = \begin{bmatrix} 2 \\ 6 \end{bmatrix}}$

$AA^T = U \Lambda U^-$ $\qquad \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$

$= U \sqrt{\Lambda} \cdot \sqrt{\Lambda} U^{-1}$

$\Rightarrow A = U \sqrt{\Lambda}$.

$\left[\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} - \lambda I\right] = 0$

$\begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix} = 0$

$(2-\lambda)^2 - 1 = 0$

$2-\lambda = \pm 1$

$\lambda = 3, 1$.

$\lambda = 1$. $\qquad \lambda = 3$.

$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} u_1 = 0$. $\qquad \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} u_2 = 0$.

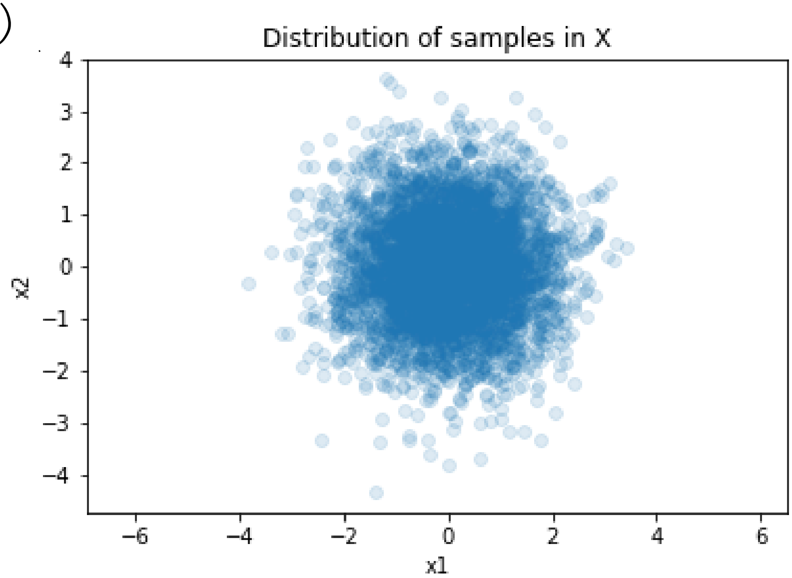
$u_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ $\qquad u_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$. $\sqrt{\Lambda} = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{3} \end{bmatrix}$

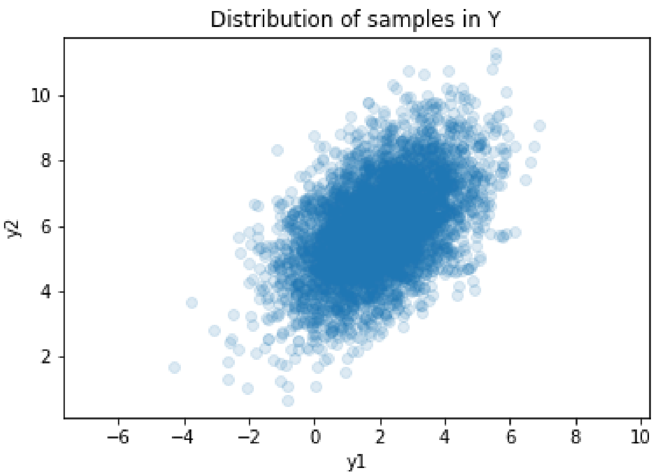
$A = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{3} \end{bmatrix}$

$= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & \sqrt{3} \\ -1 & \sqrt{3} \end{bmatrix} = \boxed{\begin{bmatrix} \frac{1}{\sqrt{2}} & \sqrt{\frac{3}{2}} \\ -\frac{1}{\sqrt{2}} & \sqrt{\frac{3}{2}} \end{bmatrix}}$

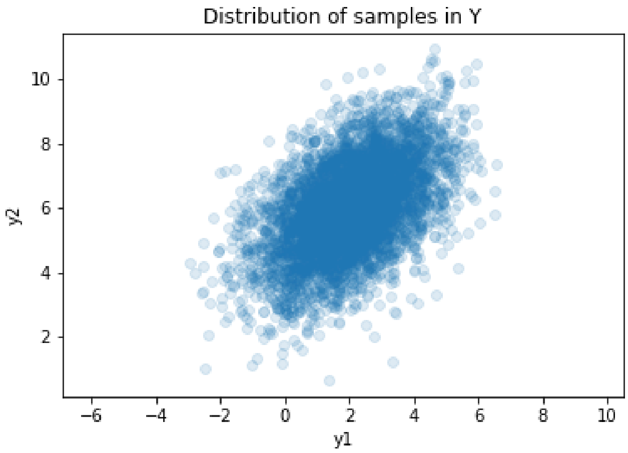
2 c) i)



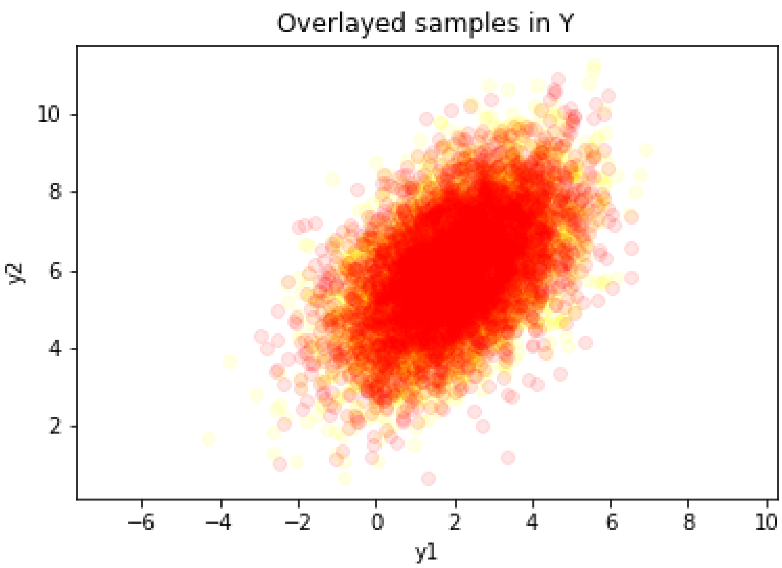
(ii)



A obtained by hand



A obtained from Python.



Two scatter plot overlayed on each other.

Exercise 3. $y = \beta_0 + \beta_1 L_1(x) + \beta_2 L_2(x) + \cdots + \beta_p L_p(x) + \epsilon$

b) $\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \|y - X\beta\|^2$

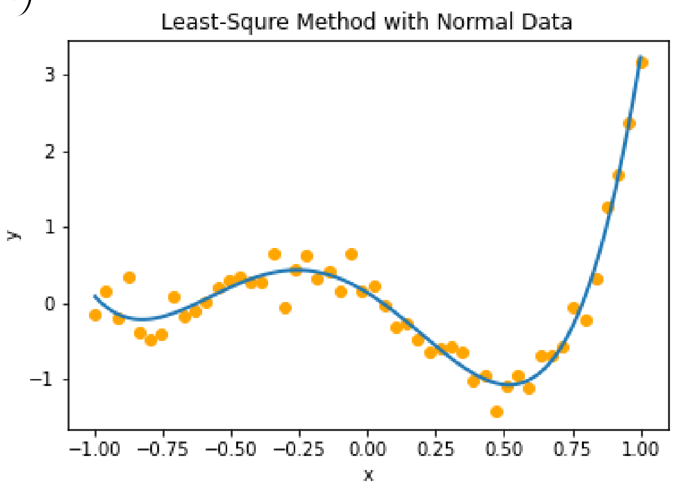
$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \quad y = \begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(N)} \end{bmatrix}, \quad X = \begin{bmatrix} x_{(0)}^0 & \cdots & x_{(0)}^p \\ & \ddots & \\ x_{(N)}^0 & \cdots & x_{(N)}^p \end{bmatrix}$$

using least-squared method,

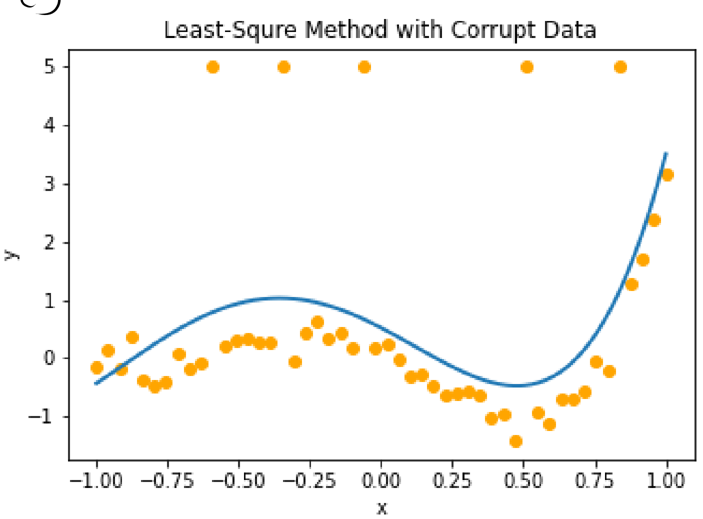
$$\hat{\beta} = (X^T X)^{-1} X^T \cdot y$$

d). The least-square method is significantly affected by the outliers, and the optimization result is drawn to the outliers.

a)



c)



e) $\hat{\beta} = \underset{\beta}{\arg\min} \|y - X\beta\|_1$

expressed in $\boxed{\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax < b. \end{array}}$

The original problem is equivalent to

$$\underset{\beta, \{u_n\}}{\text{minimize}} \sum_{n=1}^{N} u_n$$

$$y_n = [x_{(n)}^0 \; x_{(n)}^1 \cdots x_{(n)}^p] \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \end{bmatrix}$$

s.t. $u_n = |y_n - \phi_n^T \beta|$, where $\phi_n = \begin{bmatrix} x_{(n)}^0 \\ \vdots \\ x_{(n)}^p \end{bmatrix}$.

$\Downarrow$.

$$\min_{\beta, \{u_n\}} \sum_{n=1}^{N} u_n$$

$$\sum_{n=1}^{N} u_n = \underbrace{[\underbrace{0 \cdots 0}_{p} \underbrace{1 \cdots 1}_{N}]}_{c^T} \underbrace{\begin{bmatrix} \beta \\ \cdots \\ \{u_n\} \end{bmatrix}}_{x}.$$

s.t. $u_n \geq -(y_n - \phi_n^T \beta)$

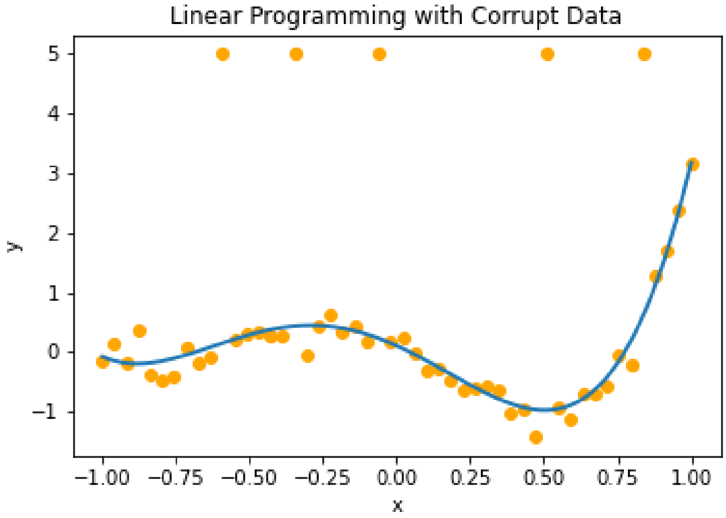
$u_n \geq (y_n - \phi_n^T \beta)$

with c & x defined, it is easy to find A & b by rearranging the constraint

$$\begin{cases} \phi_n^T \beta - u_n \leq y_n \\ -\phi_n^T \beta - u_n \leq -y_n \end{cases}$$

$$A = \begin{bmatrix} \phi_1^T & \\ \vdots & -I \\ \phi_n^T & \\ \hline -\phi_1^T & \\ \vdots & -I \\ -\phi_n^T & \end{bmatrix}, \quad b = \begin{bmatrix} \{y_n\} \\ \cdots \\ -\{y_n\} \end{bmatrix}$$

f)



Linear Programming with Corrupt Data
y
x
5
4
3
2
1
0
−1
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00