HW1

Ruijie Song

Feb.4.2021

Exercise 1: Histogram and Cross-Validation

a)

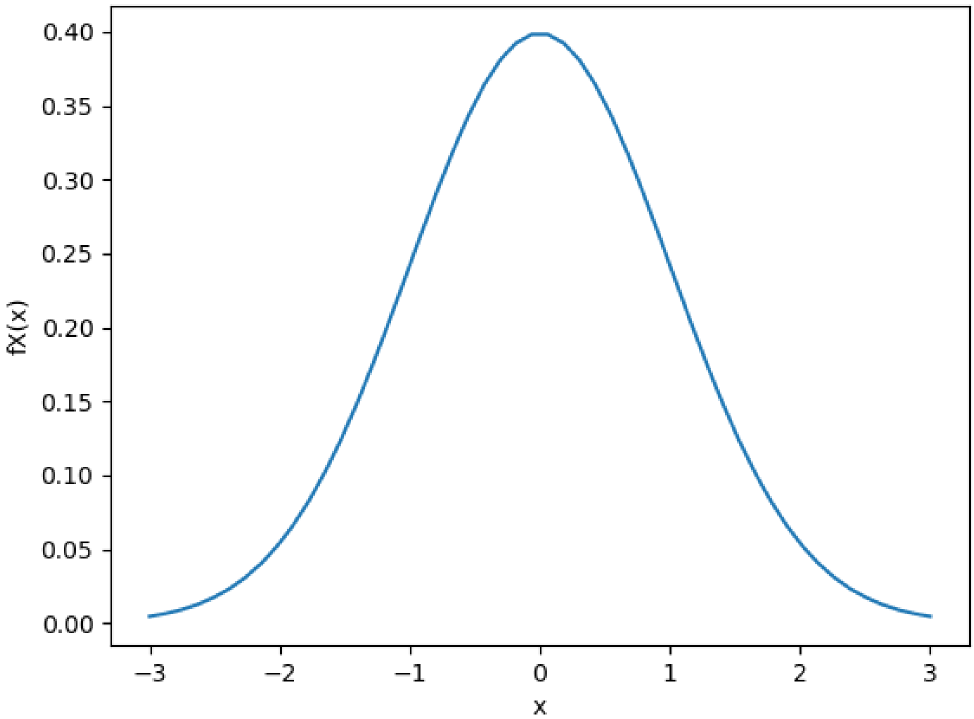


Figure 1. fX

b)
ii)

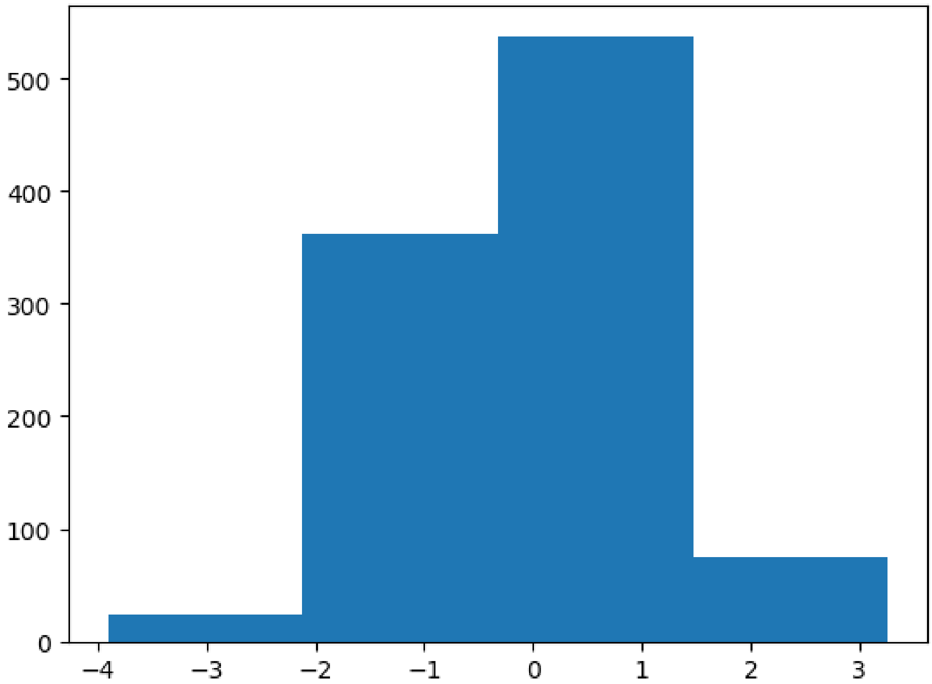


Figure 2. m=4

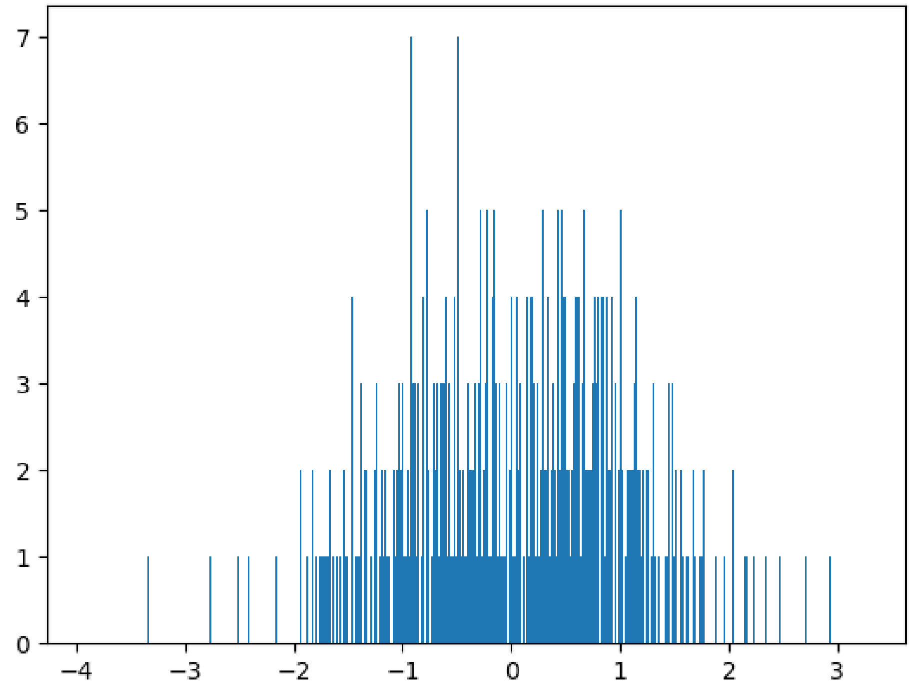


Figure 3. m=1000

iii)
mean = -0.021858244152191
standard deviation = 1.0469359084378456

iv)

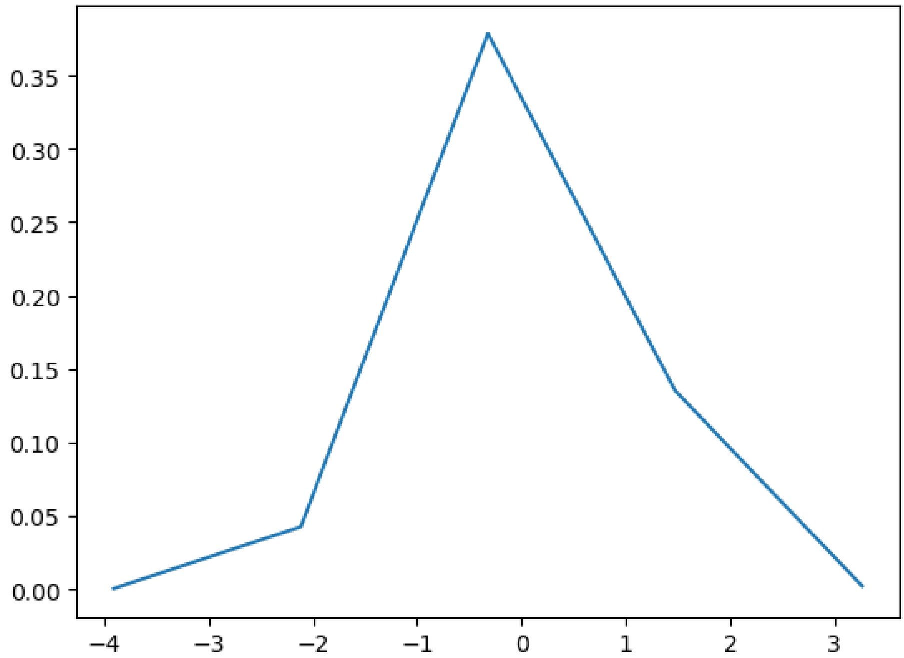


Figure 4. m=4

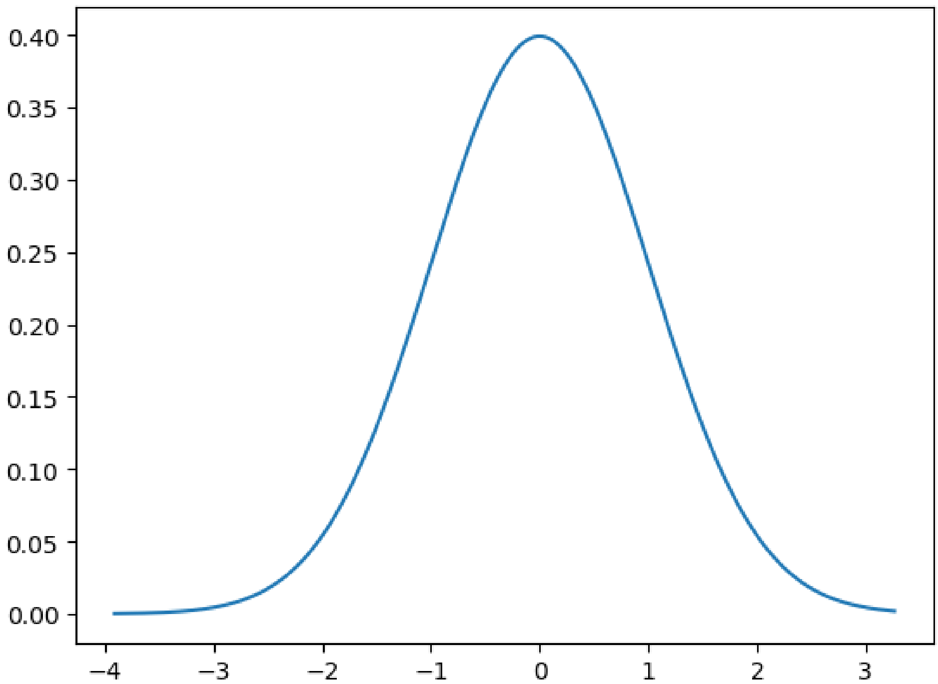


Figure 5. m=1000

c)
i)

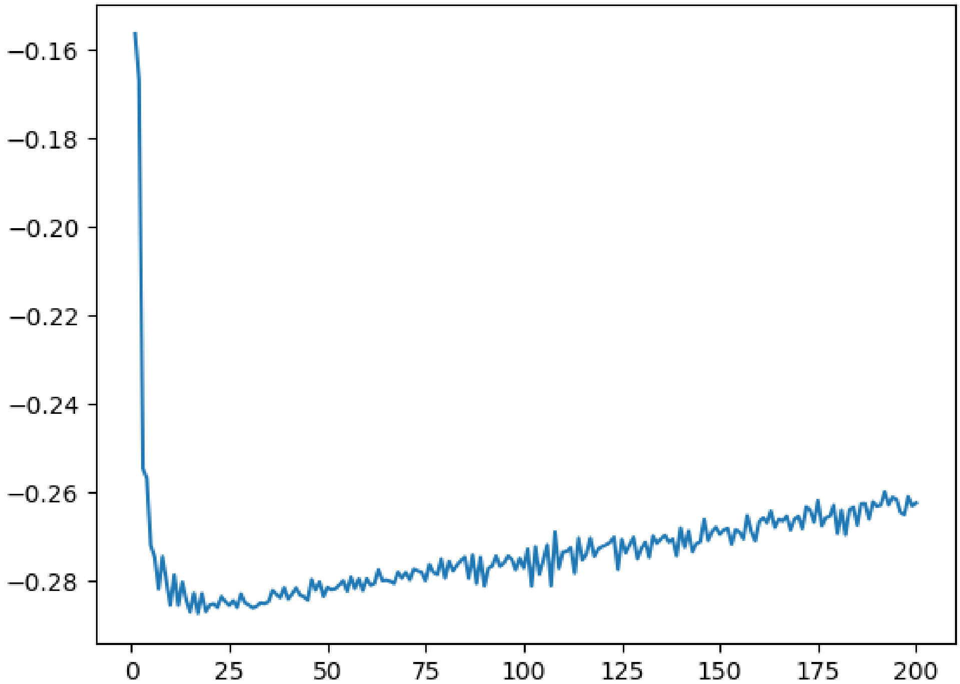


Figure 6. J(h)

ii)

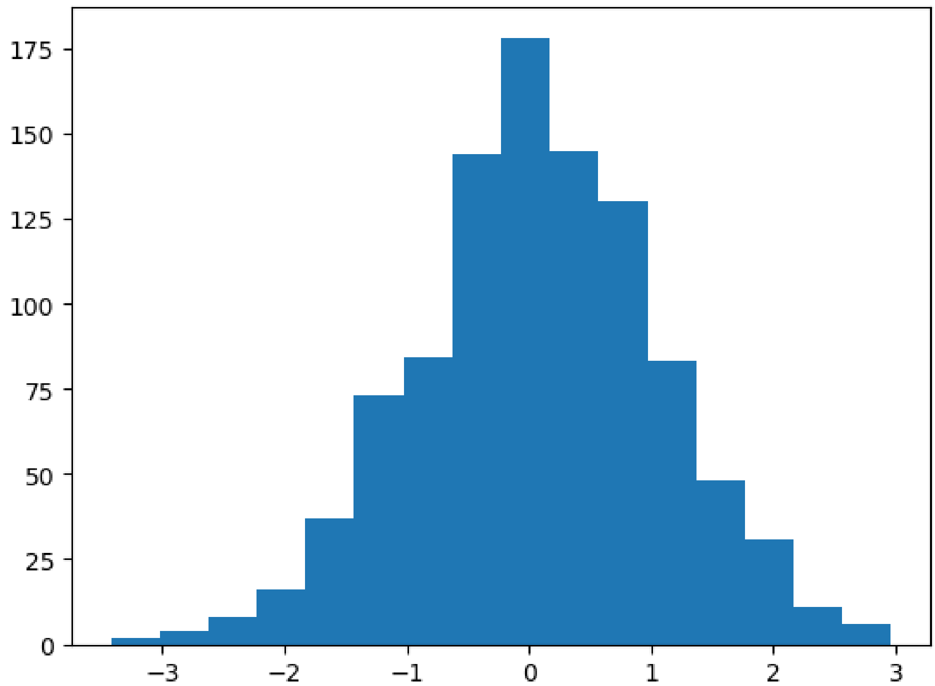


Figure 7. histogram of your data with that m∗

m* = 16

iii)

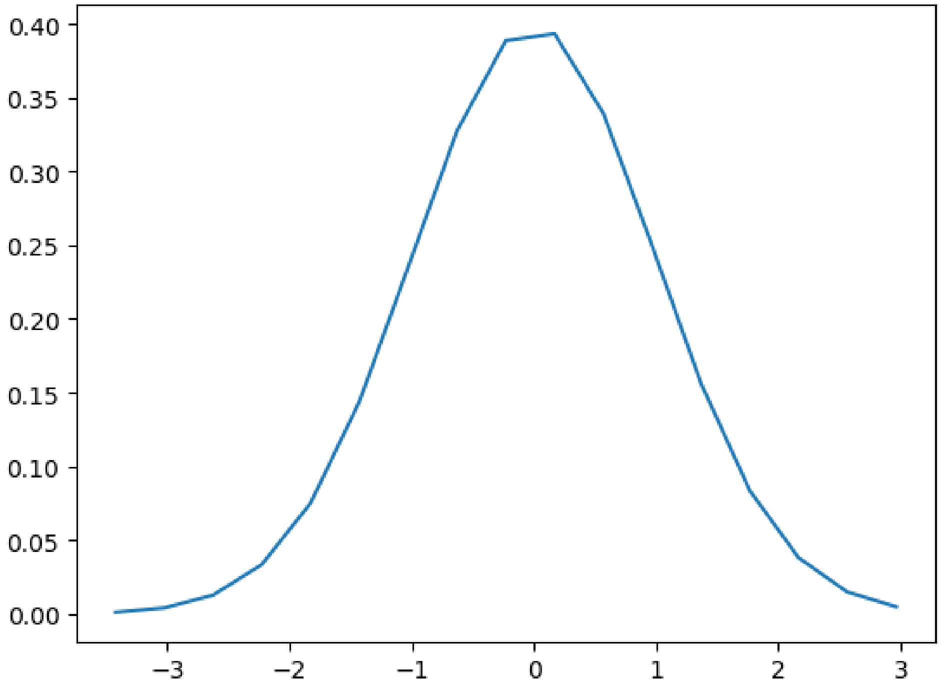


Figure 8. Gaussian curve

Exercise 2: Gaussian Whitening

a)

i)

2.

a) $X \sim N(\mu, \Sigma)$ $\quad f_X(x) = \frac{1}{\sqrt{(2\pi)^2 |\Sigma|}} \exp\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\}$

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

i) $|\Sigma| = 4 - 1 = 3$ $\quad x - \mu = \begin{bmatrix} x_1 - 2 \\ x_2 - 6 \end{bmatrix}$

$\Sigma^{-1} = \begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix}$

$$\therefore f_X(x) = \frac{1}{\sqrt{(2\pi)^2 3}} \exp\left\{-\frac{1}{2}[x_1-2, x_2-6]\begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix}\begin{bmatrix} x_1-2 \\ x_2-6 \end{bmatrix}\right\}$$

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

ii)

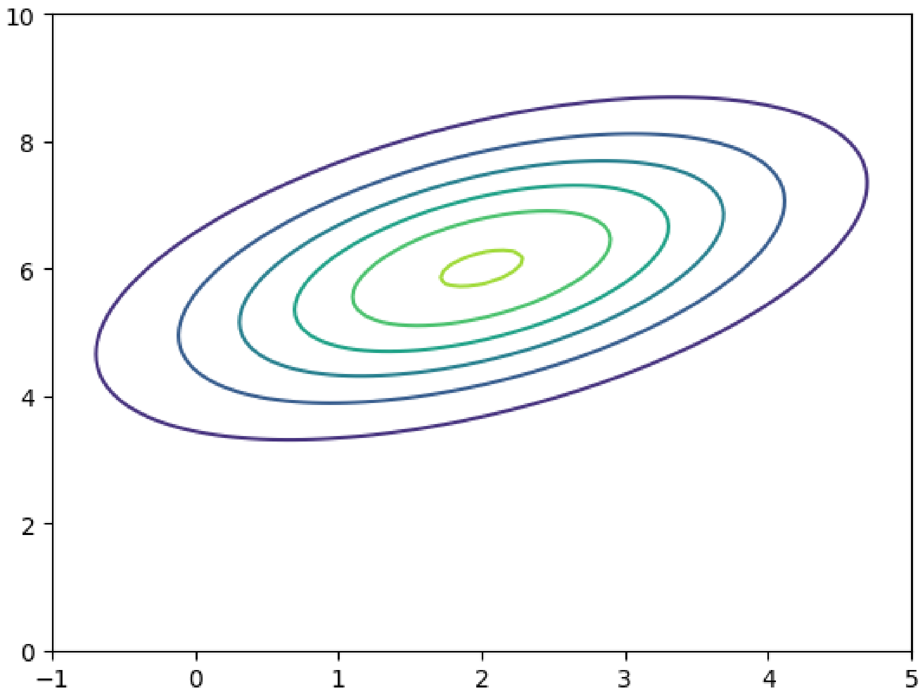


Figure 9. The contour of fX

b)

b) $A \in \mathbb{R}^{d\times d}$ $\quad b \in \mathbb{R}^d$ $\quad Y = A\bar{X} + b$ $\quad \bar{X}$: r.v $\sim N(0, I)$

$\mu_Y \overset{def}{=} E[Y]$ $\quad \Sigma_Y \overset{def}{=} E[(Y-\mu_Y)(Y-\mu_Y)^T]$

i) $E[Y] = E[A\bar{X} + b] = A \cdot \underbrace{E[\bar{X}]}_{0} + b = b = \mu_Y$

$$\Sigma_Y = E[(A\bar{X} + \cancel{b} - \cancel{b})(A\bar{X} + \cancel{b} - \cancel{b})^T] = E[(A\bar{X})(A\bar{X})^T]$$

$$= E(A\bar{X}\bar{X}^T A^T) = A\, E[\bar{X}\bar{X}^T]\, A^T = AA^T$$

ii) symmetric if $\Sigma = \Sigma^T$, positive semi-definite if $x^T \Sigma x \geq 0$ for any $x \in \mathbb{R}^n$

~~AA^T~~ $\Sigma_Y = AA^T$ must be symmetric since $A \in \mathbb{R}^{d\times d}$

Let x be a non-zero vector

$$x^T \Sigma x = x^T A^T A x = (Ax)^T (Ax) = \|Ax\|^2 \geq 0.$$

$\therefore$ Σ is symmetric positive semi-definite when $A \in \mathbb{R}^{d\times d}$

iii) $x^T \Sigma x > 0$ for $x \in \mathbb{R}^n$ when $\|Ax\| \neq 0$

iv) $Y \sim N(\mu_Y, \Sigma_Y)$ $\quad \mu_Y = \begin{bmatrix} 2 \\ 6 \end{bmatrix}$, $\Sigma_Y = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$

$b = \mu_Y = \begin{bmatrix} 2 \\ 6 \end{bmatrix}$ $\quad AA^T = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$

$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ $\quad \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$

$\begin{bmatrix} 2a+c & 2b+d \\ a+2c & b+2d \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ $\quad \begin{cases} \cancel{2}a + c = \cancel{a}\ 0 \\ \cancel{2}b + d = \cancel{b} \\ a + \cancel{2}c = \cancel{c} \\ b + \cancel{2}d = \cancel{d}\ 0 \end{cases}$

$a + c = 0$ $\quad a = -c$

$b + d = 0$ $\quad b = -d$ $\quad \therefore A = \begin{bmatrix} a & b \\ -a & -b \end{bmatrix}$

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

$$\Sigma^{\frac{1}{2}} = A = U\Lambda^{\frac{1}{2}}U^T = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}$$

c)
i)

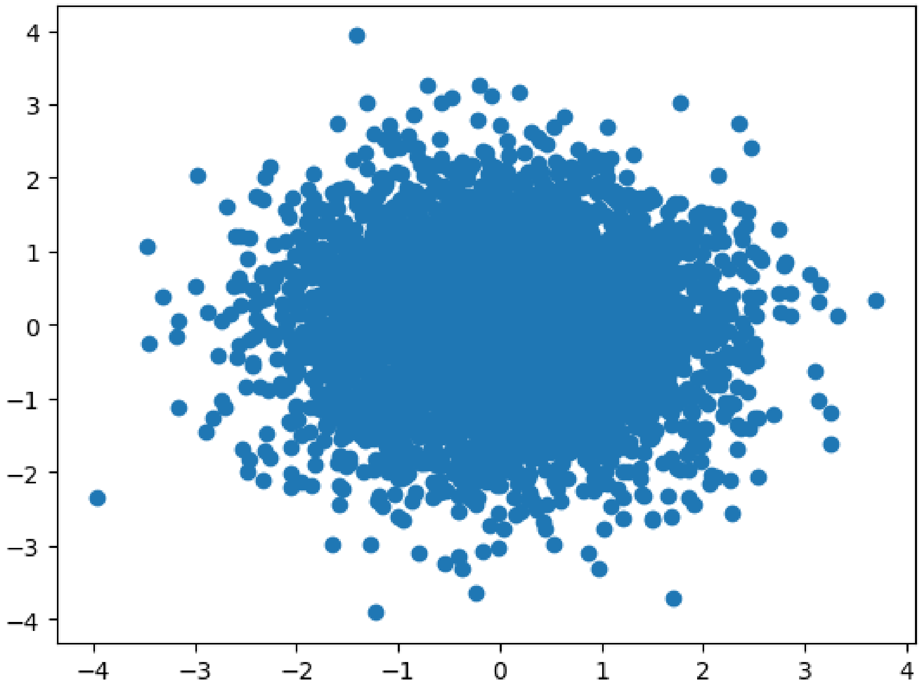


Figure 10. scatter plot

ii)

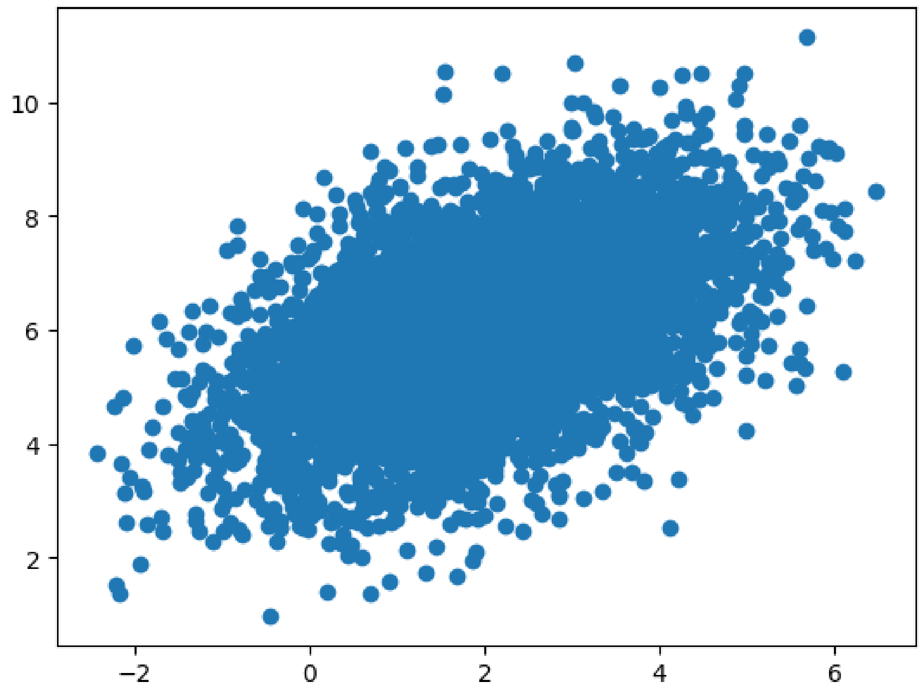


Figure 11. data calculated by numpy.random.multivariate_normal

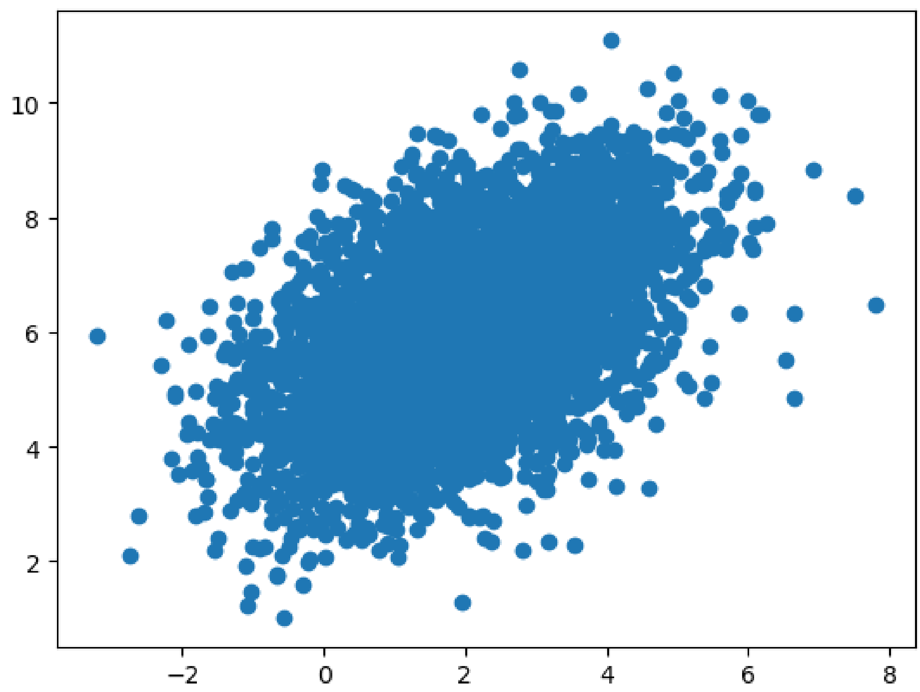
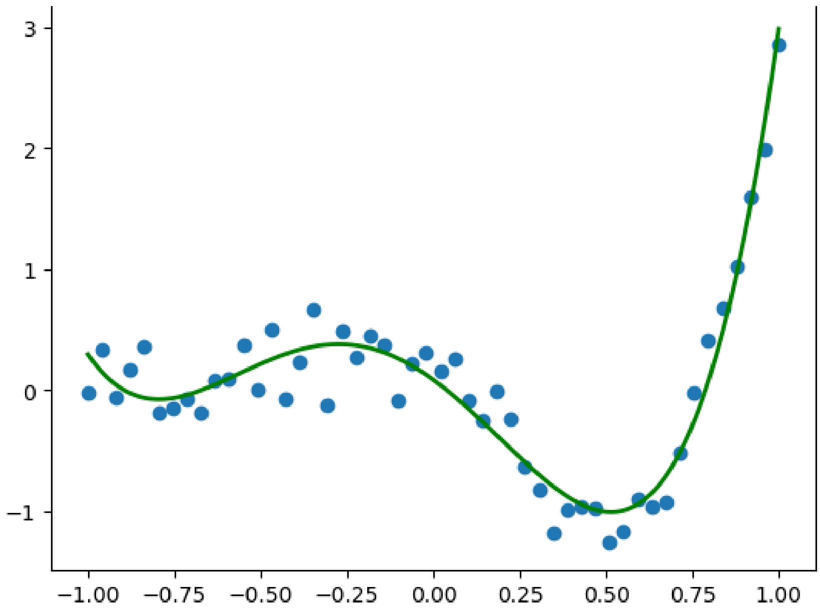


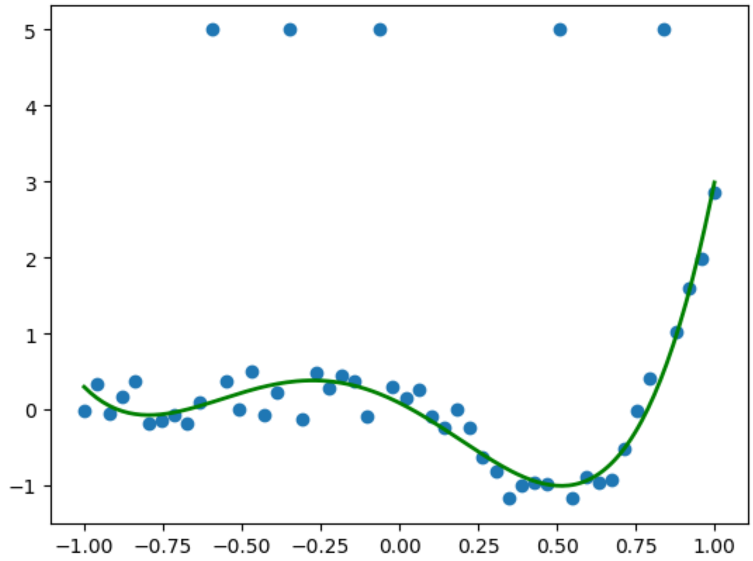
Figure 12. data calculated by affine transformation

Exercise 3: Linear Regression

a & c.



d.



There are a few differences between those 2 plot. However, the outliers do not affect much.

b & e.

3. b) $\hat{\beta} = \arg\min_{\beta} \|y - X\beta\|^2$

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

X: data

$y = g_{\beta}(X) = \beta^T X$

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

minimize$_x$ $c^T x$

s.t.: $Ax \le b$

$$\Downarrow$$

$$\min_{\beta, \cancel{\{u_n\}}} \sum_{n=1}^{N} |y - X\beta|$$

~~s.t. $u_n \ge -(y - X\beta)$~~

~~$u_n \ge y - X\beta$~~

$$\Downarrow$$

$$\min_{\beta, \{u_n\}} \sum_{n=1}^{N} u_n \quad \text{s.t.} \quad u_n = |y - X\beta|$$

$$\Downarrow$$

$$\min_{\beta, \{u_n\}} \sum_{n=1}^{N} u_n \quad \text{s.t.} \quad u_n \ge -(y - X\beta), \quad u_n \ge (y - X\beta)$$

$$\Downarrow$$

$$\min_{\beta, u} \Big(\underbrace{[0 \cdots 0, 1 \cdots 1]}_{c} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \\ u_1 \\ \vdots \\ u_N \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} X_1 & -1 & & \\ \vdots & & \ddots & \\ X_N & & & -1 \\ -X_1 & -1 & & \\ \vdots & & \ddots & \\ -X_N & & & -1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \\ u_1 \\ \vdots \\ u_N \end{bmatrix}}_{x} \le \underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_N \\ -y_1 \\ \vdots \\ -y_N \end{bmatrix}}_{b}$$

f.

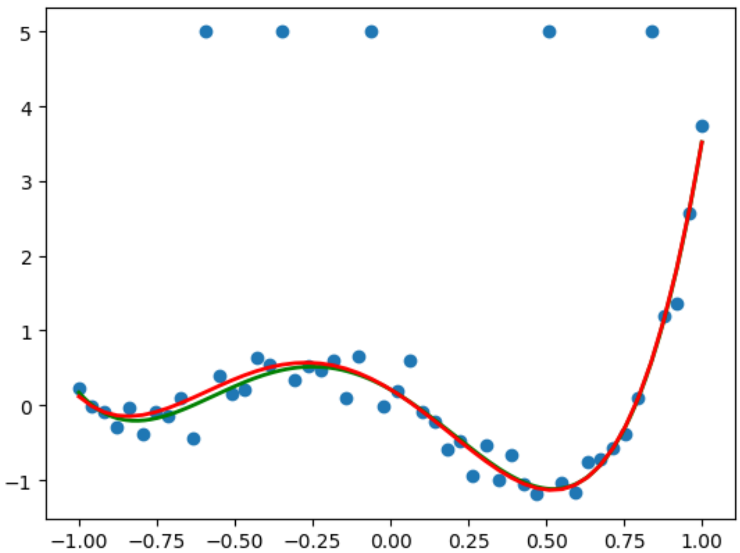


Figure 13. The green curve is for d, The red curve is for f