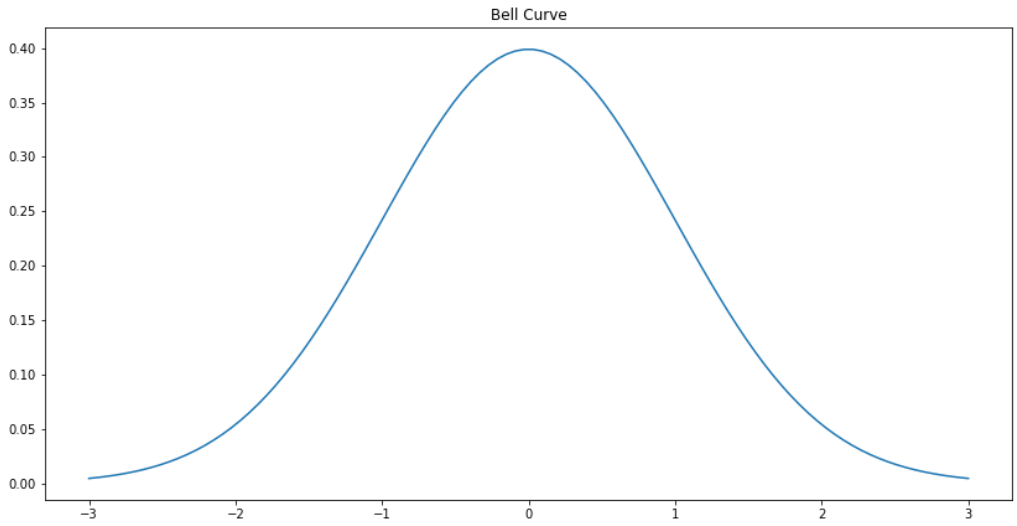
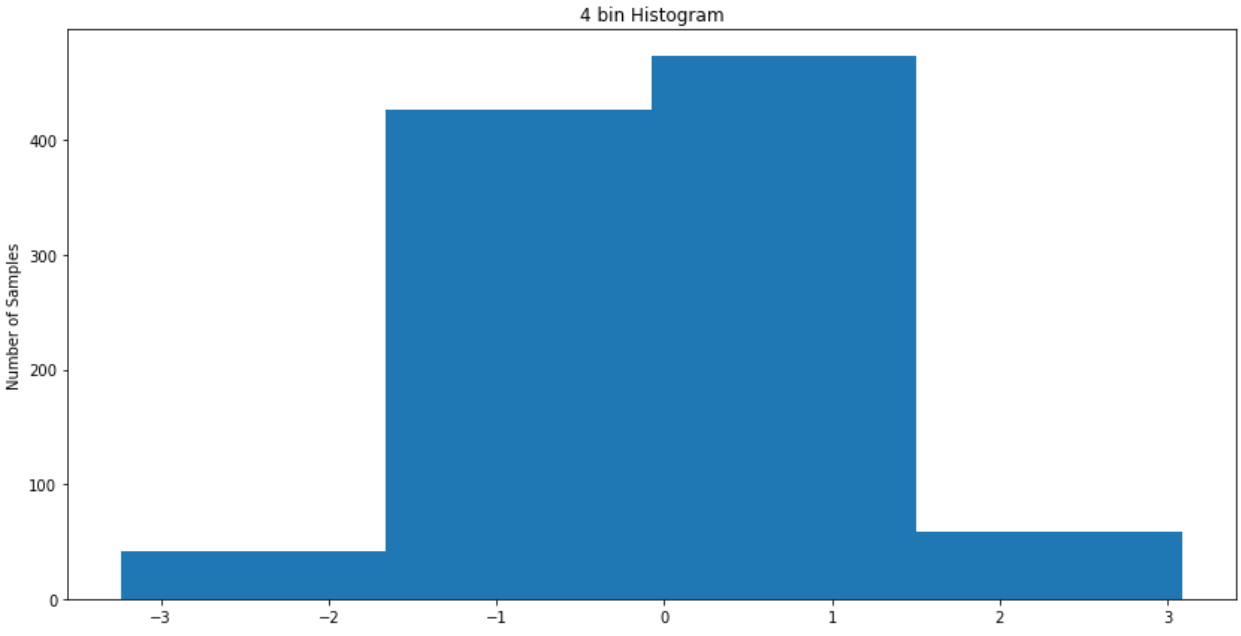
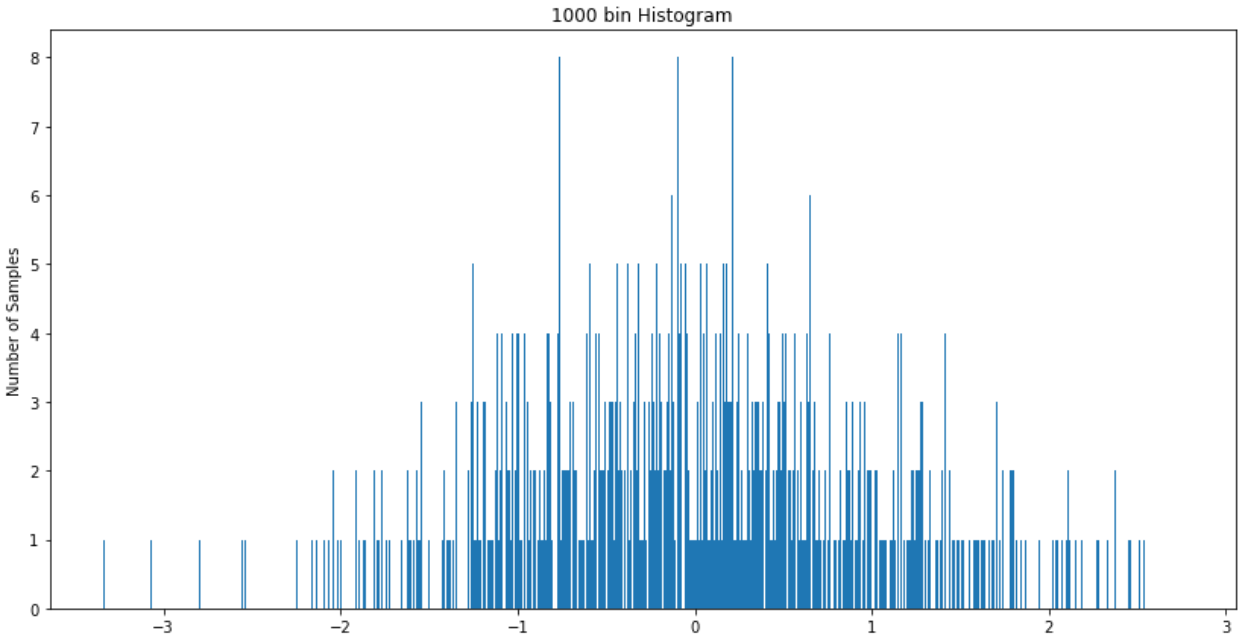
ECE 50024
Homework 1

Parth Sagar Hasabnis phasabni@purdue.edu

Exercise 1:

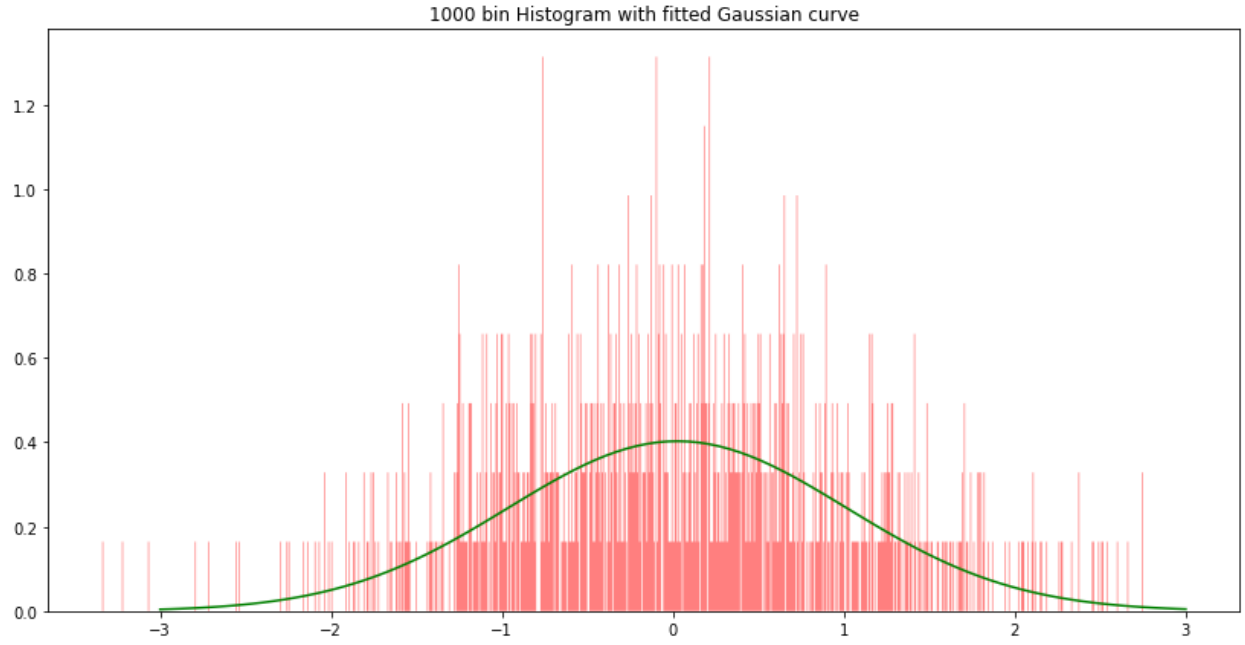






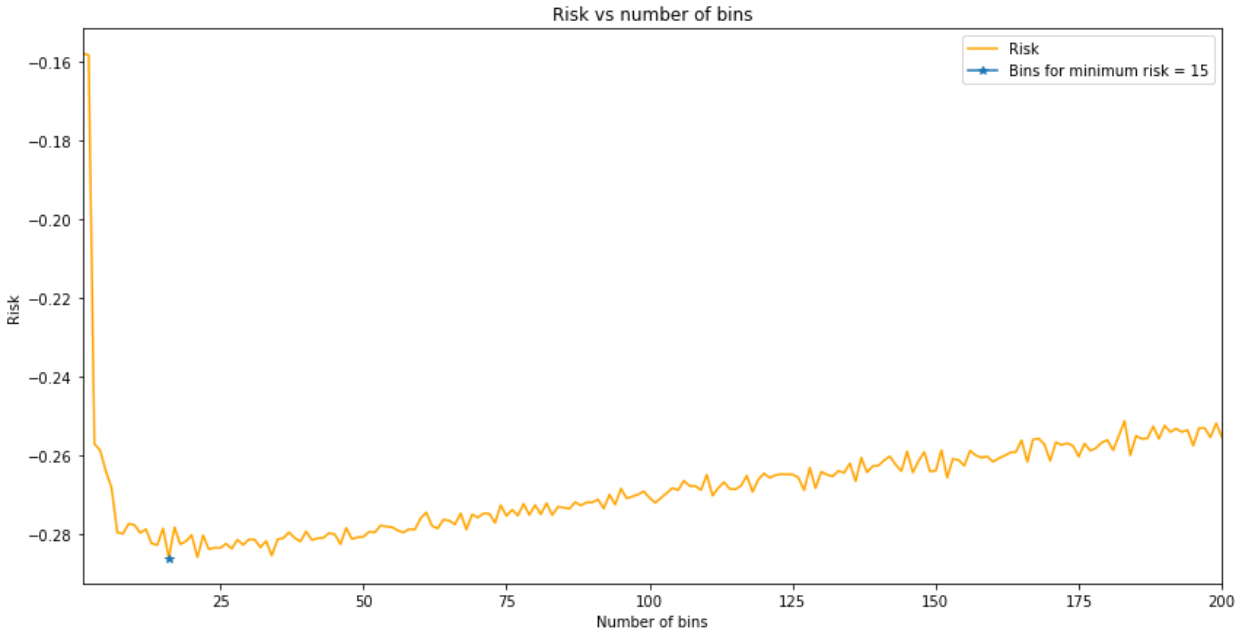
Mean of the distribution is: 0.026005414594233493.

Standard Deviation of the distribution is: 0.9907009680534483.

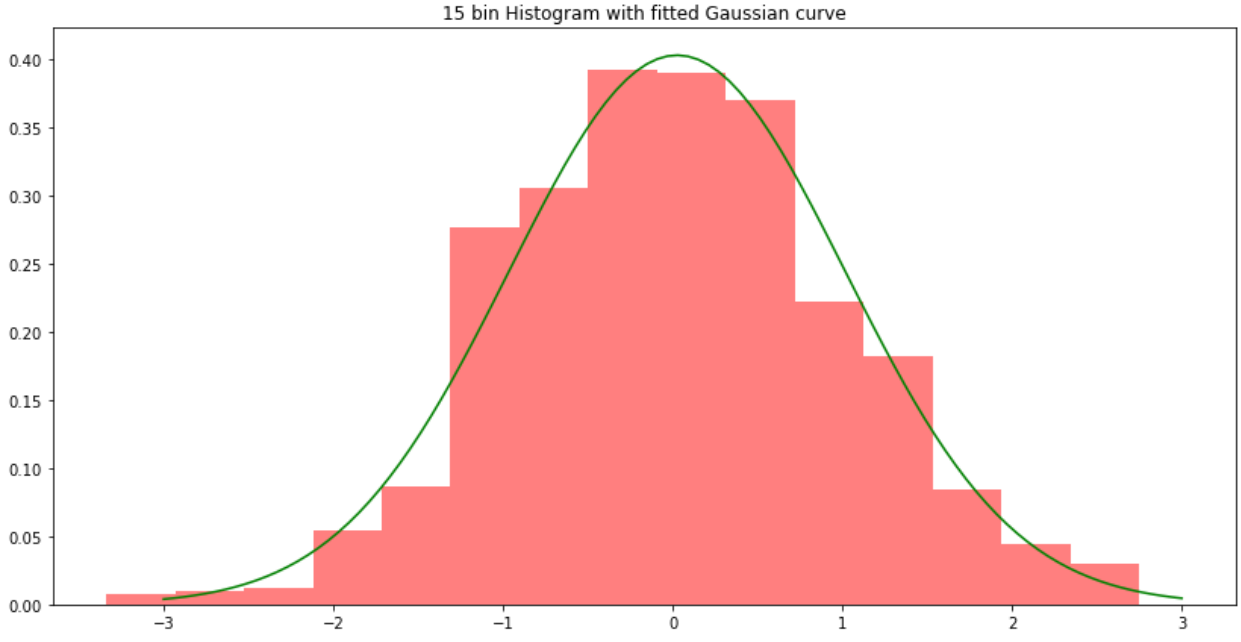


Exercise 1

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)

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

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

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

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

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

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

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

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

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

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

b)

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

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

$$Y = AX + b$$

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

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

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

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

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

$$\mu_Y = b$$

$$\Sigma_Y = \mathbb{E}\left[(Y-\mu_Y)(Y-\mu_Y)^T\right]$$

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

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

$$= \mathbb{E}\left[AXX^TA^T\right]$$

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

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

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

$$= \Sigma_X$$

$$\therefore \quad \Sigma_Y = A\,\Sigma_X\,A^T$$

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,

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

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

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

$\therefore$ 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 $v = \vec{0}_{N\times 1}$

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

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

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

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

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

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

$$\lambda = 1, \; 3$$

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

$\lambda_1 = 1$

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

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

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

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

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

$$\therefore V_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$\therefore$ 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 $V = \begin{bmatrix} 1 & 1 \\ v_1 & v_2 \\ 1 & 1 \end{bmatrix}$

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

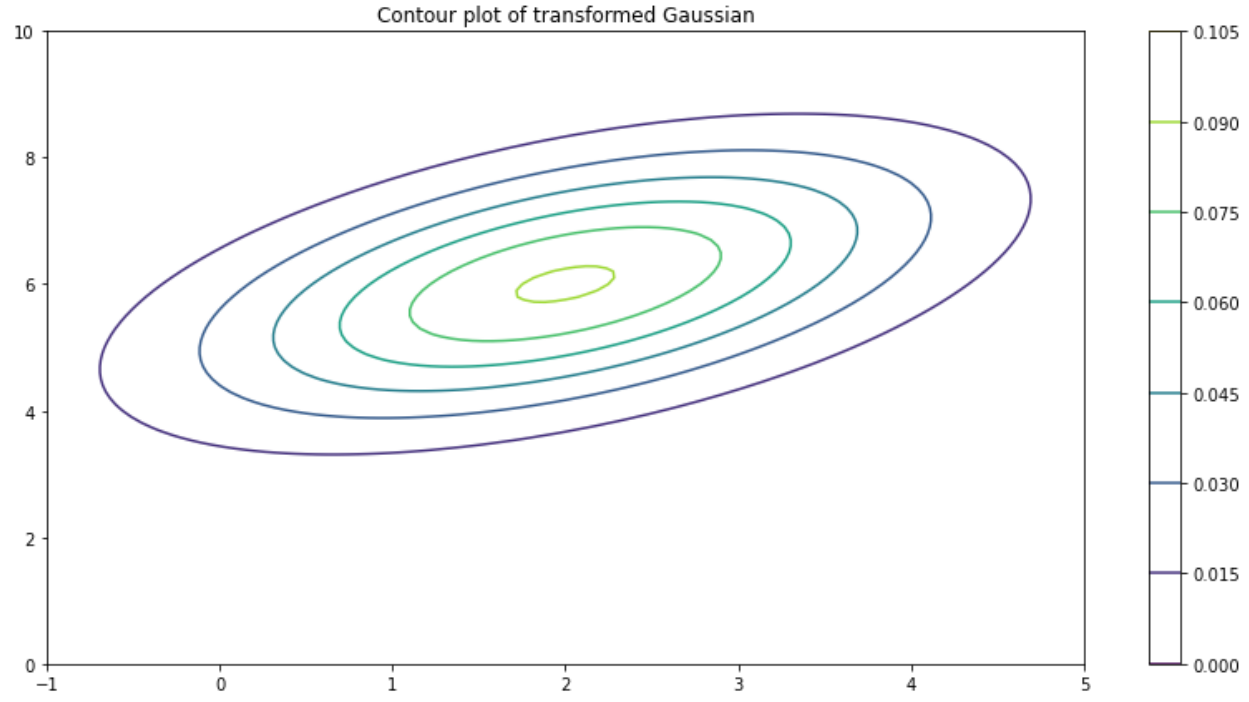
$$\Sigma_Y = V \Lambda V^{-1}$$
$$= A \cdot A^T$$

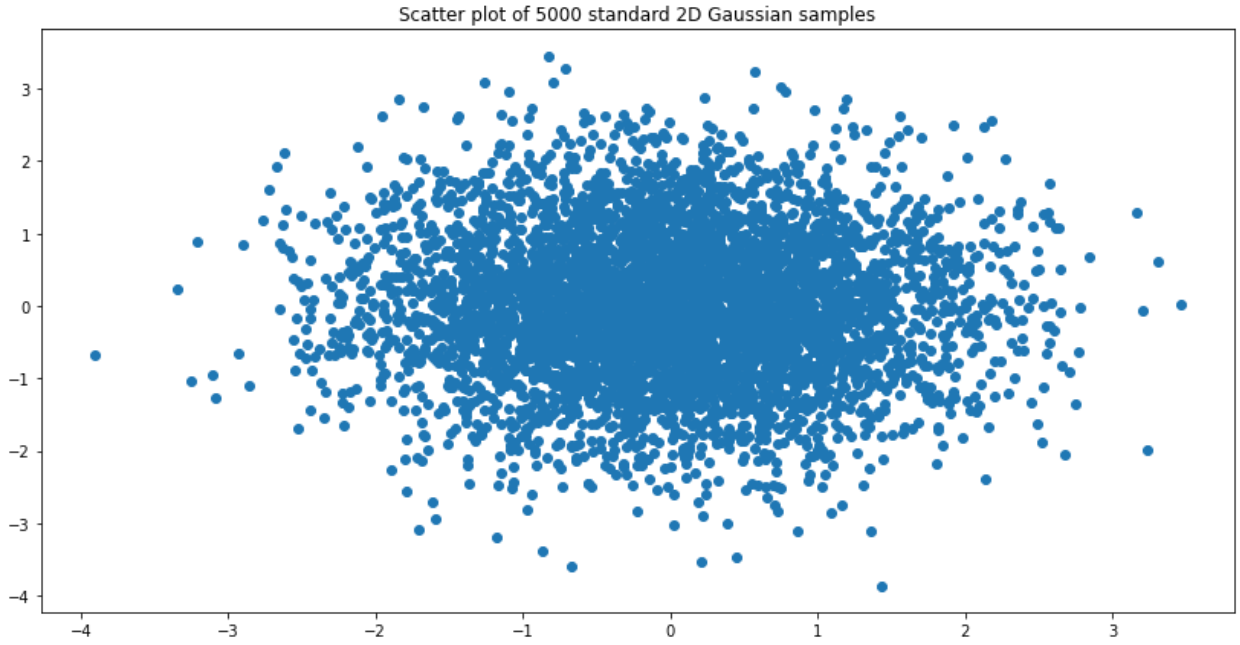
$$\therefore A = V \Lambda^{1/2} V^{-1}$$

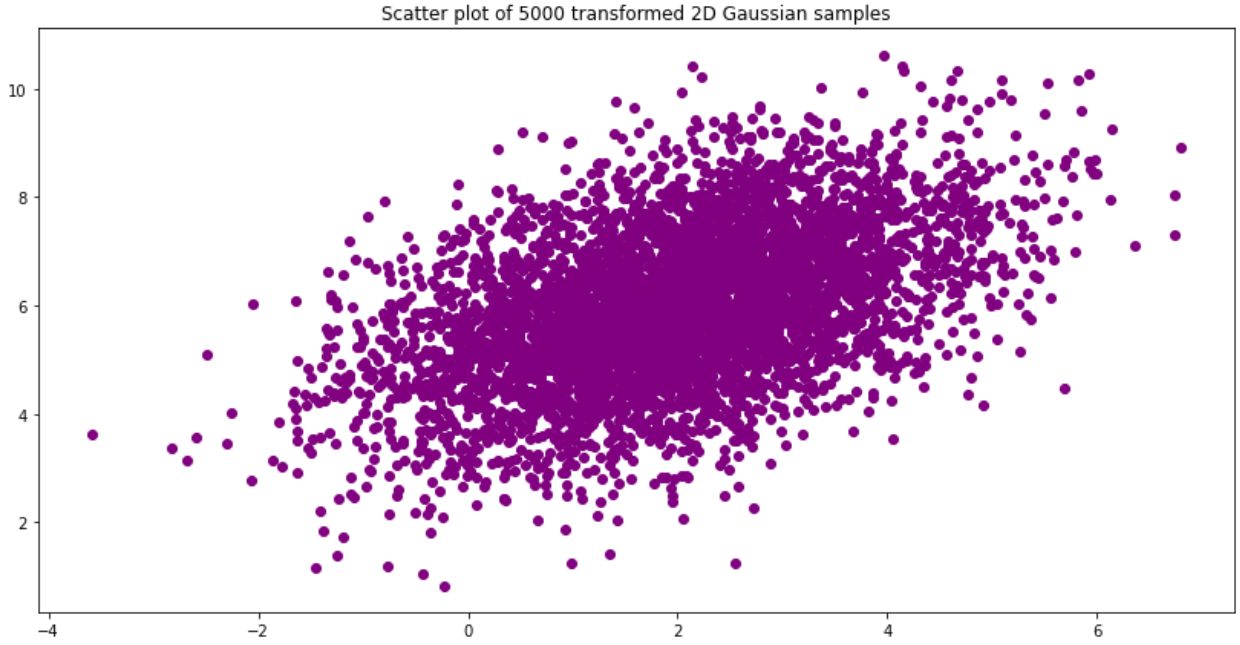
$$= \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{bmatrix} \begin{bmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

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

c)



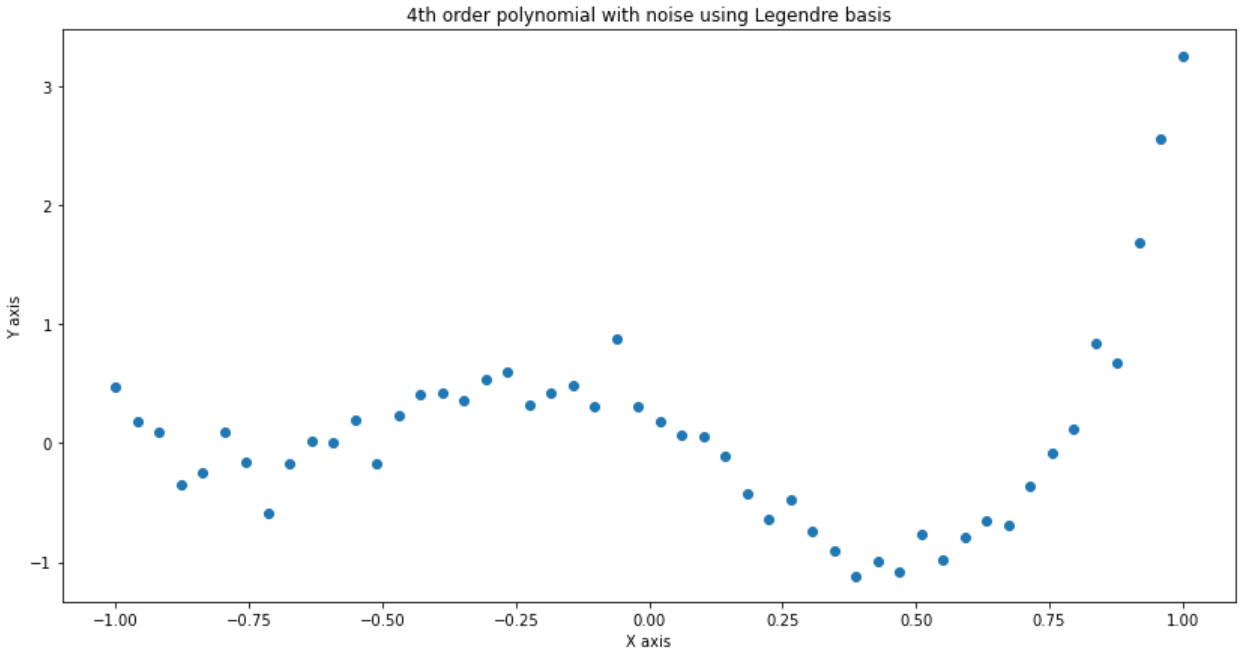




The results from parts (c)(i) and (ii) are in line with the theoritical findings. The plot of (c)(ii) represents a 2D transformed gaussian quite effectively. Similarly, the A matrix calculated from the transformed samples matches with its hand-calculated counterpart.

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

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

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

$L_p(x_n)$ 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

$$\beta = \begin{bmatrix} -0.001 \\ 0.01 \\ 0.55 \\ 1.5 \\ 1.3 \end{bmatrix}$$

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

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

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

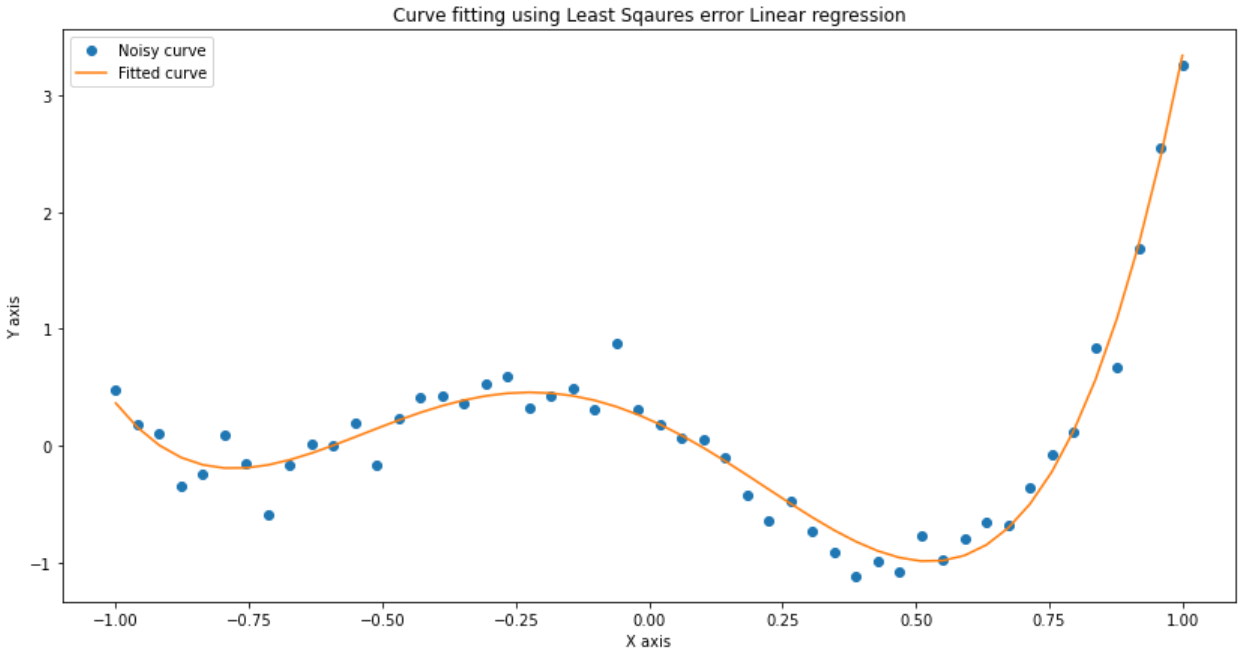
equating it to zero,

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

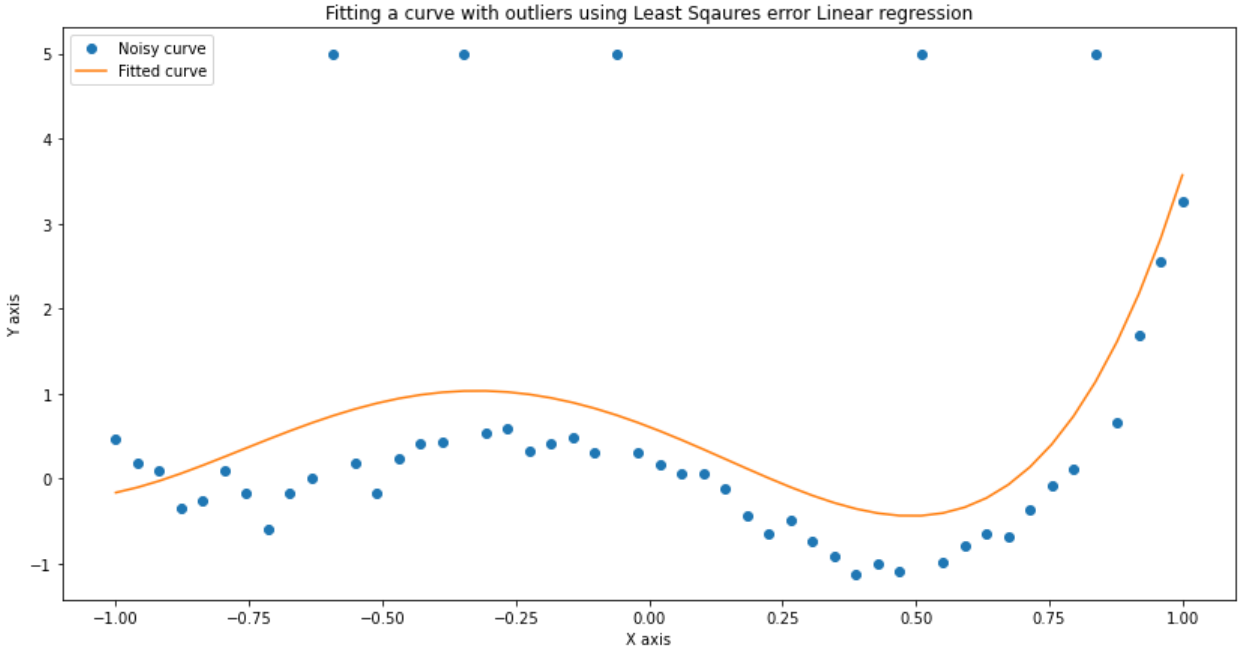
$$X^T X\beta = X^T y$$

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

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 $u_n = |y_n - x_n^T \beta|$

∴ 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 v_n as:

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

$$= \underbrace{[0 \dots 0 \; 1 \dots 1]}_{1 \times (55)} \begin{bmatrix} \beta \\ v \end{bmatrix}$$

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

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

$$\begin{bmatrix} x_1^T & -1 & 0 & \dots & 0 \\ x_2^T & 0 & -1 & \dots & 0 \\ \vdots & & & & \\ x_{50}^T & 0 & 0 & \dots & -1 \end{bmatrix} \begin{bmatrix} \beta \\ v_1 \\ \vdots \\ v_{50} \end{bmatrix} \leq \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{50} \end{bmatrix}$$

$$[X \; -I] \begin{bmatrix} \beta \\ v \end{bmatrix} \leq y$$

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

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

$$[-X \; -I] \begin{bmatrix} \beta \\ v \end{bmatrix} \leq -y$$

∴ Our problem can now be defined as

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

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

It is of the form:

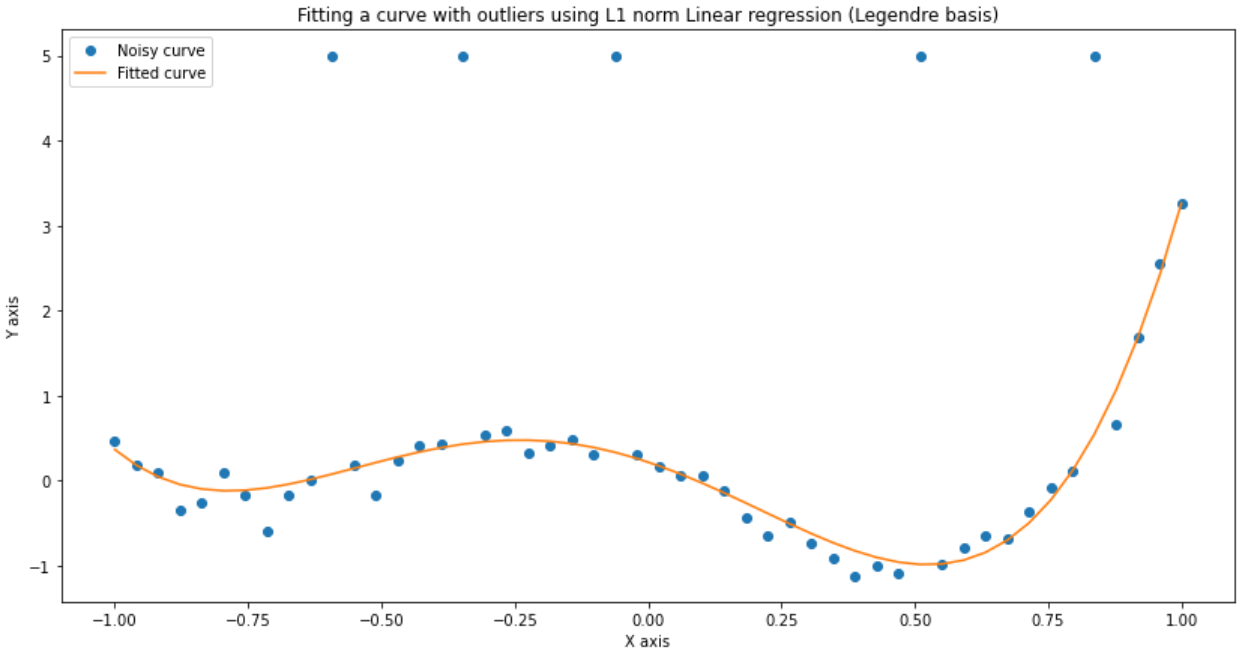
$$\underset{x}{\text{minimize}} \quad c^T x$$

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

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

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

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

Conditional Generative Adversarial Networks

Parth Sagar Hasabnis, Master of Science in ECE[1]

Abstract

This document provides a basic paper template and submission guidelines. Abstracts must be a single paragraph, ideally between 4–6 sentences long. Gross violations will trigger corrections at the camera-ready phase.

1. Electronic Submission

Submission to ICML 2021 will be entirely electronic, via a web site (not email). Information about the submission process and LaTeX templates are available on the conference web site at:

http://icml.cc/

The guidelines below will be enforced for initial submissions and camera-ready copies. Here is a brief summary:

- Submissions must be in PDF.
- Submitted papers can be up to eight pages long, not including references, plus unlimited space for references. Accepted papers can be up to nine pages long, not including references, to allow authors to address reviewer comments. Any paper exceeding this length will automatically be rejected.
- **Do not include author information or acknowledgements** in your initial submission.
- Your paper should be in **10 point Times font**.
- Make sure your PDF file only uses Type-1 fonts.
- Place figure captions *under* the figure (and omit titles from inside the graphic file itself). Place table captions *over* the table.
- References must include page numbers whenever possible and be as complete as possible. Place multiple citations in chronological order.
- Do not alter the style template; in particular, do not compress the paper format by reducing the vertical spaces.
- Keep your abstract brief and self-contained, one paragraph and roughly 4–6 sentences. Gross violations will require correction at the camera-ready phase. The title should have content words capitalized.

1.1. Submitting Papers

Paper Deadline: The deadline for paper submission that is advertised on the conference website is strict. If your full, anonymized, submission does not reach us on time, it will not be considered for publication.

Anonymous Submission: ICML uses double-blind review: no identifying author information may appear on the title page or in the paper itself. Section 2.3 gives further details.

Simultaneous Submission: ICML will not accept any paper which, at the time of submission, is under review for another conference or has already been published. This policy also applies to papers that overlap substantially in technical content with conference papers under review or previously published. ICML submissions must not be submitted to other conferences and journals during ICML's review period. Informal publications, such as technical reports or papers in workshop proceedings which do not appear in print, do not fall under these restrictions.

Authors must provide their manuscripts in **PDF** format. Furthermore, please make sure that files contain only embedded Type-1 fonts (e.g., using the program `pdffonts` in linux or using File/DocumentProperties/Fonts in Acrobat). Other fonts (like Type-3) might come from graphics files imported into the document.

Authors using **Word** must convert their document to PDF. Most of the latest versions of Word have the facility to do this automatically. Submissions will not be accepted in Word format or any format other than PDF. Really. We're not joking. Don't send Word.

Those who use **LaTeX** should avoid including Type-3 fonts. Those using `latex` and `dvips` may need the following two commands:

```
dvips -Ppdf -tletter -G0 -o paper.ps paper.dvi
```

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

```
ps2pdf paper.ps
```

It is a zero following the "-G", which tells dvips to use the config.pdf file. Newer TEX distributions don't always need this option.

Using `pdflatex` rather than `latex`, often gives better results. This program avoids the Type-3 font problem, and supports more advanced features in the `microtype` package.

Graphics files should be a reasonable size, and included from an appropriate format. Use vector formats (.eps/.pdf) for plots, lossless bitmap formats (.png) for raster graphics with sharp lines, and jpeg for photo-like images.

The style file uses the `hyperref` package to make clickable links in documents. If this causes problems for you, add `nohyperref` as one of the options to the `icml2021` usepackage statement.

1.2. Submitting Final Camera-Ready Copy

The final versions of papers accepted for publication should follow the same format and naming convention as initial submissions, except that author information (names and affiliations) should be given. See Section 2.3.2 for formatting instructions.

The footnote, "Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute." must be modified to "*Proceedings of the 38^{th} International Conference on Machine Learning*, Online, PMLR 139, 2021. Copyright 2021 by the author(s)."

For those using the **LaTeX** style file, this change (and others) is handled automatically by simply changing `\usepackage{icml2021}` to

```
\usepackage[accepted]{icml2021}
```

Authors using **Word** must edit the footnote on the first page of the document themselves.

Camera-ready copies should have the title of the paper as running head on each page except the first one. The running title consists of a single line centered above a horizontal rule which is 1 point thick. The running head should be centered, bold and in 9 point type. The rule should be 10 points above the main text. For those using the **LaTeX** style file, the original title is automatically set as running head using the `fancyhdr` package which is included in the ICML 2021 style file package. In case that the original title exceeds the size restrictions, a shorter form can be supplied by using

```
\icmltitlerunning{...}
```

just before `\begin{document}`. Authors using **Word** must edit the header of the document themselves.

2. Format of the Paper

All submissions must follow the specified format.

2.1. Dimensions

The text of the paper should be formatted in two columns, with an overall width of 6.75 inches, height of 9.0 inches, and 0.25 inches between the columns. The left margin should be 0.75 inches and the top margin 1.0 inch (2.54 cm). The right and bottom margins will depend on whether you print on US letter or A4 paper, but all final versions must be produced for US letter size.

The paper body should be set in 10 point type with a vertical spacing of 11 points. Please use Times typeface throughout the text.

2.2. Title

The paper title should be set in 14 point bold type and centered between two horizontal rules that are 1 point thick, with 1.0 inch between the top rule and the top edge of the page. Capitalize the first letter of content words and put the rest of the title in lower case.

2.3. Author Information for Submission

ICML uses double-blind review, so author information must not appear. If you are using LaTeX and the `icml2021.sty` file, use `\icmlauthor{...}` to specify authors and `\icmlaffiliation{...}` to specify affiliations. (Read the TeX code used to produce this document for an example usage.) The author information will not be printed unless `accepted` is passed as an argument to the style file. Submissions that include the author information will not be reviewed.

2.3.1. Self-Citations

If you are citing published papers for which you are an author, refer to yourself in the third person. In particular, do not use phrases that reveal your identity (e.g., "in previous work (**?**), we have shown ...").

Do not anonymize citations in the reference section. The only exception are manuscripts that are not yet published (e.g., under submission). If you choose to refer to such unpublished manuscripts (**?**), anonymized copies have to be submitted as Supplementary Material via CMT. However, keep in mind that an ICML paper should be self contained and should contain sufficient detail for the reviewers to evaluate the work. In particular, reviewers are not required to look at the Supplementary Material when writing their review.

2.3.2. Camera-Ready Author Information

If a paper is accepted, a final camera-ready copy must be prepared. For camera-ready papers, author information should start 0.3 inches below the bottom rule surrounding the title. The authors' names should appear in 10 point bold type, in a row, separated by white space, and centered. Author names should not be broken across lines. Unbolded superscripted numbers, starting 1, should be used to refer to affiliations.

Affiliations should be numbered in the order of appearance. A single footnote block of text should be used to list all the affiliations. (Academic affiliations should list Department, University, City, State/Region, Country. Similarly for industrial affiliations.)

Each distinct affiliations should be listed once. If an author has multiple affiliations, multiple superscripts should be placed after the name, separated by thin spaces. If the authors would like to highlight equal contribution by multiple first authors, those authors should have an asterisk placed after their name in superscript, and the term "*Equal contribution" should be placed in the footnote block ahead of the list of affiliations. A list of corresponding authors and their emails (in the format Full Name <email@domain.com>) can follow the list of affiliations. Ideally only one or two names should be listed.

A sample file with author names is included in the ICML2021 style file package. Turn on the `[accepted]` option to the stylefile to see the names rendered. All of the guidelines above are implemented by the LaTeX style file.

2.4. Abstract

The paper abstract should begin in the left column, 0.4 inches below the final address. The heading 'Abstract' should be centered, bold, and in 11 point type. The abstract body should use 10 point type, with a vertical spacing of 11 points, and should be indented 0.25 inches more than normal on left-hand and right-hand margins. Insert 0.4 inches of blank space after the body. Keep your abstract brief and self-contained, limiting it to one paragraph and roughly 4–6 sentences. Gross violations will require correction at the camera-ready phase.

2.5. Partitioning the Text

You should organize your paper into sections and paragraphs to help readers place a structure on the material and understand its contributions.

2.5.1. Sections and Subsections

Section headings should be numbered, flush left, and set in 11 pt bold type with the content words capitalized. Leave 0.25 inches of space before the heading and 0.15 inches after the heading.

Similarly, subsection headings should be numbered, flush left, and set in 10 pt bold type with the content words capitalized. Leave 0.2 inches of space before the heading and 0.13 inches afterward.

Finally, subsubsection headings should be numbered, flush left, and set in 10 pt small caps with the content words capitalized. Leave 0.18 inches of space before the heading and 0.1 inches after the heading.

Please use no more than three levels of headings.

2.5.2. Paragraphs and Footnotes

Within each section or subsection, you should further partition the paper into paragraphs. Do not indent the first line of a given paragraph, but insert a blank line between succeeding ones.

You can use footnotes[1] to provide readers with additional information about a topic without interrupting the flow of the paper. Indicate footnotes with a number in the text where the point is most relevant. Place the footnote in 9 point type at the bottom of the column in which it appears. Precede the first footnote in a column with a horizontal rule of 0.8 inches.[2]

2.6. Figures

You may want to include figures in the paper to illustrate your approach and results. Such artwork should be centered, legible, and separated from the text. Lines should be dark and at least 0.5 points thick for purposes of reproduction, and text should not appear on a gray background.

Label all distinct components of each figure. If the figure takes the form of a graph, then give a name for each axis and include a legend that briefly describes each curve. Do not include a title inside the figure; instead, the caption should serve this function.

Number figures sequentially, placing the figure number and caption *after* the graphics, with at least 0.1 inches of space before the caption and 0.1 inches after it, as in Figure 1. The figure caption should be set in 9 point type and centered unless it runs two or more lines, in which case it should be flush left. You may float figures to the top or bottom of a column, and you may set wide figures across both columns (use the environment `figure*` in LaTeX). Always place two-column figures at the top or bottom of the page.

[1] Footnotes should be complete sentences.

[2] Multiple footnotes can appear in each column, in the same order as they appear in the text, but spread them across columns and pages if possible.

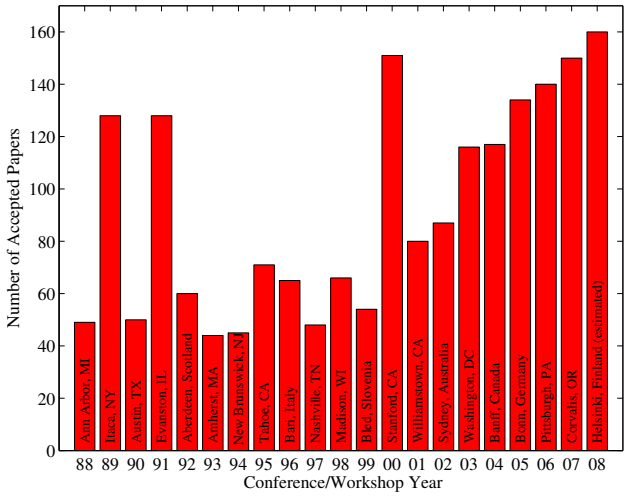


Figure 1. Historical locations and number of accepted papers for International Machine Learning Conferences (ICML 1993 – ICML 2008) and International Workshops on Machine Learning (ML 1988 – ML 1992). At the time this figure was produced, the number of accepted papers for ICML 2008 was unknown and instead estimated.

Algorithm 1 Bubble Sort

Input: data x_i, size m
repeat
  Initialize $noChange = true$.
  for $i = 1$ **to** $m - 1$ **do**
    if $x_i > x_{i+1}$ **then**
      Swap x_i and x_{i+1}
      $noChange = false$
    end if
  end for
until $noChange$ is $true$

2.7. Algorithms

If you are using LaTeX, please use the "algorithm" and "algorithmic" environments to format pseudocode. These require the corresponding stylefiles, algorithm.sty and algorithmic.sty, which are supplied with this package. Algorithm 1 shows an example.

2.8. Tables

You may also want to include tables that summarize material. Like figures, these should be centered, legible, and numbered consecutively. However, place the title *above* the table with at least 0.1 inches of space before the title and the same after it, as in Table 1. The table title should be set in 9 point type and centered unless it runs two or more lines, in which case it should be flush left.

Table 1. Classification accuracies for naive Bayes and flexible Bayes on various data sets.

DATA SET	NAIVE	FLEXIBLE	BETTER?
BREAST	95.9± 0.2	96.7± 0.2	√
CLEVELAND	83.3± 0.6	80.0± 0.6	×
GLASS2	61.9± 1.4	83.8± 0.7	√
CREDIT	74.8± 0.5	78.3± 0.6	
HORSE	73.3± 0.9	69.7± 1.0	×
META	67.1± 0.6	76.5± 0.5	√
PIMA	75.1± 0.6	73.9± 0.5	
VEHICLE	44.9± 0.6	61.5± 0.4	√

Tables contain textual material, whereas figures contain graphical material. Specify the contents of each row and column in the table's topmost row. Again, you may float tables to a column's top or bottom, and set wide tables across both columns. Place two-column tables at the top or bottom of the page.

2.9. Citations and References

Please use APA reference format regardless of your formatter or word processor. If you rely on the LaTeX bibliographic facility, use `natbib.sty` and `icml2021.bst` included in the style-file package to obtain this format.

Citations within the text should include the authors' last names and year. If the authors' names are included in the sentence, place only the year in parentheses, for example when referencing Arthur Samuel's pioneering work (**?**). Otherwise place the entire reference in parentheses with the authors and year separated by a comma (**?**). List multiple references separated by semicolons (**???**). Use the 'et al.' construct only for citations with three or more authors or after listing all authors to a publication in an earlier reference (**?**).

Authors should cite their own work in the third person in the initial version of their paper submitted for blind review. Please refer to Section 2.3 for detailed instructions on how to cite your own papers.

Use an unnumbered first-level section heading for the references, and use a hanging indent style, with the first line of the reference flush against the left margin and subsequent lines indented by 10 points. The references at the end of this document give examples for journal articles (**?**), conference publications (**?**), book chapters (**?**), books (**?**), edited volumes (**?**), technical reports (**?**), and dissertations (**?**).

Alphabetize references by the surnames of the first authors, with single author entries preceding multiple author entries. Order references for the same authors by year of publication, with the earliest first. Make sure that each reference includes all relevant information (e.g., page numbers).

Please put some effort into making references complete, presentable, and consistent. If using bibtex, please protect capital letters of names and abbreviations in titles, for example, use {B}ayesian or {L}ipschitz in your .bib file.

Software and Data

If a paper is accepted, we strongly encourage the publication of software and data with the camera-ready version of the paper whenever appropriate. This can be done by including a URL in the camera-ready copy. However, **do not** include URLs that reveal your institution or identity in your submission for review. Instead, provide an anonymous URL or upload the material as "Supplementary Material" into the CMT reviewing system. Note that reviewers are not required to look at this material when writing their review.

Acknowledgements

Do not include acknowledgements in the initial version of the paper submitted for blind review.

If a paper is accepted, the final camera-ready version can (and probably should) include acknowledgements. In this case, please place such acknowledgements in an unnumbered section at the end of the paper. Typically, this will include thanks to reviewers who gave useful comments, to colleagues who contributed to the ideas, and to funding agencies and corporate sponsors that provided financial support.

A. Do *not* have an appendix here

Do not put content after the references. Put anything that you might normally include after the references in a separate supplementary file.

We recommend that you build supplementary material in a separate document. If you must create one PDF and cut it up, please be careful to use a tool that doesn't alter the margins, and that doesn't aggressively rewrite the PDF file. pdftk usually works fine.

Please do not use Apple's preview to cut off supplementary material. In previous years it has altered margins, and created headaches at the camera-ready stage.

APPENDIX

Exercise 1

In [ ]:
```
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats
```

a)

In [ ]:
```
figs=(14,7)
var = 1
sigma = np.sqrt(var)
mean = 0

x = np.linspace(-3, 3, 100)
f = 1/np.sqrt((2*np.pi*var)) * np.exp(-1*np.square(x - mean)/(2*var))

fig, ax1 = plt.subplots(figsize=figs)
ax1.plot(x,f)
plt.title("Bell Curve")
plt.savefig("Bell.png")
```

b)

In [ ]:
```
s1 = np.random.normal(mean, sigma, 1000) # Part b i)
fig, ax2 = plt.subplots(figsize=(14,7))

ax2.hist(s1, bins=4)
plt.title("4 bin Histogram")
plt.ylabel("Number of Samples")
plt.show()
```

In [ ]:
```
s2 = np.random.normal(mean, sigma, 1000)
fig, ax3 = plt.subplots(figsize=(14,7))

ax3.hist(s2, bins=1000)
plt.title("1000 bin Histogram")
plt.ylabel("Number of Samples")
plt.show()
```

In [ ]:
```
mu, var = scipy.stats.norm.fit(s1)
print("Mean of the distribution is:", mu)
print("Standard Deviation of the distribution is:", np.sqrt(var))
```

c)

In [ ]:
```
s3 = scipy.stats.norm.pdf(x, mu, np.sqrt(var))
fig, ax3 = plt.subplots(figsize=(14,7))
ax3.hist(s1, density=True, histtype='stepfilled', alpha=0.5, bins=4, color='red')
ax3.plot(x,s3, color='green')
```

```
plt.title("4 bin Histogram with fitted Gaussian curve")
plt.show()
```

In [ ]:
```
fig, ax4 = plt.subplots(figsize=(14,7))
ax4.hist(s2, density=True, histtype='stepfilled', alpha=0.5, bins=1000, color='red')
ax4.plot(x,s3, color='green')
plt.title("1000 bin Histogram with fitted Gaussian curve")
plt.show()
```

In [ ]:
```
J_array = []
n = 1000

for m in range(1,201):
    hist, bins = np.histogram(s1, m)
    h = (max(s1) - min(s1))/m

    p = 0
    for j in range(0, m):
        p = p + (hist[j]/n)**2

    J = 2/(h*(n-1)) - (n+1)/(h*(n-1))*p

    J_array.append(J)

h_star = np.argmin(J_array)

print("Optimal number of bins = {0}".format(h_star))
```

In [ ]:
```
fig, ax5 = plt.subplots(figsize=(14,7))
x_axis = np.arange(1,201)
ax5.plot(x_axis, J_array, color='orange')
ax5.plot(h_star+1, J_array[h_star], marker='*')
plt.xlim((1,200))
plt.title("Risk vs number of bins")
plt.xlabel("Number of bins")
plt.ylabel("Risk")
plt.legend(["Risk", "Bins for minimum risk = {0}".format(h_star)])
plt.show()
```

In [ ]:
```
fig, ax5 = plt.subplots(figsize=(14,7))
ax5.hist(s2, density=True, histtype='stepfilled', alpha=0.5, bins=h_star, color='red')
ax5.plot(x,s3, color='green')
plt.title("{0} bin Histogram with fitted Gaussian curve".format(h_star))
plt.show()
```

In [ ]:
```

```

Exercise 2

In [ ]:
```
import matplotlib.pyplot as plt
import numpy as np
from   scipy.stats import multivariate_normal

figs=(14,7)
```

In [ ]:
```
n = 101
x1 = (-1,5)
x2 = (0,10)
mean = np.array([2, 6])
cov = np.array([[2, 1],[1, 2]])

range_mat = [ [0 for _ in range(n)] for _ in range(n)]
X = [ [0 for _ in range(n)] for _ in range(n)]
Y = [ [0 for _ in range(n)] for _ in range(n)]
for i in range(n):
    for j in range(n):
        val = (x1[0] + j*(x1[1] - x1[0])/(n-1), x2[0] + i*(x2[1] - x2[0])/(n-1))
        range_mat[i][j] = val
        X[i][j] = x1[0] + j*(x1[1] - x1[0])/(n-1)
        Y[i][j] = x2[0] + i*(x2[1] - x2[0])/(n-1)


det = np.linalg.det(cov)
F = [ [0 for _ in range(n)] for _ in range(n)]
for i in range(np.shape(F)[0]):
    for j in range(np.shape(F)[1]):
        matrix = np.matmul((range_mat[i][j] - mean).T, np.linalg.inv(cov))
        matrix = np.matmul(matrix,(range_mat[i][j] - mean))
        F[i][j] = 1/np.sqrt(((2*np.pi)**2)*det)*np.exp(-1/2*matrix)

fig, ax1 = plt.subplots(figsize=figs)
plt.contour(X, Y, F)
plt.colorbar()
plt.title("Contour plot of transformed Gaussian")
plt.show()
```

Q2 c)

In [ ]:
```
m = np.array([0,0])
c = np.array([[1, 0],[0, 1]])
Z = np.random.multivariate_normal(m, c, 5000)
fig, ax2 = plt.subplots(figsize=figs)
plt.scatter(Z[:,0], Z[:,1])
plt.title("Scatter plot of 5000 standard 2D Gaussian samples")
plt.show()
```

In [ ]:
```
from scipy.linalg import fractional_matrix_power

w, v = np.linalg.eig(cov)
lamb = np.diag(w)
```

```
A_dash = np.sqrt(lamb)
A = np.matmul(v,A_dash)
A = np.matmul(A, np.linalg.inv(v))

print(A)
```

In [ ]:
```
from numpy import matlib
Y = np.dot(A, Z.T) + matlib.repmat(mean,5000,1).T
Y = Y.T
fig, ax3 = plt.subplots(figsize=figs)
plt.scatter(Y[:,0], Y[:,1], color='purple')
plt.title("Scatter plot of 5000 transformed 2D Gaussian samples")
plt.show()
```

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

In [ ]:
```
import numpy as np
from scipy.special import eval_legendre
import matplotlib.pyplot as plt
from scipy.optimize import linprog

figs=(14,7)
```

In [ ]:
```
N = 50
x = np.linspace(-1,1,N)
L1 = eval_legendre(1,x)
L2 = eval_legendre(2,x)
L3 = eval_legendre(3,x)
L4 = eval_legendre(4,x)
```

In [ ]:
```
b0 = -0.001
b1 = 0.01
b2 = 0.55
b3 = 1.5
b4 = 1.2
b = np.array([b0, b1, b2, b3, b4])
y = b[0] + b[1]*L1 + b[2]*L2 + b[3]*L3 + b[4]*L4 + 0.2*np.random.randn(N)
```

In [ ]:
```
fig, ax1 = plt.subplots(figsize=figs)
plt.plot(x, y, 'o')
plt.title("4th order polynomial with noise using Legendre basis")
plt.xlabel("X axis")
plt.ylabel("Y axis")
```

In [ ]:
```
X = np.column_stack((eval_legendre(0,x), eval_legendre(1,x), eval_legendre(2,x), eval_le
b_hat= np.linalg.lstsq(X, y, rcond=None)[0]
y_hat = b_hat[0] + b_hat[1]*L1 + b_hat[2]*L2 + b_hat[3]*L3 + b_hat[4]*L4
```

In [ ]:
```
fig, ax2 = plt.subplots(figsize=figs)
plt.plot(x, y, 'o')
plt.plot(x, y_hat)
plt.title("Curve fitting using Least Sqaures error Linear regression")
plt.xlabel("X axis")
plt.ylabel("Y axis")
plt.legend(["Noisy curve", "Fitted curve"])
```

In [ ]:
```
idx = [10,16,23,37,45]
y[idx] = 5
X = np.column_stack((eval_legendre(0,x), eval_legendre(1,x), eval_legendre(2,x), eval_le
b_hat= np.linalg.lstsq(X, y, rcond=None)[0]
y_hat = b_hat[0] + b_hat[1]*L1 + b_hat[2]*L2 + b_hat[3]*L3 + b_hat[4]*L4

fig, ax3 = plt.subplots(figsize=figs)
plt.plot(x, y, 'o')
plt.plot(x, y_hat)
plt.title("Fitting a curve with outliers using Least Sqaures error Linear regression")
plt.xlabel("X axis")
```

```
plt.ylabel("Y axis")
plt.legend(["Noisy curve", "Fitted curve"])
```

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

In [ ]:
```
A = np.vstack((np.hstack((X, -np.eye(N))), np.hstack((-X, -np.eye(N)))))
b = np.hstack((y,-y))
c = np.hstack((np.zeros(5), np.ones(N)))
res = linprog(c, A, b, bounds=(None,None), method="revised simplex")
beta = res.x
t = np.linspace(-1,1,N)
y_hat = beta[0] + beta[1]*L1 + beta[2]*L2 + beta[3]*L3 + beta[4]*L4

fig, ax4 = plt.subplots(figsize=figs)
plt.plot(x,y,'o')
plt.plot(t,y_hat)
plt.title("Fitting a curve with outliers using L1 norm Linear regression (Legendre basis
plt.xlabel("X axis")
plt.ylabel("Y axis")
plt.legend(["Noisy curve", "Fitted curve"])
```

In [ ]:
```
X = np.column_stack((np.ones(x.shape), x, x**2, x**3, x**4))
A = np.vstack((np.hstack((X, -np.eye(N))), np.hstack((-X, -np.eye(N)))))
b = np.hstack((y,-y))
c = np.hstack((np.zeros(5), np.ones(N)))
res = linprog(c, A, b, bounds=(None,None), method="revised simplex")
beta = res.x
t = np.linspace(-1,1,N)
y_hat = beta[0] + beta[1]*x + beta[2]*(x**2) + beta[3]*(x**3) + beta[4]*(x**4)

fig, ax4 = plt.subplots(figsize=figs)
plt.plot(x,y,'o')
plt.plot(t,y_hat)
plt.title("Fitting a curve with outliers using L1 norm Linear regression (Ordinary basis
plt.xlabel("X axis")
plt.ylabel("Y axis")
plt.legend(["Noisy curve", "Fitted curve"])
```

In [ ]:
```
```