1. Consider the function $f(x) = -\sum_i^m log(b_i - a_i^T x)$. Compute ∇f and $\nabla^2 f$. Write down the first three terms of the Taylor series expansion of $f(x)$ around some x_0.

$$
\begin{aligned}
\frac{\partial f}{\partial x_j} &= -\sum_i^m \frac{1}{b_i - a_i^T x}\frac{\partial(b_i - a_i^T x)}{\partial x_j}, j = 1, .., n \\
&= -\sum_i^m \frac{1}{b_i - a_i^T x}(-a_{ij}) \\
&= \sum_i^m \frac{a_{ij}}{b_i - a_i^T x}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 f}{\partial x_{jk}} &= \frac{\partial \sum_i^m \frac{a_{ij}}{b_i - a_i^T x}}{\partial x_k}, j = 1, .., n, , k = 1, .., n \\
&= -\sum_i^m \frac{a_{ij}}{(b_i - a_i^T x)^2}\frac{\partial(b_i - a_i^T x)}{\partial x_k} \\
&= \sum_i^m \frac{a_{ij}}{(b_i - a_i^T x)^2}(a_{ik}) \\
&= \sum_i^m \frac{a_{ij}a_{ik}}{(b_i - a_i^T x)^2} \\
f(x) &= \sum_{i=0}^{\infty} \frac{f^i(x_0)(x - x_0)^i}{i!} \\
&= f(x_0) + f^{(1)}(x_0)(x - x_0) + \frac{1}{2}f^{(2)}(x_0)(x - x_0)^2 + o((x - a)^3) \\
&= (-\sum_i^m log(b_i - a_i^T x_0))_: + (\sum_i^m \frac{a_{ij}}{b_i - a_i^T x})^T(x - x_0) + \frac{1}{2}(x - x_0)^T(\sum_i^m \frac{a_{ij}a_{ik}}{(b_i - a_i^T x)^2})_{:,:}(x - x_0)
\end{aligned}
$$

2. 2.5 from textbook.
   What is the distance between two parallel hyperplanes $\{x \in \mathbb{R}^n : a^T x = b_1\}$ and $\{x \in R^n : a^T x = b_2\}$?
   Shortest displacement will be in the direction of a with scaling factor β, thus we solve for $|\beta|$ in the following.

$$
\begin{aligned}
a^T x_1 &= b_1 \\
a^T x_2 &= b_2 \\
x_1 &= x_2 + \beta a, \beta \in \mathbb{R} \\
a^T(x_2 + \beta a) &= b_1 \\
a^T x_2 + a^T \beta a) &= b_1 \\
b_2 + \beta a^T a &= b_1 \\
\beta &= \frac{b_1 - b_2}{a^T a} \\
|\beta| &= \frac{|b_1 - b_2|}{\|a\|_2^2}
\end{aligned}
$$

3. 2.14(a) from textbook.
   Expanded and restricted sets. Let $S \subseteq R^n$, and let $\|\dot{\|}$ be a norm on R^n.
   (a) For $a \geq 0$ we define S_a as $\{x : dist(x, S) \leq a\}$, where $dist(x, S) = \inf_{y \in S} \|x - y\|$. We refer to S_a as S expanded or extended by a. Show that if S is convex, then S_a is convex.

$$
\begin{aligned}
S \text{ is convex } &\implies (\forall y_1, y_2)(\forall \theta \in [0,1])\theta y_1 + (1-\theta) y_2 \in S, \text{substitute into } dist(x, S): \\
dist(x, S) &= \inf_{y_1, y_2 \in S} \|x - (\theta y_1 + (1-\theta) y_2)\| \\
&\text{let } x = \theta x_1 + (1-\theta) x_2 \text{ with no assumptions} \\
dist(x, S) &= \inf_{y_1, y_2 \in S} \|\theta(x_1 - y_1) + (1-\theta)(x_2 - y_2)\| \\
dist(x, S) &\leq \inf_{y_1 \in S} \|\theta(x_1 - y_1)\| + \inf_{y_2 \in S} \|(1-\theta)(x_2 - y_2)\| \text{ via triangular inequality} \\
&\leq |\theta| \inf_{y_1 \in S} \|x_1 - y_1\| + |(1-\theta)| \inf_{y_2 \in S} \|x_2 - y_2\|
\end{aligned}
$$

suppose $x_1, x_2 \in S_a \implies \inf_{y_1 \in S} \|x_1 - y_1\| \leq a \wedge \inf_{y_2 \in S} \|x_2 - y_2\| \leq a$ to satisfy definition of $S_a,\ then:$

$$
dist(x, S) \leq \theta a + (1-\theta) a = a
$$

$$
(\forall x_1, x_2 \in S_a)\{x = \theta x_1 + (1-\theta) x_2 : dist(x, S) \leq a\} \implies x_1, x_2 \in S_a \ \square
$$

4. Problem 3.14 from textbook.
Convex-concave functions and saddle-points. We say the function $f : R^n \times R^m \to R$ is convex-concave if $f(x, z)$ is a concave function of z, for each fixed x, and a convex function of x, for each fixed z. We also require its domain to have the product form $dom f = A \times B$, where $A \subseteq R^n$ and $B \subseteq R^m$ are convex.

(a) Give a second-order condition for a twice differentiable function $f : R^n \times R^m \to R$ to be convex-concave, in terms of its Hessian $\nabla^2 f(x, z)$.

$$(\forall z)\nabla^2 f(x, z) \preceq 0 \text{ in x} \wedge (\forall x)\nabla^2 f(x, z) \succeq 0 \text{ in z}$$

(b) Suppose that $f : R^n \times R^m \to R$ is convex-concave and differentiable, with $\nabla f(\tilde{x}, \tilde{z}) = 0$. Show that the saddle-point property holds: for all x, z, we have

$$f(\tilde{x}, z) \leq f(\tilde{x}, \tilde{z}) \leq f(x, \tilde{z})$$

Using 1st order condition for concavity and fixing z:

$$(\forall x_1, x_2) f(x_2, z) \geq f(x_1, z) + \nabla f(x_1, z) \begin{bmatrix} x_2 - x_1 \\ 0 \end{bmatrix}$$
$$\text{let } z = \tilde{z}, x_1 = \tilde{x}$$
$$(\forall x_2) f(x_2, \tilde{z}) \geq f(\tilde{x}, \tilde{z}) + \nabla f(\tilde{x}, \tilde{z}) \begin{bmatrix} x_2 - \tilde{x} \\ 0 \end{bmatrix}$$
$$\nabla f(\tilde{x}, \tilde{z}) = 0$$
$$(\forall x_2) f(x_2, \tilde{z}) \geq f(\tilde{x}, \tilde{z})$$

Using 1st order condition for convexity and fixing x:

$$(\forall z_1, z_2) f(x, z_2) \leq f(x, z_1) + \nabla f(x, z_1) \begin{bmatrix} z_2 - z_1 \\ 0 \end{bmatrix}$$
$$\text{let } x = \tilde{x}, z_1 = \tilde{z}$$
$$(\forall z_2) f(\tilde{x}, z_2) \leq f(\tilde{x}, \tilde{z}) + \nabla f(\tilde{x}, \tilde{z}) \begin{bmatrix} 0 \\ z_2 - \tilde{z} \end{bmatrix}$$
$$(\forall z_2) f(\tilde{x}, z_2) \leq f(\tilde{x}, \tilde{z})$$

Then:

$$(\forall z, \forall x) f(\tilde{x}, z) \leq f(\tilde{x}, \tilde{z}) \leq f(x, \tilde{z})$$

Show that this implies that f satisfies the strong max-min property:

$$\sup_z \inf_x f(x,z) = \inf_x \sup_z f(x,z)$$

(and their common value is $f(\tilde{x},\tilde{z})$).

From part B:

$$(\forall x_2) f(x_2,\tilde{z}) \geq f(\tilde{x},\tilde{z})$$
$$(\forall z_2) f(\tilde{x},z_2) \leq f(\tilde{x},\tilde{z})$$

Supremum and infimum in terms of results from part B:

$$\sup_z f(x,z) = f(x,\tilde{z}) : (\forall z) f(x,\tilde{z}) \geq f(x,z)$$
$$\inf_x f(x,z) = f(\tilde{x},z) : (\forall x) f(\tilde{x},z) \leq f(x,z)$$

Using substitution:

$$\sup_z \inf_x f(x,z) = \sup_z f(\tilde{x},z) = f(\tilde{x},\tilde{z})$$
$$\inf_x \sup_z f(x,z) = \inf_x f(x,\tilde{z}) = f(\tilde{x},\tilde{z})$$
$$\sup_z \inf_x f(x,z) = \inf_x \sup_z f(x,z) = f(\tilde{x},\tilde{z})$$

(c) Now suppose that $f : R^n \times R^m \to \mathbb{R}$ is differentiable, but not necessarily convex-concave, and the saddle-point property holds at $\tilde{x},\tilde{z}$:

$$f(\tilde{x},z) \leq f(\tilde{x},\tilde{z}) \leq f(x,\tilde{z})$$

for all x,z. Show that $\nabla f(\tilde{x},\tilde{z}) = 0$.

Using 1st order conditions:

$$(\forall x)\ f(x,\tilde{z}) \geq f(\tilde{x},\tilde{z}) + \nabla f(\tilde{x},\tilde{z}) \begin{bmatrix} x-\tilde{x} \\ 0 \end{bmatrix}$$
$$(\forall z)\ f(\tilde{x},z) \leq f(\tilde{x},\tilde{z}) + \nabla f(\tilde{x},\tilde{z}) \begin{bmatrix} 0 \\ z-\tilde{z} \end{bmatrix}$$
$$(\forall x,z)\ f(\tilde{x},z) - \nabla f(\tilde{x},\tilde{z}) \begin{bmatrix} 0 \\ z-\tilde{z} \end{bmatrix} \leq f(\tilde{x},\tilde{z}) \leq f(x,\tilde{z}) - \nabla f(\tilde{x},\tilde{z}) \begin{bmatrix} x-\tilde{x} \\ 0 \end{bmatrix}$$
$$(\forall x,z)\ f(\tilde{x},z) \leq f(\tilde{x},\tilde{x}) \leq f(x,\tilde{z}), \implies$$
$$(\forall z)\ 0 \leq \nabla f(\tilde{x},\tilde{z}) \begin{bmatrix} 0 \\ z-\tilde{z} \end{bmatrix}$$
$$(\forall x)\ 0 \leq -\nabla f(\tilde{x},\tilde{z}) \begin{bmatrix} x-\tilde{x} \\ 0 \end{bmatrix}, \implies$$
$$\nabla f(\tilde{x},\tilde{z}) = 0$$

5. Problem 3.16(a-c) from textbook.

   For each of the following functions determine whether it is convex, concave, quasiconvex, or quasiconcave.

   (a) $f(x) = e^x - 1$ on $\mathbb{R}$
   Using 2nd order condition:

   $$(\forall x \in \mathbb{R}) f^{(2)}(x) = e^x > 0 \implies f \text{ is convex}$$

   $f(x) = e^x - 1$ is monotone so $\{x : f(x) \leq \alpha\}$ and $\{x : f(x) \geq \alpha\}$ are convex $\implies f$ is quasiconvex and quasiconcave.
   So f is convex, quasiconvex, quasiconcave.

   (b) $f(x_1, x_2) = x_1 x_2$ on $\mathbb{R}^2_{++}$
   Using 2nd order condition:

   $$\nabla^2 f(x_1, x_2) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

   $$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \not\succeq 0 \wedge \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \not\preceq 0 \implies f \text{ not convex, not concave}$$

   Super- and sub-level sets:
   $x_1 x_2 = \alpha$, $x_2 = \frac{\alpha}{x_1}$
   epi $x_1 \mapsto \frac{\alpha}{x_1}$ is convex $\implies \{x_1, x_2 : x_1 x_2 \geq \alpha\}$ is convex, thus quasiconcave.
   hypo $x_1 \mapsto \frac{\alpha}{x_1}$ is not convex $\implies \{x_1, x_2 : x_1 x_2 \leq \alpha\}$ is not convex, thus not quasiconvex.
   So f is quasiconcave.

   (c) $f(x_1, x_2) = 1/(x_1 x_2)$ on $\mathbb{R}^2_{++}$

   $$\nabla^2 f(x_1, x_2) = \begin{bmatrix} \frac{2}{x_1^3 x_2} & \frac{1}{x_1^2 x_2^2} \\ \frac{1}{x_1^2 x_2^2} & \frac{2}{x_1 x_2^3} \end{bmatrix}$$

   $$\begin{bmatrix} \frac{2}{x_1^3 x_2} & \frac{1}{x_1^2 x_2^2} \\ \frac{1}{x_1^2 x_2^2} & \frac{2}{x_1 x_2^3} \end{bmatrix} \succeq 0 \implies f \text{ is convex} \implies f \text{ quasiconvex}$$

   Super- and sub-level sets:
   $\frac{1}{x_1 x_2} = \alpha$, $\frac{1}{\alpha x_1} = x_2$
   epi $x_1 \mapsto \frac{1}{\alpha x_1}$ is convex $\implies \{\frac{1}{x_1 x_2} \leq \alpha\}$ is convex, thus quasiconvex.
   hypo $x_1 \mapsto \frac{1}{\alpha x_1}$ is not convex $\implies \{\frac{1}{x_1 x_2} \geq \alpha\}$ is not convex, thus not quasiconcave.
   So f is convex and quasiconvex.

6. Problem 3.32(a) from textbook.
   Products and ratios of convex functions. In general the product or ratio of two convex functions is not convex. However, there are some results that apply to functions on R. Prove the following.

   (a) If f and g are convex, both nondecreasing (or nonincreasing), and positive functions on an interval, then fg is convex.

$$
\begin{aligned}
(fg)' &= f'g + fg' \\
(fg)'' &= f''g + f'g' + f'g' + fg'' \\
f'' \geq 0 \wedge g \geq 0 &\implies f''g \geq 0 \\
g'' \geq 0 \wedge f \geq 0 &\implies fg'' \geq 0 \\
f' \geq 0 \wedge g' \geq 0 &\implies f'g' \geq 0 \\
f' \leq 0 \wedge g' \leq 0 &\implies f'g' \geq 0 \\
(fg)'' \geq 0 &\implies f \text{ is convex, via 2nd order condition}
\end{aligned}
$$

7. Consider the function $f(x, y) = x^2 + y^2 + \beta xy + x + 2y$. Find (x^*, y^*) for which $\nabla f = 0$. Express your answer as a function of β. For which values of β is the (x^*, y^*) a global minimum of $f(x, y)$?

$$
\begin{aligned}
\nabla f &= \begin{bmatrix} 2x + \beta y + 1 \\ 2y + \beta x + 2 \end{bmatrix} = 0 \\
&2x + \beta y + 1 = 0 \\
&2y + \beta x + 2 = 0 \\
y &= \frac{-2 - \beta x}{2} \\
&2x - \beta - \frac{\beta^2 x}{2} + 1 = 0 \\
&x(2 - \frac{\beta^2}{2}) = \beta - 1 \\
x &= \frac{2(\beta - 1)}{4 - \beta^2} \implies |\beta| \neq 2 \text{ for } x^* \text{ to exist} \\
x^* &= \frac{2(\beta - 1)}{4 - \beta^2} \\
y^* &= -1 - \frac{\beta x^*}{2} \\
\nabla^2 f &= \begin{bmatrix} 2 & \beta \\ \beta & 2 \end{bmatrix} \succeq 0 \\
&|\beta| \leq 2 \text{ via Gersgorin Circle theorem so eigenvalues } \geq 0 \\
&|\beta| \leq 2 \wedge |\beta| \neq 2 \implies \beta \in (-2, 2)
\end{aligned}
$$

8. A function $f(x)$ is strongly convex with a positive factor m if $\nabla^2 f(x) \succeq mI$, for all x, where I denotes the identity matrix. Another equivalent definition of a m-strongly convex function, with respect to l_2-norm $\|\cdot\|_2$, is given by $f(y) \geq f(x) + \nabla f(x)^T(yx) + \frac{m}{2}\|y-x\|_2^2$ for all x, y.

   (a) Assume f (x) is a strongly convex function with factor m. Is f (x) also a strictly convex function?

   $(\forall x)\ \nabla^2 f(x) - mI \succeq 0 \iff \nabla^2 f(x) - mI$ is SPD $\iff (\forall i)\lambda_i^{'}(\nabla^2 f(x) - mI) \geq 0$
   mI is symmetric $\implies (\forall x)\nabla^2 f(x)$ is symmetric
   eigenvalues of the orginal hessian:
   $(\forall x)\ (\forall i)\lambda_i(\nabla^2 f(x)) \geq \lambda_n^{'}(\nabla^2 f(x) - mI) + m \geq 0 + m = m$
   $m > 0 \implies (\forall x)(\forall i)\lambda_i(\nabla^2 f(x)) > 0$
   $(\forall x)\nabla^2 f(x) = \nabla^2 f(x)^T \wedge (\forall i)\lambda_i(\nabla^2 f(x)) > 0 \implies (\forall x)\nabla^2 f(x) \succ 0 \implies f$ strictly convex

   (b) Assume g(x) is a strictly convex function. Is g(x) also a strongly convex function? Find the largest factor of strong convexity.

   Hint: You may assume that the eigen values of the hessian matrix, represented by $\lambda_1 \geq \lambda_2 \geq .. \geq \lambda_n$, are given and known. You may describe the largest strong convexity factor in terms of the eigen values of the hessian matrix.

   let $\lambda_i(A)$ denote ith sorted eigenvalue of A

   $g(x)$ strictly convex $\not\Longrightarrow (\forall x)\nabla^2 g(x) \succ 0($ eg: $f(x) = x^4, \nabla^2 f(x)|_{x=0} \not\succ 0)$

   case for $\nabla^2 g(x)$ *not PD* :
   $(\forall x)\nabla^2 g(x) \not\succ 0 \implies (\exists x)(\exists i)\lambda_i(\nabla^2 g(x)) = \lambda_n(\nabla^2 g(x)) \leq 0$
   $(\exists x)(\exists i)\lambda_i(\nabla^2 g(x) - mI) \geq \lambda_n(\nabla^2 g(x)) - m < 0$
   $m > 0 \wedge \lambda_n(\nabla^2 g(x)) \leq 0 \implies (\exists x)(\exists i)\lambda_i(\nabla^2 g(x) - mI) < 0 \implies (\forall x)\nabla^2 g(x) - mI \not\succeq 0$
   $\implies g(x)$ not strongly convex

   case for $\nabla^2 g(x)$ *PD* :
   $(\forall x)\nabla^2 g(x) \succ 0 \implies (\forall x)(\forall i)\lambda_i(\nabla^2 g(x)) \geq \lambda_n(\nabla^2 g(x)) > 0$
   $(\forall x)(\forall i)\lambda_i(\nabla^2 g(x) - mI) \geq \lambda_n(\nabla^2 g(x)) - m, m > 0, \lambda_n(\nabla^2 g(x)) > 0$
   let $m \leq \lambda_n(\nabla^2 g(x)) \implies (\forall x)(\forall i)\lambda_i(\nabla^2 g(x) - mI) \geq 0$
   $\implies (\forall x)\nabla^2 g(x) - mI \succeq 0 \implies g(x)$ strongly convex
   largest strong convexity factor: $m = \lambda_n(\nabla^2 g(x))$

9. In this problem, we are given a set of data points $(x_i, y_i), i = 1, .., 100$. We wish to fit a quadratic model, $y_i = ax_i^2 + bx_i + c + n_i$, to the data. Here, (a, b, c) are the parameters to be determined and n_i is the unknown observation noise. The (x_i, y_i) points are contained in a file hw1data.mat available on the course webpage. You may load the data to MATLAB using the command load hw1data and view them using scatter(x,y,'+'). Please use the same data set and find the maximum likelihood estimate of (a, b, c) assuming n_i's are i.i.d., and

   (a) $n_i \sim \mathcal{N}(0, 1)$;

$$\begin{aligned}
loss &= l = log\Pi_i p(z_i) \\
z_i &= y_i - (ax_i^2 + bx_i + c) \\
p(z_i) &= \frac{1}{(2\pi\sigma^2)^{0.5}} exp(\frac{-z^2}{2\sigma^2}) \\
l &= \sum_i logp(z_i) \\
l &= \sum_i log(\frac{1}{(2\pi\sigma^2)^{0.5}} exp(\frac{-z_i^2}{2\sigma^2})) \\
l &= \sum_i -\frac{1}{2} log(2\pi\sigma^2) + \frac{-z_i^2}{2\sigma^2} \\
l &= \sum_i -\frac{1}{2} log(2\pi) + \frac{-z_i^2}{2\sigma^2} \\
\frac{\partial l}{\partial a} &= \sum_i -\frac{1}{\sigma^2} z_i(-x_i^2) = 0 \\
\frac{\partial l}{\partial b} &= \sum_i -\frac{1}{\sigma^2} z_i(-x_i) = 0 \\
\frac{\partial l}{\partial c} &= \sum_i -\frac{1}{\sigma^2} z_i(-1) = 0
\end{aligned}$$

expand and rearrange:

$$\begin{bmatrix} \sum_i x_i^4 & \sum_i x_i^3 & \sum_i x_i^2 \\ \sum_i x_i^2 & \sum_i x_i & \sum_i 1 \\ \sum_i x_i^3 & \sum_i x_i^2 & \sum_i x_i \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum_i y_i x_i^2 \\ \sum_i y_i \\ \sum_i y_i x_i \end{bmatrix}$$

solve:

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 0.0120 \\ 3.1288 \\ -42.3550 \end{bmatrix}$$

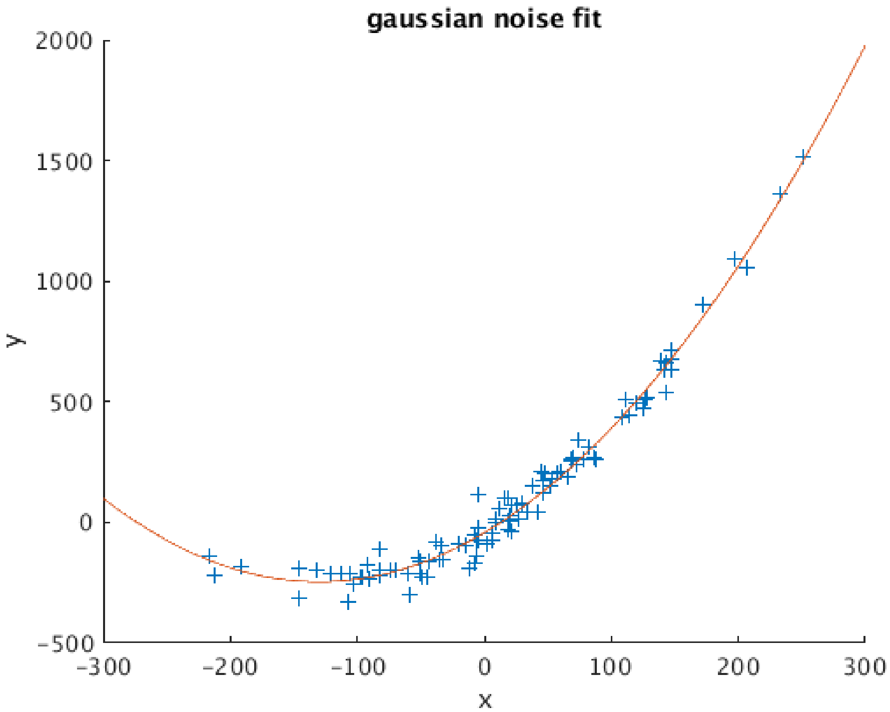

Figure 1: Data and Fitted Curve

(b) n_i is always positive and $n_i \sim e^{-z}$ for $z \geq 0$.

$$l = log\Pi_i p(z_i), z_i \geq 0$$
$$l = log\Pi_i exp(-z_i)$$
$$l = \sum_i logexp(-z_i)$$
$$l = \sum_i -z_i$$
$$\max_{a,b,c} l = \min_{a,b,c} -l = \sum_i z_i,\ s.t.\ z_i \geq 0$$

linear programming formulation:

$$\min -l = \min \left[-\sum_i x_i^2 \quad -\sum_i x_i \quad -length(x)\right] \begin{bmatrix} a \\ b \\ c \end{bmatrix},\ s.t.:$$

$$\left[power\_elementwise(x,2) \quad x \quad ones(length(x),1)\right] \begin{bmatrix} a \\ b \\ c \end{bmatrix} \leq y$$

solve:

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 0.0113 \\ 3.1734 \\ -153.6072 \end{bmatrix}$$

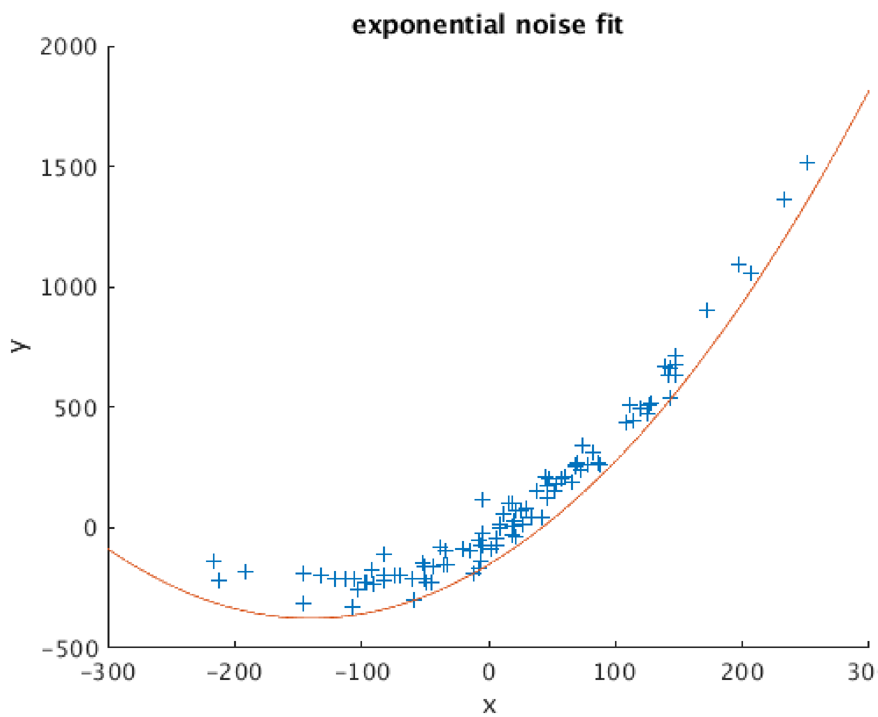

Figure 2: Data and Fitted Curve

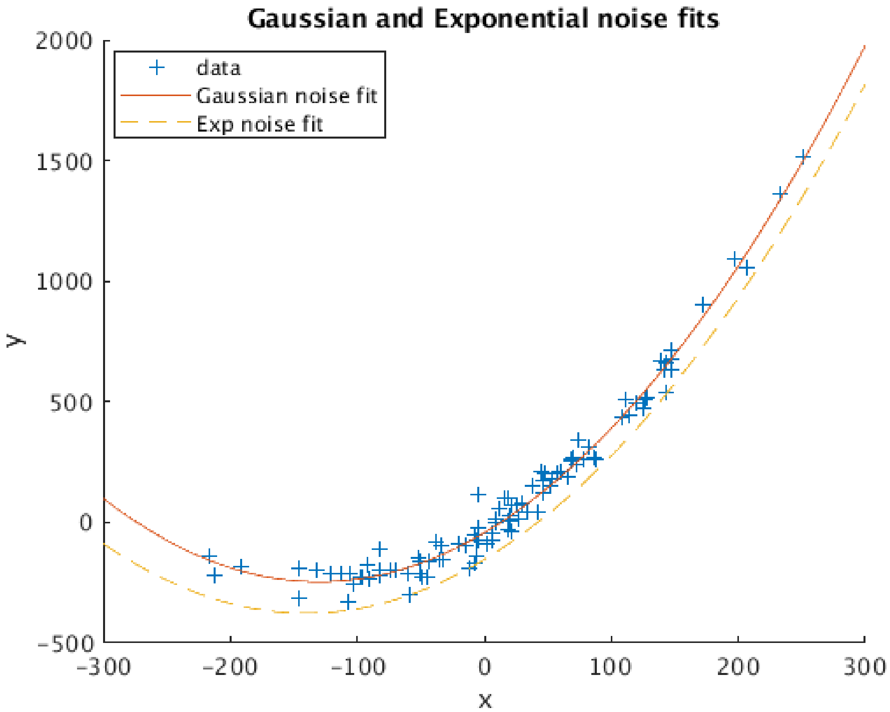

Figure 3: Data and Fitted Curves