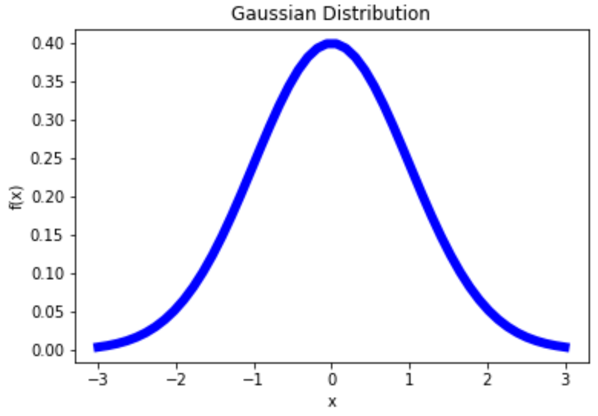
Problem-1: Histogram & Cross-Validation

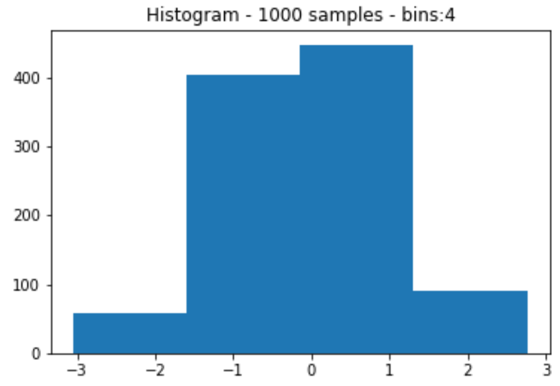
ⓐ

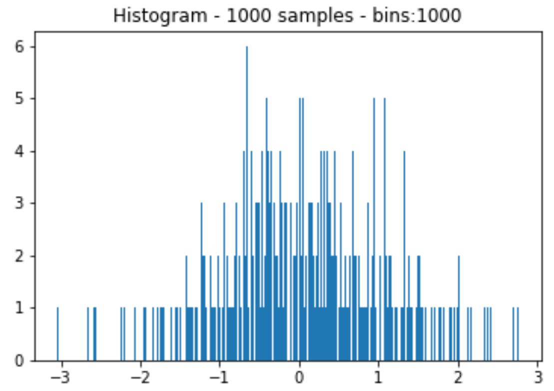


ⓑ Drawing N=1000 samples

(ii) Bins = 4

Bins = 1000





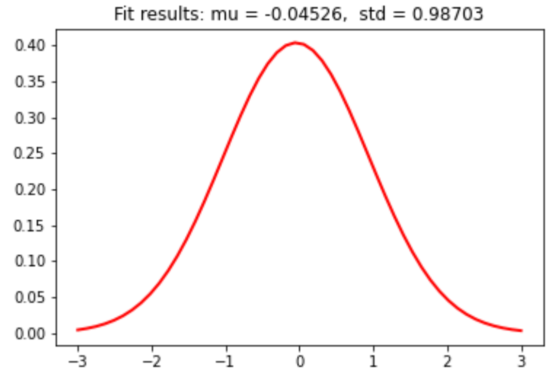
(iii) Estimating the mean & std. deviation of our data

$\mu = -0.04526$

$\sigma = 0.98703$

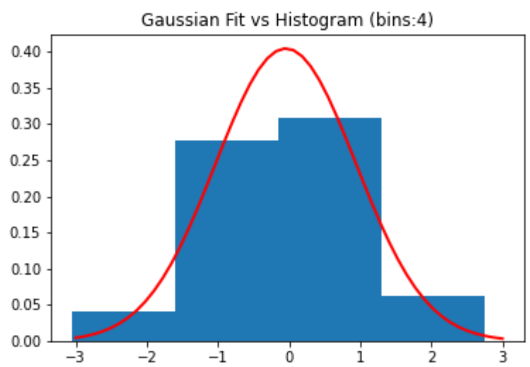
(iv) Gaussian curve vs Histogram

(I) fitted Gaussian Curve

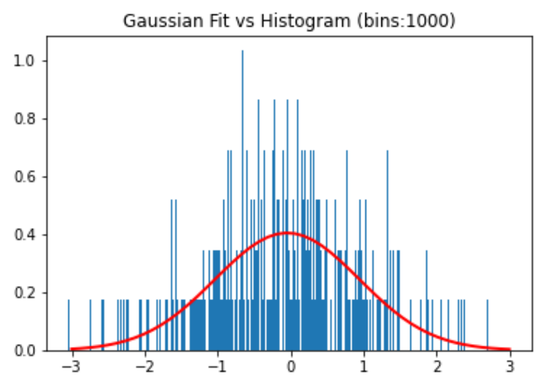


(II) Fitted Gaussian vs Histogram

ⓐ Bins = 4



ⓑ Bins = 1000



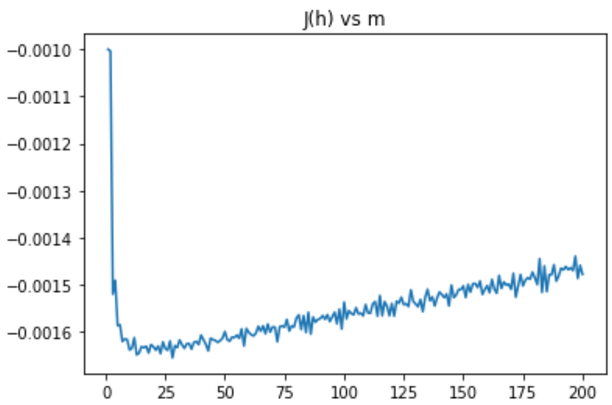
→ red line depicts fitted gaussian curve

(v) Are two histograms representative of our data's distribution?

→ The two histograms are not the ideal representation of our data's distribution.

(c) Cross Validation

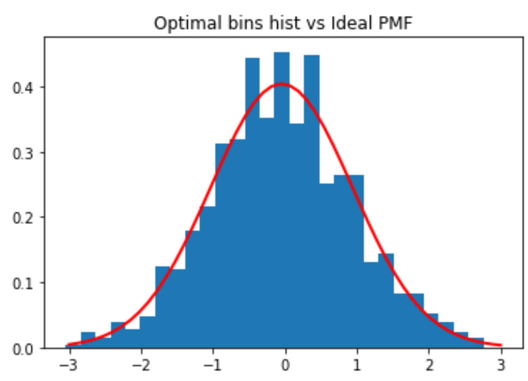
(i) Plot $\hat{J}(h)$



(ii) Find m^* that minimizes $\hat{J}(h)$

The "m" that minimizes $J(h)$ is $\boxed{m=28}$

(iii) Optimal bins vs histogram with optimal bins



Problem-2: Gaussian Whitening

(a) $|\Sigma| = (4-1) = 3$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

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

$$x-\mu = \begin{bmatrix} x_1 - 2 \\ x_2 - 6 \end{bmatrix}$$

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

$$= \frac{1}{3}\begin{bmatrix} 2x_1 - x_2 + 2 \\ 2x_2 - x_1 - 10 \end{bmatrix}$$

$$(x-\mu)^T \Sigma^{-1}(x-\mu) \Rightarrow \begin{bmatrix} (x_1-2) & (x_2-6) \end{bmatrix} \times \frac{1}{3}\begin{bmatrix} 2x_1 - x_2 + 2 \\ 2x_2 - x_1 - 10 \end{bmatrix}$$

$$= \frac{(x_1-2)(2x_1-x_2+2) + (x_2-6)(2x_2-x_1-10)}{3}$$

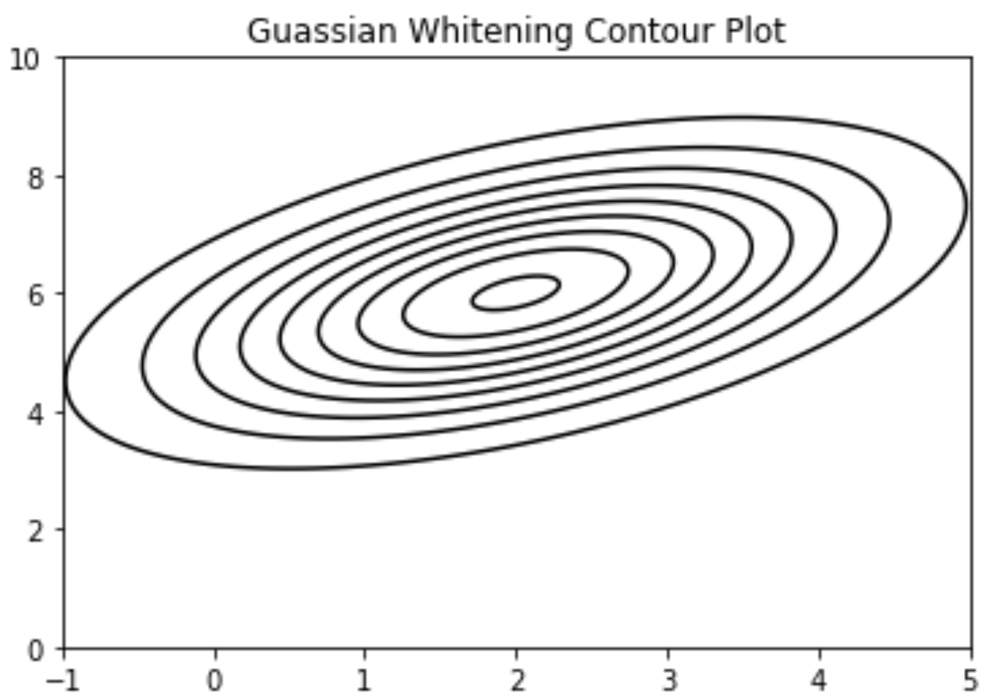
$$= \frac{\left(2x_1^2 - x_1x_2 + \overset{-2x_1}{\cancel{2x_1} - \cancel{4x_1}} + 2x_2 - 4\right) + \left(2x_2^2 - x_1x_2 - \overset{-22x_2}{\cancel{10x_2} - \cancel{12x_2}} + 6x_1 + 60\right)}{3}$$

$$= \boxed{\left(\frac{2x_1^2 + 2x_2^2 - 2x_1x_2 + 4x_1 - 20x_2 + 56}{3}\right)}$$

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

$$\boxed{f_X(x) = \frac{1}{\sqrt{12\pi^2}} \exp\left\{\left(\frac{x_1x_2 + 10x_2 - 2x_1 - x_1^2 - x_2^2 - 28}{3}\right)\right\}}$$

(ii) Contour Plot of $f_X(x)$



(b) PDF of $X \sim \mathcal{N}(0, I)$:

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

here $\mu = 0$, $\Sigma = I$

Given $Y = AX + b$

(i) $\mu_y \triangleq \mathbb{E}[Y] \Rightarrow \mathbb{E}[AX+b] = A\,\mathbb{E}[X] + \mathbb{E}[b]$

$$= A \cancelto{0}{\mu_X} + b \cancelto{1}{\int_{-\infty}^{\infty} f_X(x)\,dx}$$

So $\mu_y = 0 + b \Rightarrow \boxed{\mu_y = b}$

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

$$\triangleq \mathbb{E}\left[(Y-\mu_y)(Y^T - \mu_y^T)\right]$$

$$\Rightarrow \mathbb{E}\left[YY^T - \mu_y Y^T - Y\mu_y^T + \mu_y \mu_y^T\right]$$

$$\Rightarrow \mathbb{E}\left[(AX+b)(X^TA^T+b^T) - \mu_y(X^TA^T+b^T) \ldots \right.$$

$$\left. \ldots - (AX+b)\mu_y^T + \mu_y\mu_y^T\right]$$

$$\Rightarrow \mathbb{E}\left[(AXX^TA^T) + \cancel{AXb^T} + \cancel{bX^TA^T} + \cancel{bb^T} - \cancel{bX^TA^T} - \cancel{bb^T} - \cancel{AXb^T} - \cancel{bb^T} + \cancel{bb^T}\right]$$

[since $\mu_y = b$]

$\Rightarrow \mathbb{E}[A x x^T A^T]$

$\Rightarrow A\,\mathbb{E}[x x^T] A^T$ ——————— (1)

We know $\Sigma_x = I$ i.e $\Sigma_x = \mathbb{E}[(x-0)(x-0)^T]$ as $\mu_x = 0$

so $I = \mathbb{E}[x x^T]$ ——— (2)

using (2) in (1) we get

$$\boxed{\Sigma_y = A A^T}$$

(ii) Show Σ_y is symmetric positive semi-definite

i.e $\Sigma_y = \Sigma_y^T$

(a) $\Sigma_y = A A^T \quad \Sigma_y^T = (A A^T)^T = A A^T$

$\therefore \Sigma_y = \Sigma_y^T$ so Σ_y is symmetric.

(b) Positive Semi definite

To prove: $v^T \Sigma_y v \geq 0 \quad \forall\, v \in \mathbb{R}^d$

$$v^T \Sigma_y v = v^T\,\mathbb{E}[(Y-\mu_y)(Y-\mu_y)^T]\,v$$
$$= \mathbb{E}[\,v^T (Y-\mu_y)(Y-\mu_y)^T v\,]$$

$$= \mathbb{E}[b^T b] \quad (\text{where } b = (Y - \mu_y)^T v)$$

$$= \mathbb{E}[\|b\|^2] \quad \text{i.e} \geq 0$$

so $v^T \Sigma_y v \geq 0 \therefore \Sigma_y$ is positive semi definite

(iii) A & Σ_y? Such that Σ_y = sym. & positive definite

If Σ_y is positive definite then $v^T \Sigma_y v > 0$

i.e $v^T (A A^T) v > 0$ for all $v \in \mathbb{R}^n$ as $\Sigma_y = A A^T$.

So $(v^T A)(A^T v) > 0$; let $d = A^T v$

i.e $d^T . d > 0$, we know $d^T . d \geq 0$

but for $d^T . d > 0$ it means that $d = A^T v$ can't be zero.

So the nullspace of A^T should be empty such that $A^T v$ never goes to zero for any $v \in \mathbb{R}^n$.

That means "A" should be invertible

(iv) Given $\mu_y = \begin{bmatrix} 2 \\ 6 \end{bmatrix}$ & $\Sigma_y = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$

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

Finding A:

$\Sigma_y = U \Lambda U^T$ [eigen decomposition]

$$U^T U = I \Rightarrow \begin{bmatrix} u_1 & u_3 \\ u_2 & u_4 \end{bmatrix} \begin{bmatrix} u_1 & u_2 \\ u_3 & u_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\Rightarrow \boxed{u_1^2 + u_3^2 = 1} \quad \boxed{u_1 u_2 + u_3 u_4 = 0}$$

$$\boxed{u_2^2 + u_4^2 = 1}$$

$$\underbrace{u_3^2 = 1 - u_1^2} \qquad \underbrace{u_4 = \left(\frac{-u_1 u_2}{u_3} \right)}$$

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} u_1 & u_2 \\ u_3 & u_4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} u_1 & u_3 \\ u_2 & u_4 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} u_1 & 3u_2 \\ u_3 & 3u_4 \end{bmatrix} \begin{bmatrix} u_1 & u_3 \\ u_2 & u_4 \end{bmatrix} = \begin{bmatrix} u_1^2 + 3u_2^2 & \cdot \\ u_1 u_3 + 3u_4 u_2 & u_3^2 + 3u_4^2 \end{bmatrix}$$

$$\boxed{2 = u_1^2 + 3u_2^2} \qquad \boxed{1 = u_1 u_3 + 3u_4 u_2}$$

$$\boxed{2 = u_3^2 + 3u_4^2} \rightarrow 2 = (1 - u_1^2) + 3(1 - u_2^2)$$

$$\Rightarrow 2 = 4 - u_1^2 - 3u_2^2$$

$$1 = u_1 v_3 + 3u_2\left(\frac{-u_1 u_2}{u_3}\right)$$

$$\Rightarrow \quad \frac{u_3}{u_1} = u_3^2 - 3u_2^2 = 1 - u_1^2 - 3u_2^2$$

$$= 1 - (2) = -1$$

$$\boxed{u_1 = -u_3}$$

So $\quad u_1 = u_3 = \pm \frac{1}{\sqrt{2}}$

Lets say $\quad u_1 = \left(\frac{1}{\sqrt{2}}\right) \quad u_3 = \left(\frac{-1}{\sqrt{2}}\right)$

$u_4 u_2 = \frac{1}{2} \quad \& \quad u_2^2 = \frac{1}{2}$

So $\quad u_2 = +\frac{1}{\sqrt{2}} \;\&\; u_4 = +\frac{1}{\sqrt{2}}$

(or)

$u_2 = \left(\frac{-1}{\sqrt{2}}\right) \;\&\; u_4 = \left(\frac{-1}{\sqrt{2}}\right)$

u_1	u_3
$\frac{1}{\sqrt{2}}$	$\frac{-1}{\sqrt{2}}$
$\frac{-1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$

So $\quad U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$ (or) $\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$

(a) $\cancel{\begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}}$ (b) $\cancel{\begin{bmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}}$

$$\overbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}}^{U} \overbrace{\begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}}^{\lambda} \overbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}}^{V^T} = AA^T$$

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

$$\underbrace{\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix}}_{A^T} = \begin{bmatrix} a_{11}^2 + a_{12}^2 & a_{11}a_{21} + a_{12}a_{22} \\ a_{21}a_{11} + a_{22}a_{12} & a_{21}^2 + a_{22}^2 \end{bmatrix}$$

$1 = a_{11}a_{21} + a_{12}a_{22}$

$2 = a_{11}^2 + a_{12}^2$

$2 = a_{21}^2 + a_{22}^2$

$a_{11} = \frac{\sqrt{3}}{\sqrt{2}}$ $a_{12} = \frac{1}{\sqrt{2}}$

$a_{21} = \frac{\sqrt{3}}{\sqrt{2}}$ $a_{22} = \frac{-1}{\sqrt{2}}$

$a_{11}\, a_{21} = \frac{\sqrt{3}}{\sqrt{2}} \cdot \frac{\sqrt{3}}{\sqrt{2}}$

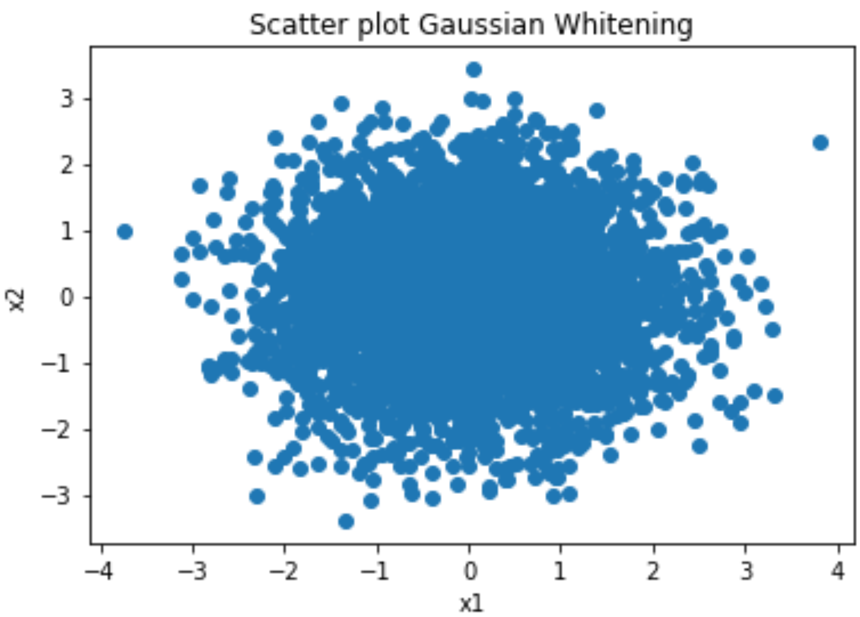
$a_{12}\, a_{22} = \frac{1}{\sqrt{2}} \cdot \frac{-1}{\sqrt{2}}$

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

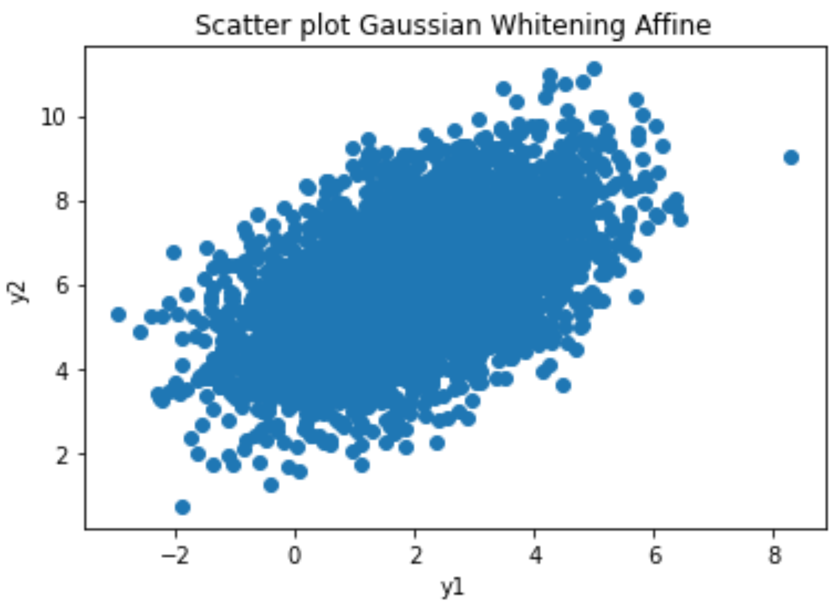
This is one possible solution for "A"

(c)

(i) 2D standard normal distribution



(ii) Affine transformation



$$AA^T = U\Lambda U^T$$

$$= (U\sqrt{\Lambda})(\sqrt{\Lambda}^T U^T)$$

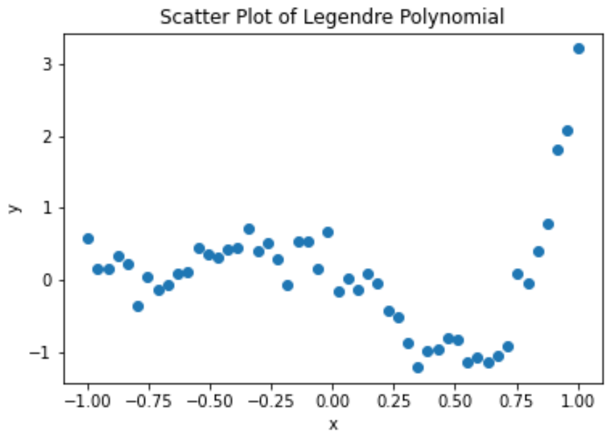
$$= (U\sqrt{\Lambda})(U\sqrt{\Lambda})^T$$

So $\boxed{A = U\sqrt{\Lambda}}$

U & Λ are obtained from eigen value decomposition of Σ_y.

Problem-3:

(a) Scatter Plot Data



(b) Given $N=50$

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

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{50} \end{bmatrix}$$

$y_{50\times1}$

where $y_i = \beta_0 + \beta_1 L_1(x_i) + \beta_2 L_2(x_i) + \beta_3 L_3(x_i) + \beta_4 L_4(x_i)$

& $x_i = i^{th}$ element in np.linspace(−1,1,50)

$$X = \begin{bmatrix} L_0(x_1) & L_1(x_1) & L_2(x_1) & L_3(x_1) & L_4(x_1) \\ L_0(x_2) & L_1(x_2) & L_2(x_2) & L_3(x_2) & L_4(x_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ L_0(x_{50}) & L_1(x_{50}) & L_2(x_{50}) & L_3(x_{50}) & L_4(x_{50}) \end{bmatrix}$$

$X_{50 \times 5}$

$$\beta = [\beta_0 \quad \beta_1 \quad \beta_2 \quad \beta_3 \quad \beta_4]^T$$

$\beta_{5\times 1}$

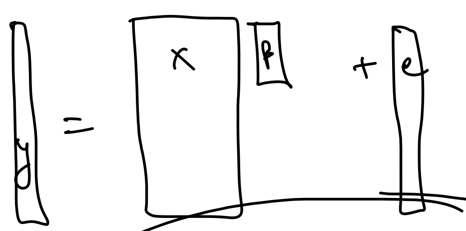
$\beta_0 = -0.001 \quad \beta_1 = 0.01 \quad \beta_2 = 0.55 \quad \beta_3 = 1.5 \quad \beta_4 = 1.2$

Deriving optimal solution:

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

$$\underset{50\times 1}{y} = \underset{50\times 5}{X}\ \underset{5\times 1}{\beta} + e$$

Over-determined system



$N > d \quad N = 50 \quad d = 5$

→ Using python code to find rank of X, i.e numpy.linalg.matrix_rank(X) we get rank = 6

→ As $\operatorname{rank}(X) = d$ the optimal solution derivation is as follows:

$$\text{Where } J = \|X\beta - y\|^2$$

$$\nabla_\theta J = 2X^T(X\beta - \bar{y}) = 0$$

$$\Rightarrow \quad X^T X \beta = X^T y$$

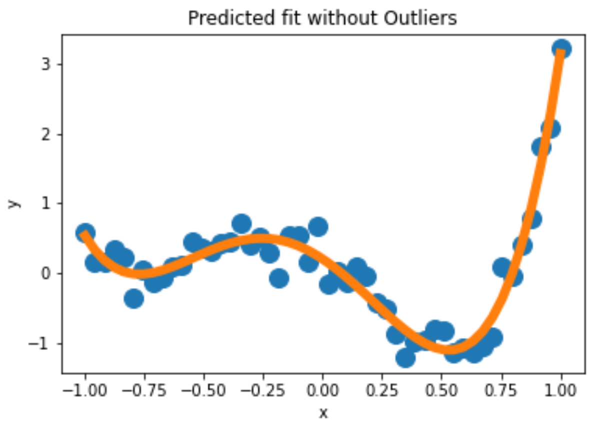
$$\Rightarrow \quad \boxed{\beta = (X^T X)^{-1} X^T y}$$

$(X^T X)$ is invertible because rank of $(X) = d$.

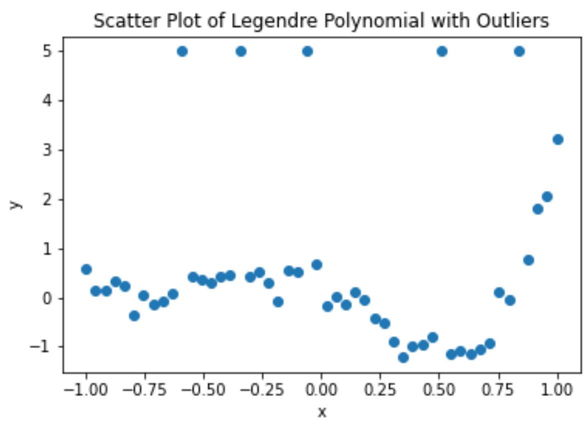
So now the predicted curve $(\hat{y})$ would be

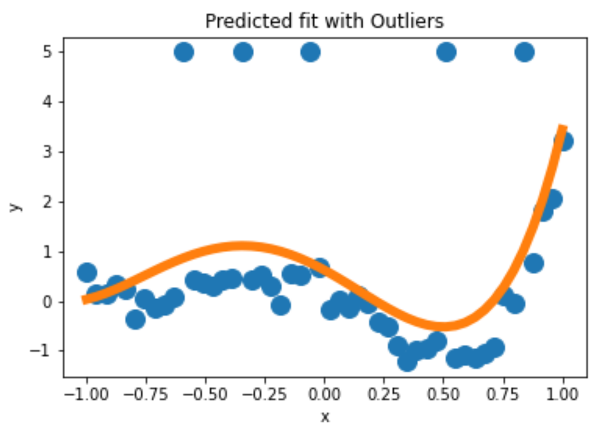
$$\hat{y} = X\hat{\beta} \quad \text{where } \hat{\beta} = (X^T X)^{-1} X^T y$$

c) Predicted curve with Scatter plot



d) Predicted curve with outliers





Comment on difference:

The prediction curve has shifted upwards due to the presence of the outliers, the farther the outliers are, the prediction curve would deviate from the data points.

(e) Optimization

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

In linear programing form : $\boxed{\begin{array}{l}\underset{x}{\text{minimize}}\ c^T x \\ \text{subject to } Ax \leq b\end{array}}$

Sol :

$$\min_{\bar{\beta}\cdot\bar{u}} \|\bar{u}\|_1 \quad \text{s.t} \quad \bar{u} \geq |y - X\beta|$$

$$\Downarrow$$

$$\bar{u} \geq -(y - X\beta)$$

$$\&$$

$$\bar{u} \geq (y - X\beta)$$

So $\min_{\bar{\beta}\cdot\bar{u}} \|\bar{u}\|_1 \Rightarrow \min_{\bar{\beta}\cdot\bar{u}} \begin{bmatrix} \bar{0} \\ \bar{1} \end{bmatrix}^T \begin{bmatrix} \bar{\beta} \\ \bar{u} \end{bmatrix}$

such that $\begin{bmatrix} X & -I \\ -X & -I \end{bmatrix} \begin{bmatrix} \bar{\beta} \\ \bar{u} \end{bmatrix} \leq \begin{bmatrix} \bar{y} \\ -y \end{bmatrix}$

i.e

$$\underset{[55\times 1]}{C} = \begin{bmatrix} \bar{0} \\ \bar{1} \end{bmatrix} \quad ; \text{ where } \bar{0} = [0\ 0\ 0\ 0\ 0]^T$$

$$\bar{1} = [\underbrace{1\ \ldots\ 1}_{50 \text{ elements}}]^T$$

$$\underset{[100\times 55]}{A} = \begin{bmatrix} X & -I \\ -X & -I \end{bmatrix} \quad ; \quad X = \text{same as defined in (b)}$$

$$I = [I]_{50\times 50}$$

$$\underset{[100\times1]}{B} = \begin{bmatrix} \bar{y} \\ -\bar{y} \end{bmatrix} \quad ; \quad \bar{y} = [\bar{y}]_{50\times1}$$

$$\underset{[55\times1]}{\bar{x}} = \begin{bmatrix} \bar{\beta} \\ \bar{u} \end{bmatrix} \quad ; \quad \bar{\beta} = [\bar{\beta}]_{5\times1}$$

$$\bar{u} = [\bar{u}]_{50\times1}$$

(f) Linear programing