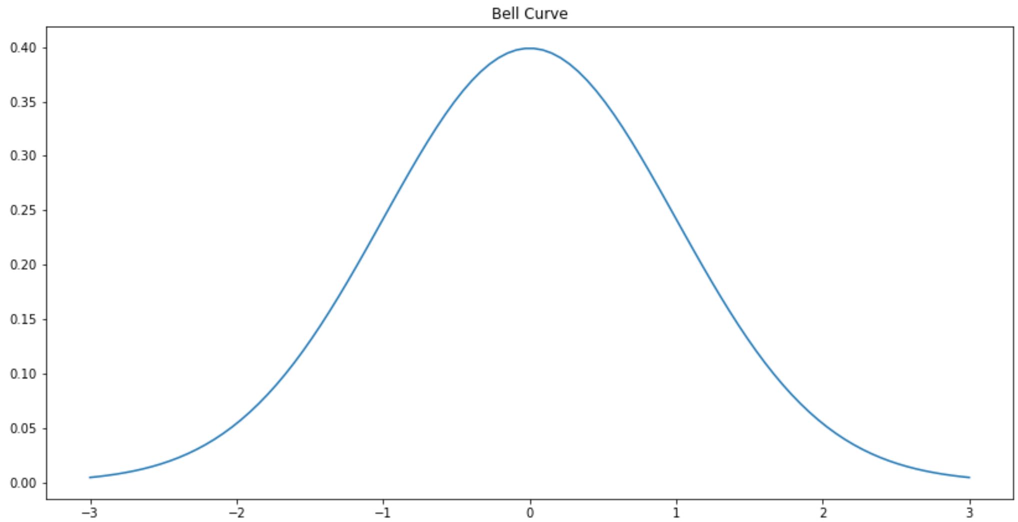
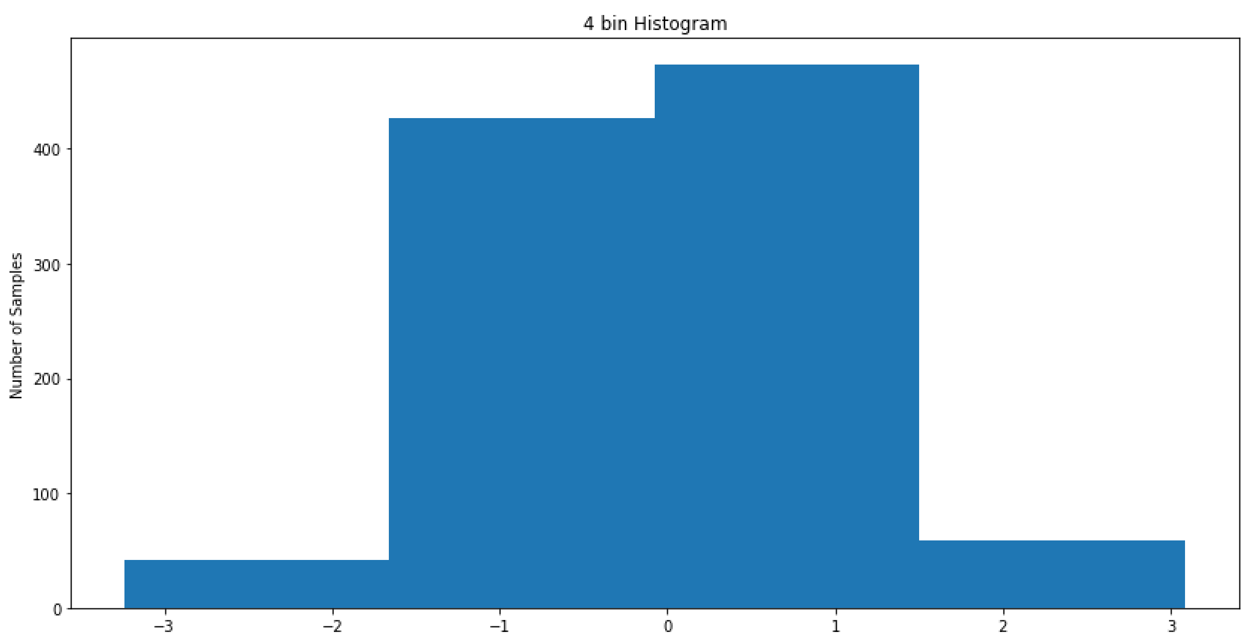
ECE 50024

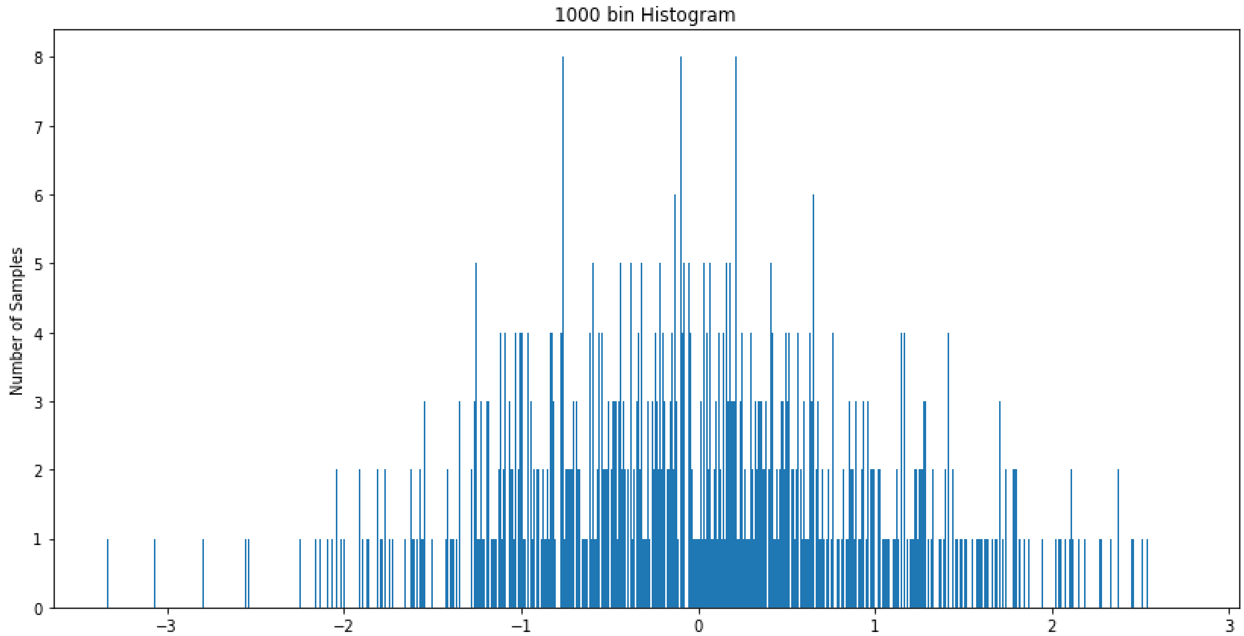
Homework 1

Parth Sagar Hasabnis phasabni@purdue.edu

Exercise 1:

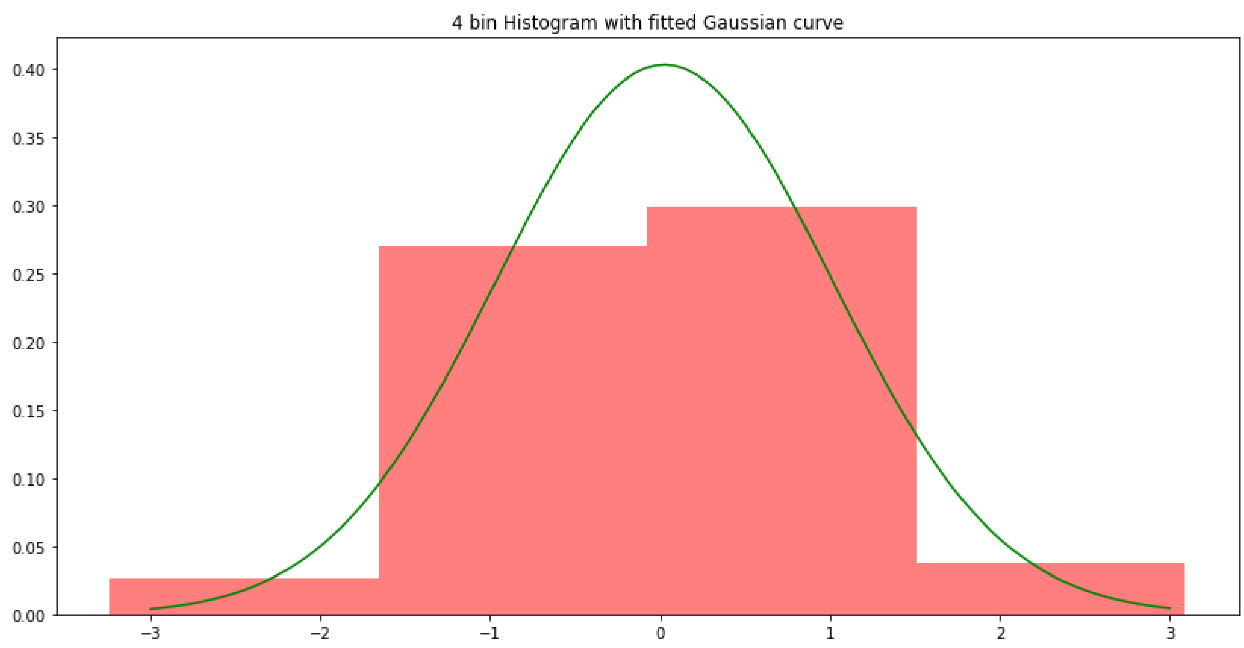


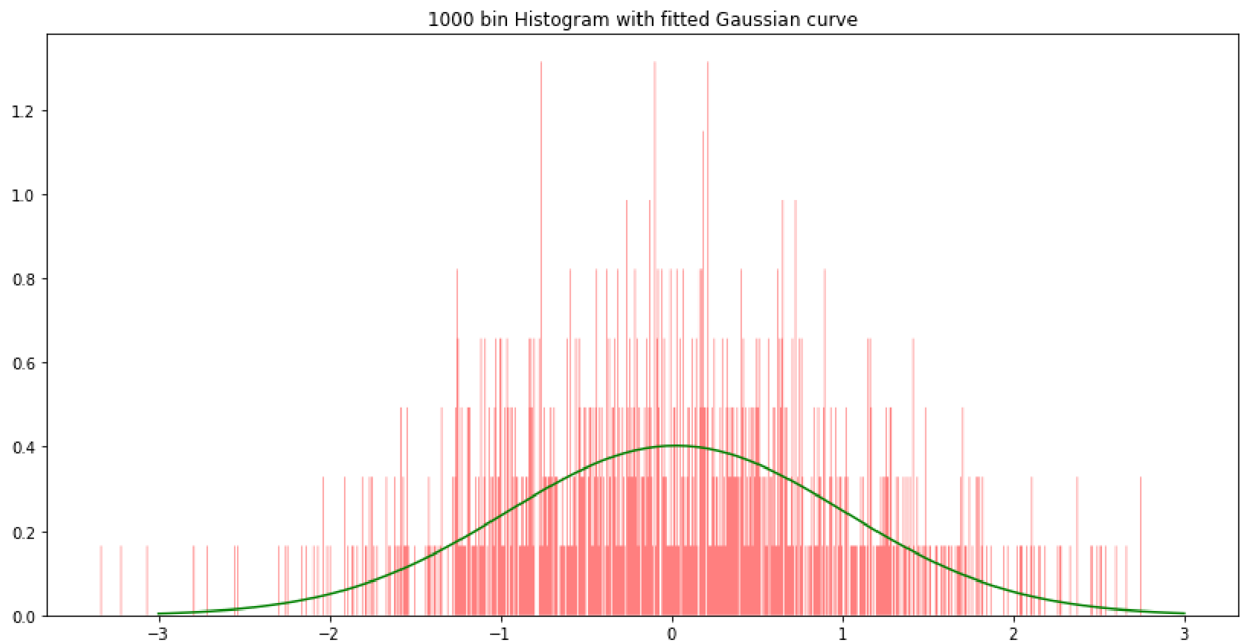




Mean of the distribution is: 0.026005414594233493.

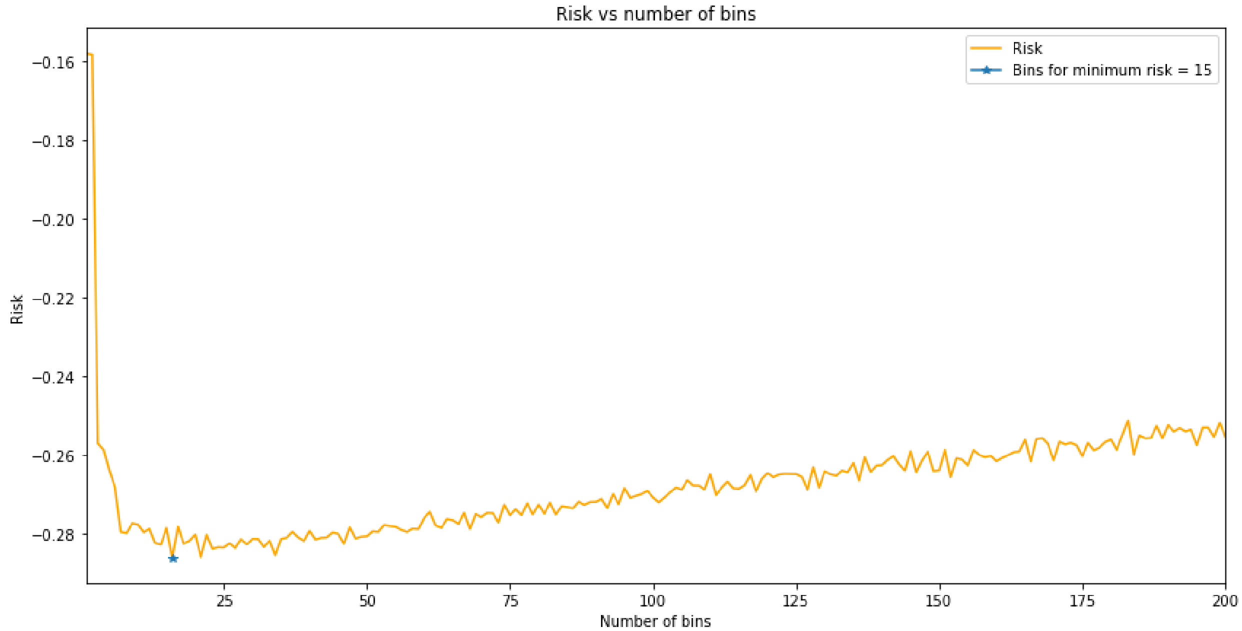
Standard Deviation of the distribution is: 0.9907009680534483.



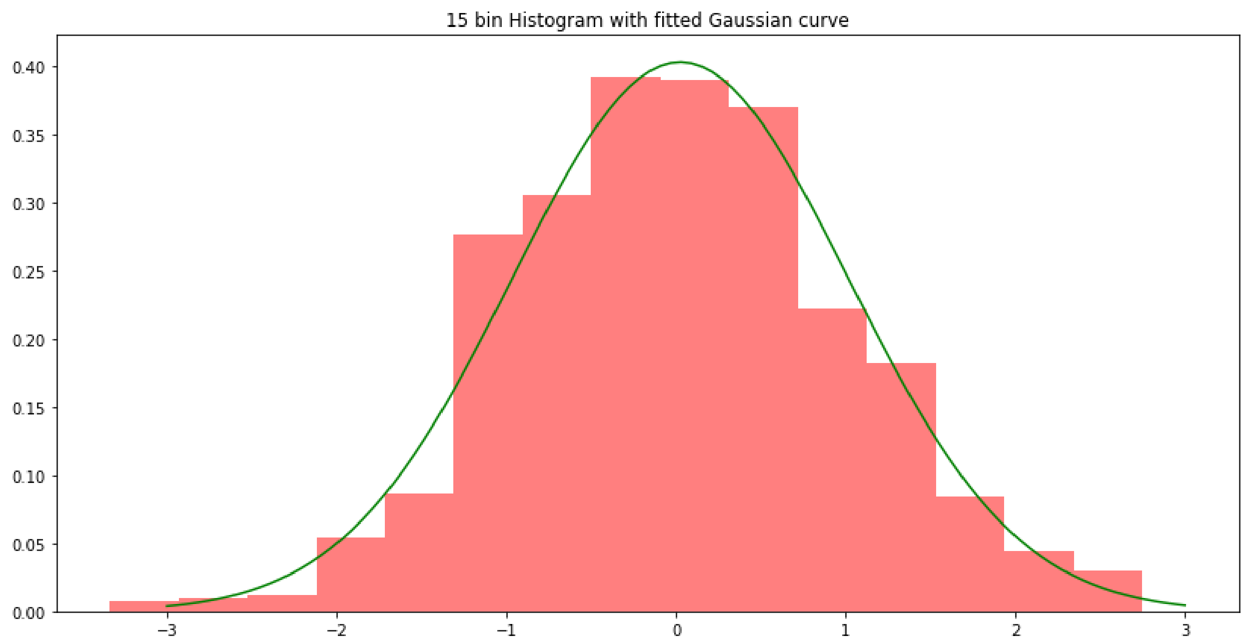


Exercise 1

b) v) The histogram with 4 bins does not represent the data very well, as the number of bins is too less. It does not provide adequate information about the mean or the variance of the data.
Likewise, 1000 is far too many bins for the given data. Even though we can obtain some information about the mean and variance, the plot appears to be noisy, and a trend in the data is not discernable



Optimal number of bins = 15



Exercise 2:

a)

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

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

$$|\Sigma| = 3 \qquad \Sigma^{-1} = \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

$$\therefore (x-\mu)^T \Sigma^{-1} (x-\mu)$$

$$= \begin{bmatrix} x_1 - 2 & x_2 - 6 \end{bmatrix} \times \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 - 2 \\ x_2 - 6 \end{bmatrix}$$

$$= \frac{1}{3} \begin{bmatrix} 2(x_1-2) - 1(x-6) & -1(x_1-2) + 2(x_2-6) \end{bmatrix} \begin{bmatrix} x_1 - 2 \\ x_2 - 6 \end{bmatrix}$$

$$= \frac{-2}{3} \left[(x_1-2)(x_2-6) - (x_1-2)^2 - (x_2-6)^2 \right]$$

$$= \frac{-2}{3} \left[-x_1^2 - x_2^2 + x_1x_2 - 2x_1 + 16x_2 - 28 \right]$$

$$\therefore f_X(x) = \frac{1}{\sqrt{(2\pi)^2 3}}$$

$$f_X(x) = \frac{1}{\sqrt{3(2\pi)^2}} \exp\left\{ \frac{1}{3} \left[\begin{array}{l} -x_1^2 - x_2^2 + x_1 x_2 \\ -2x_1 + 10x_2 - 28 \end{array} \right] \right\}$$

b)

b) ii) $X \sim N(0, I)$ $\therefore \mu_x = \vec{0}$

$$\Sigma_x = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$Y = AX + b$

$$\therefore \mathbb{E}[Y_n] = \mathbb{E}\left[\sum_{k=1}^{N} a_{nk} X_k + b_n\right]$$

$$= \sum_{k=1}^{N} a_{nk}\, \mathbb{E}[X_k] + b_n$$

$$\therefore \mu_Y = \begin{bmatrix} \mathbb{E}[Y_1] \\ \mathbb{E}[Y_2] \\ \vdots \\ \mathbb{E}[Y_n] \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{N} a_{1k}\, \mathbb{E}[X_k] + b_n \\ \sum_{k=1}^{N} a_{2k}\, \mathbb{E}[X_k] + b_2 \\ \vdots \\ \sum_{k=1}^{N} a_{Nk}\, \mathbb{E}[X_k] + b_N \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & & a_{NN} \end{bmatrix} \begin{bmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \\ \vdots \\ \mathbb{E}[X_N] \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{bmatrix}$$

$$= A\vec{0} + b$$

$$\mu_Y = b$$

$$\Sigma_y = \mathbb{E}\left[(y - \mu_y)(y - \mu_y)^T \right]$$

$$= \mathbb{E}\left[(Ax + b - b)(Ax + b - b)^T \right]$$

$$= \mathbb{E}\left[(Ax)(Ax)^T \right]$$

$$= \mathbb{E}\left[A x x^T A^T \right]$$

$$= A \, \mathbb{E}\left[x x^T \right] A^T$$

As $\mu_x = 0$ $\quad \mathbb{E}\left[x x^T \right]$

$$= \mathbb{E}\left[(x - \mu_x)(x - \mu_x)^T \right]$$

$$= \Sigma_x$$

$$\therefore \quad \Sigma_y = A \Sigma_x A^T$$

ii) Consider an element $\Sigma_{i,j}$ of Σ_Y

$$\Sigma_{i,j} = Cov(Y_i, Y_j), \quad \Sigma_{j,i} = Cov(Y_j, Y_i)$$

As covariance is commutative

$$Cov(Y_i, Y_j) = Cov(Y_j, Y_i)$$

$$\therefore \Sigma_{i,j} = \Sigma_{j,i}$$

$\therefore \Sigma_Y$ is a symmetric matrix

For a matrix to be positive semi-definite,

$$v^T \Sigma_Y v \geq 0 \quad \forall v \in R^n$$

$$\cancel{v^T \Sigma_Y v = v^T \mathbb{E}[}$$

$$\begin{aligned} v^T \Sigma_Y v &= v^T A A^T v \\ &= (v^T A)(v^T A)^T \\ &= (v^T A) \cdot (v^T A) \\ &= \|v^T A\|_2 \end{aligned}$$

$\therefore$ The L_2 norm of a matrix is always greather than or equal to 0.

$\therefore \Sigma_Y$ is a symmetric positive semi definite matrix

iii) For Σ_Y to be positive definite:

$$v^T \Sigma_Y v > 0 \quad \forall\, x \in \mathbb{R}^n - \{\vec{0}_{N\times 1}\}$$

$\therefore$ $v^T \Sigma_Y v = 0$ implies $v = \vec{0}_{N\times 1}$

$$\therefore\ v^T (AA^T) v = 0$$
$$(v^T A)(A^T v) = 0$$
$$(v^T A)(v^T A)^T = 0$$
$$(v^T A) \cdot (v^T A) = 0$$
$$\|v^T A\|_2 = 0, \quad v^T$$

For the L_2 norm of $v^T A$ to be zero when $v = \vec{0}$, A needs to have full rank

$\therefore$ For Σ_Y to be positive definite, A needs to have full rank, which implies that A needs to be invertible

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

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

$Y = AX + b \qquad X = N(0, I)$

We know that $\mu_Y = b$

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

Consider the eigen decomposition of Σ_Y

$$\Sigma_Y - \lambda I = 0$$

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

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

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

$$\lambda = 1, 3$$

$$(A - \lambda_i) v_i = 0$$

$$\lambda_1 = 1 \quad \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} = 0$$

$$v_{11} + v_{12} = 0 \qquad v_1 = -v_2$$

$$\therefore v_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$\lambda_2 = 3 \quad \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} v_{21} \\ v_{22} \end{bmatrix} = 0$$

$$-v_{21} + v_{22} = 0 \qquad v_{21} = v_{22}$$

$$\therefore v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$\therefore$ The matrix Σ_Y has eigen values $1, 3$ corresponding to eigenvectors $\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Σ_Y can be written as:

$$\Sigma_Y = V \Lambda V^{-1}$$

where $V = \begin{bmatrix} | & | \\ v_1 & v_2 \\ | & | \end{bmatrix}$

V is a matrix whose coloms give the eigen vectors of Σ_Y

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

Λ is a diagonal matrix whose elements are eigenvalues of Σ_Y

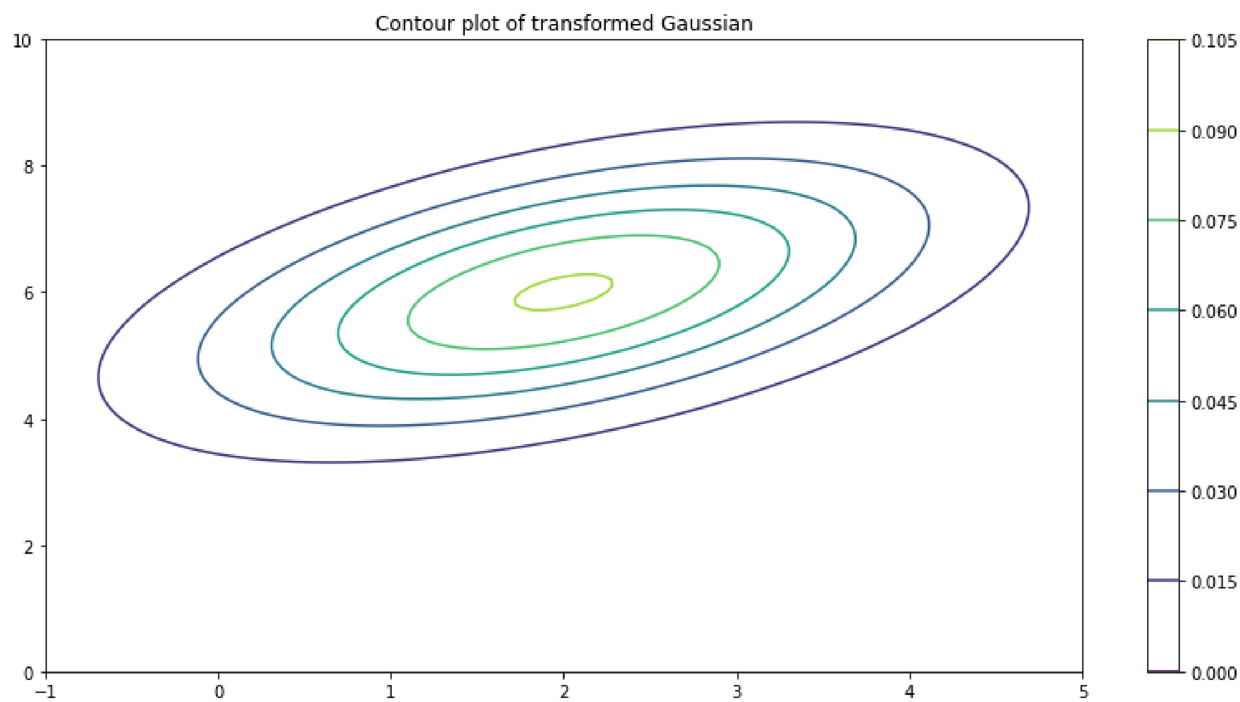
$$\Sigma_Y = V \Lambda V^{-1}$$
$$= A \cdot A^T$$

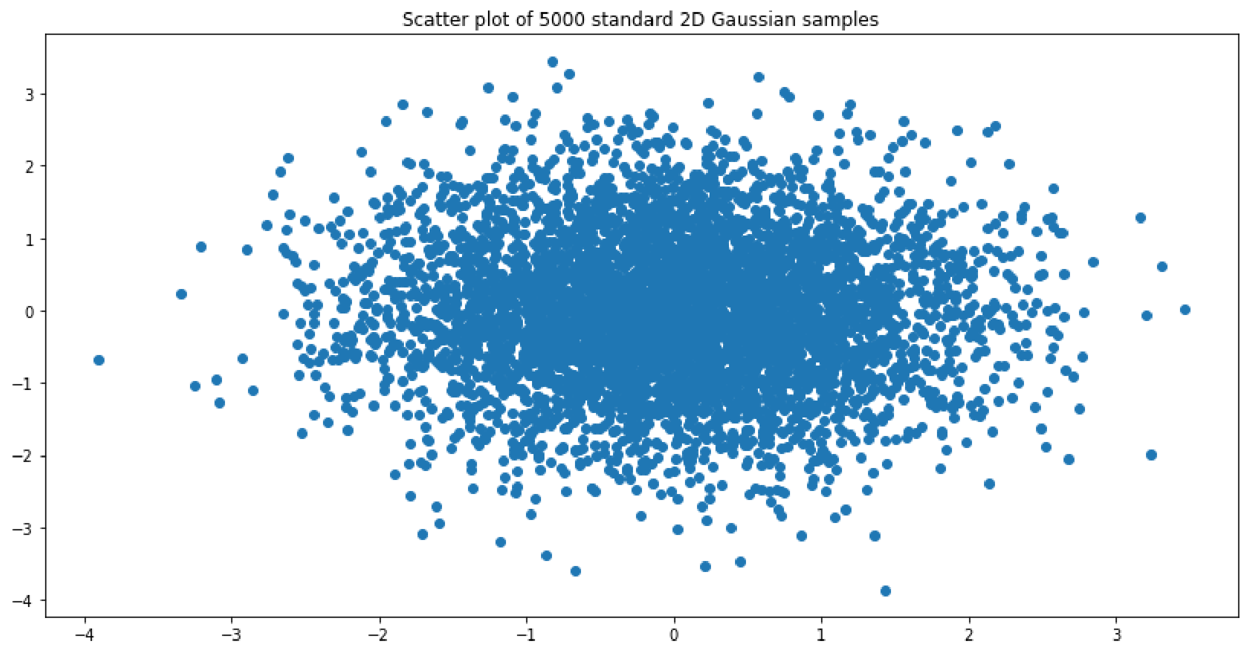
$$\therefore A = V \Lambda^{1/2} V^{-1}$$

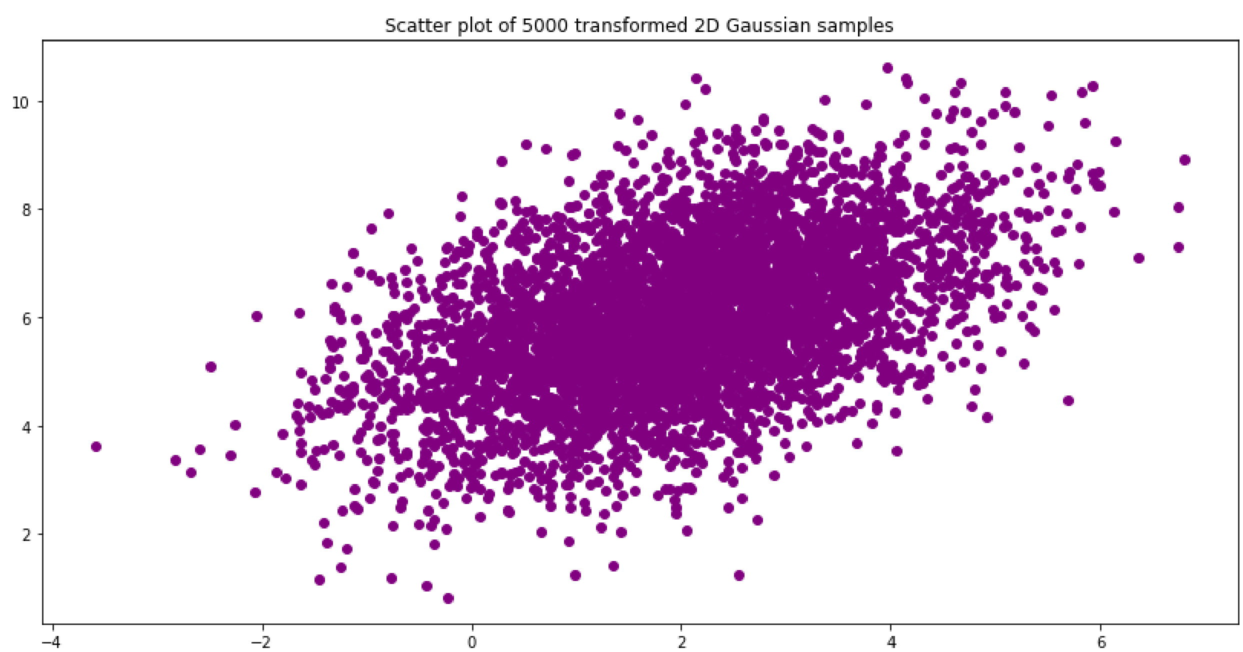
$$= \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{bmatrix} \begin{bmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

$$A = \begin{bmatrix} 1.366 & 0.366 \\ 0.366 & 1.366 \end{bmatrix}$$

c)



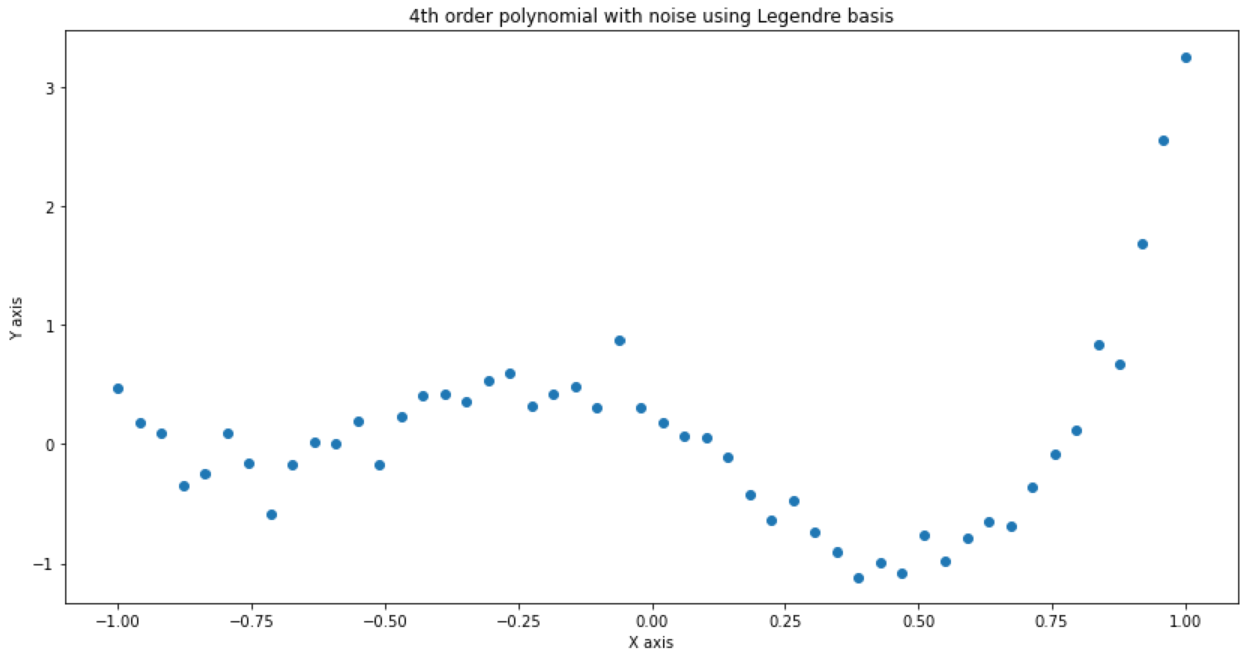




The results from parts (c)(i) and (ii) are in line with the theoritical findings. The plot of (c)(ii) represents a 2D transformed gaussian quite effectively. Similarly, the A matrix calculated from the transformed samples matches with its hand-calculated counterpart.

Exercise 3:

a) Scatter plot of data



b)

Exercise 3

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

y is the column vector whose elements are given using the following equation

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{50\times 1}$$

$y_n = \beta_0 + \beta_1 L_1(x_n) + \beta_2 L_2(x_n) + \beta_3 L_3(x_n) \cancel{+ \beta_4 (x_n} + \beta_4 L_4(x_n) + \epsilon$

$L_p(x_n)$ is the p^{th} legendre polynomial

ϵ is the noise term

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}_{50\times 1}$$

x is a vector of 50 equidistant points in the interval $[-1, 1]$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}_{5\times 1}$$

β is a column vector consisting of multiplicative coefficients

$$\beta = \begin{bmatrix} -0.001 \\ 0.01 \\ 0.55 \\ 1.5 \\ 1.2 \end{bmatrix}$$

X is the data matrix whose i^{th} row is given by the i^{th} Legendre polynomial calculated over x

$$X = \begin{bmatrix} L_0(x_1) & L_1(x_1) & \cdots & L_4(x_1) \\ L_0(x_2) & L_1(x_2) & \cdots & L_4(x_2) \\ \vdots & \vdots & & \vdots \\ L_0(x_{50}) & L_1(x_{50}) & & L_4(x_{50}) \end{bmatrix}_{50\times 5}$$

The training loss for this linear regression problem is given by:

$$\mathcal{E}_{train}(\beta) = ||y - X\beta||^2$$

taking the gradient with respect to β

$$\nabla_\beta \mathcal{E}_{train}(\beta) = \nabla_\beta \left\{ ||y - X\beta||^2 \right\}$$

$$= -2X^T(y - X\beta)$$

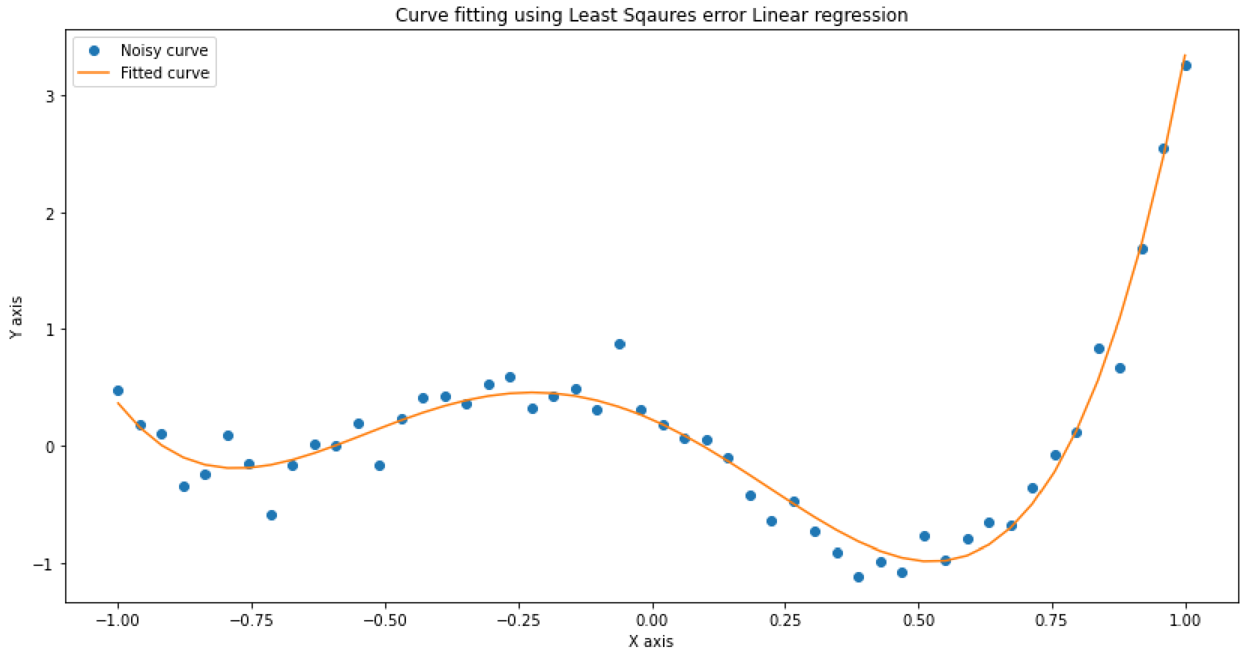
equating it to zero,

$$X^T(y - X\beta) = 0$$

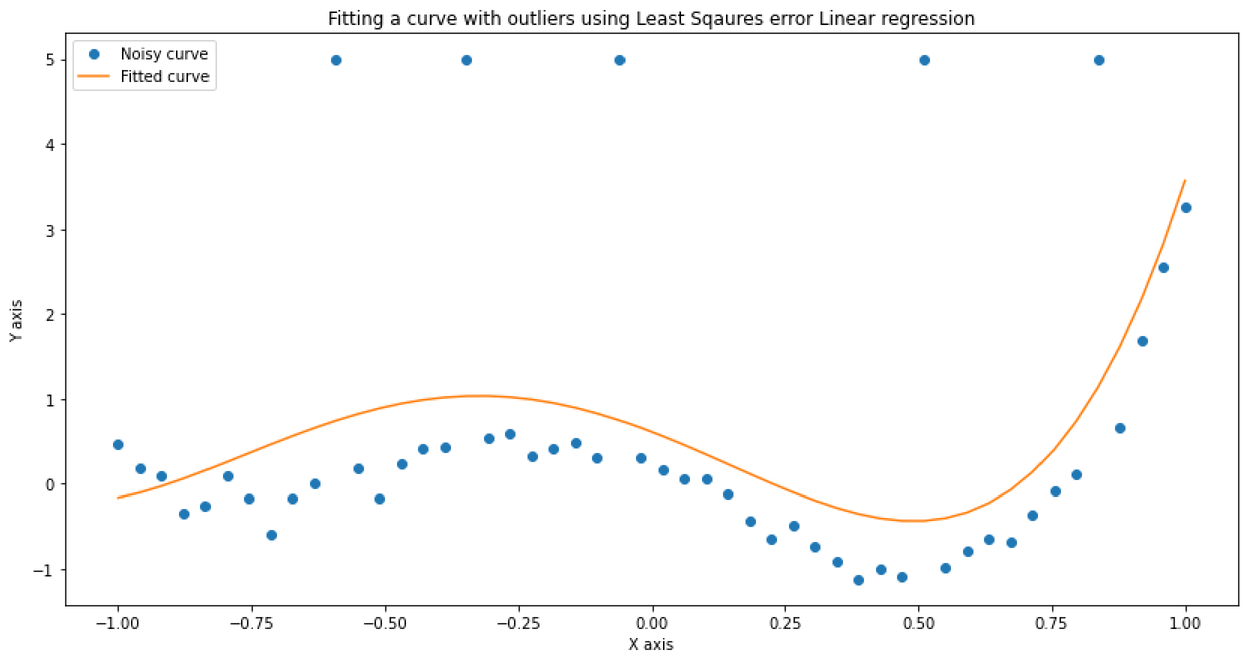
$$X^T X \beta = X^T Y$$

$$\therefore \quad \beta = (X^T X)^{-1} X^T Y$$

c) Least squares solution



d) Presence of Outliers



Observation: The existence of outliers has severely impacted the performance of Least Square regression. The fitted curve has experienced a positive offset, and no longer is representative of the data.

e) Linear Programming Problem

e) We consider the following optimization

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \|y - X\beta\|_1$$

$$\underset{\beta \in R^5}{\text{minimize}} \sum_{n=1}^{50} |y_n - x_n^T \beta|$$

we denote $U_n = |y_n - x_n^T \beta|$

$\therefore$ The problem beomes

$$\underset{\beta \in R^5, U \in R^{50}}{\text{minimize}} \sum_{n=1}^{50} U_n$$

subject to $U_n = |y_n - x_n^T \beta|$

The expression $U_n = |y_n - x_n^T \beta|$ can be written as:

$$U_n \geq -(y_n - x_n^T \beta) \;\&\; U_n \geq (y_n - x_n^T \beta)$$

We can convert this into a linear programming problem:

We can define U_n as:

$$\sum_{n=1}^{N} U_n = \sum_{p=0}^{4} (0)(\beta_p) + \sum_{n=1}^{50} (1)(U_n)$$

$$= \underbrace{\begin{bmatrix} 0 \dots 0 & 1 \dots 1 \end{bmatrix}}_{1 \times (55)} \begin{bmatrix} \beta \\ U \end{bmatrix}$$

$U_n \geq -(y_n - x_n^T \beta)$ can be written as

$$x_n^T \beta - U_n \leq y_n$$

$$\begin{bmatrix} x_1^T & -1 & 0 & \dots & 0 \\ x_2^T & 0 & -1 & \dots & 0 \\ \vdots & & & & \\ x_{50}^T & 0 & 0 & \dots & -1 \end{bmatrix} \begin{bmatrix} \beta \\ U_1 \\ \vdots \\ U_{50} \end{bmatrix} \leq \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{50} \end{bmatrix}$$

$$\begin{bmatrix} X & -I \end{bmatrix} \begin{bmatrix} \beta \\ U \end{bmatrix} \leq y$$

Similarly constraint $U_n \geq y_n - x_n^T \beta$ can be written as

$$-x_n^T \beta - U_n \leq -y_n$$

$$\begin{bmatrix} -X & -I \end{bmatrix} \begin{bmatrix} \beta \\ U \end{bmatrix} \leq -y$$

$\therefore$ Our problem can now be defined as

$$\text{minimize} \quad \begin{bmatrix} 0_5 & 1_{50} \end{bmatrix} \begin{bmatrix} \beta \\ u \end{bmatrix}$$

$$\text{subject to} \quad \begin{bmatrix} X & -I \\ -X & -I \end{bmatrix} \begin{bmatrix} \beta \\ u \end{bmatrix} \leq \begin{bmatrix} y \\ -y \end{bmatrix}$$

It is of the form:

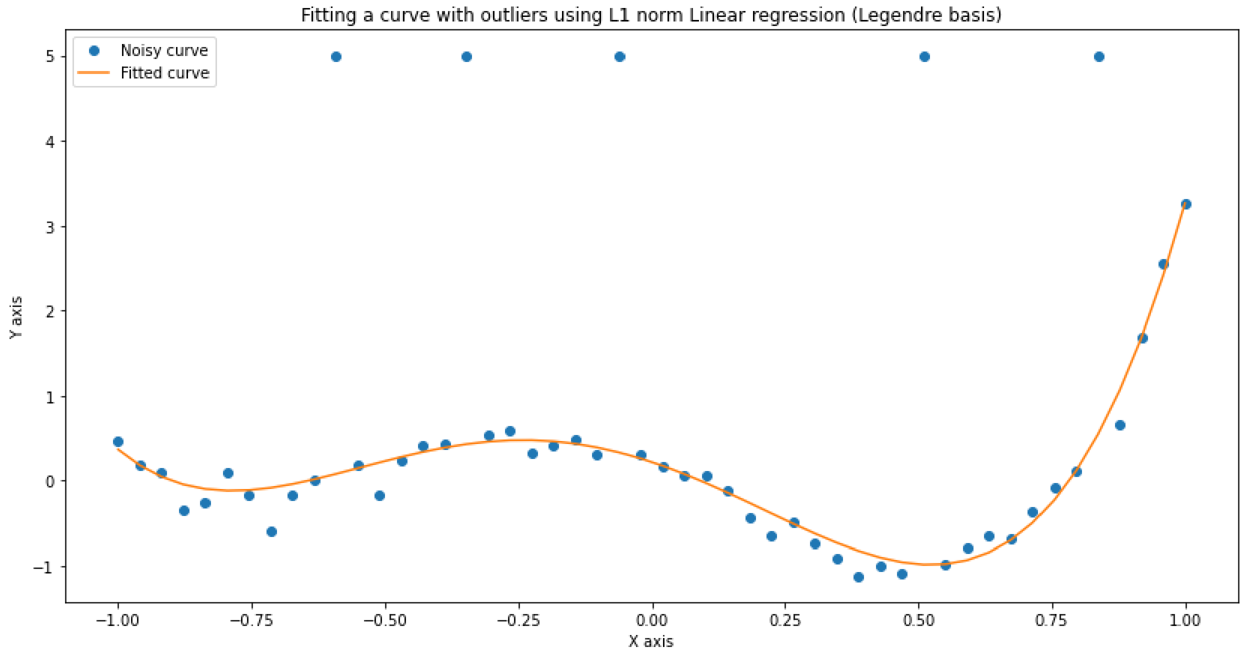
$$\underset{x}{\text{minimize}} \quad c^T x$$

$$\text{subject to} \quad Ax \leq b$$

$$c = \begin{bmatrix} 0_5 \\ 1_{50} \end{bmatrix} \qquad x = \begin{bmatrix} \beta \\ u \end{bmatrix}$$

$$A = \begin{bmatrix} X & -I \\ -X & -I \end{bmatrix} \qquad b = \begin{bmatrix} y \\ -y \end{bmatrix}$$

f)



Fitting a curve with outliers using L1 norm Linear regression (Legendre basis)
Noisy curve
Fitted curve
Y axis
X axis
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

APPENDIX