Optimization for Machine Learning (CSL4010)

Dr. Md Abu Talhamainuddin Ansary
Department of Mathematics, IIT Jodhpur

PPT 2

Convex Set and Convex Function

Convex optimization problem

- Consider the optimization problem

$$(P): \quad \min_{x \in X \subseteq \mathbb{R}^n} f(x)$$

- (P) is a convex optimization problem if f is a convex function and X is a convex set.

Convex optimization problem

- Consider the optimization problem

$$(P): \quad \min_{x \in X \subseteq \mathbb{R}^n} \; f(x)$$

- (P) is a convex optimization problem if f is a convex function and X is a convex set.
- Consider the problem

$$\begin{aligned} \min \;\; & (x_1 - 2)^2 + (x_2 - 2)^2 \\ s.t. \;\; & x_1^2 + x_2^2 = 1 \end{aligned}$$

Convex optimization problem

- Consider the optimization problem

$$(P): \quad \min_{x \in X \subseteq \mathbb{R}^n} f(x)$$

- (P) is a convex optimization problem if f is a convex function and X is a convex set.
- Consider the problem

$$\begin{aligned} \min \ & (x_1 - 2)^2 + (x_2 - 2)^2 \\ s.t. \ & x_1^2 + x_2^2 = 1 \end{aligned}$$

- Set of feasible points is $X = \{x \in \mathbb{R}^2 | x_1^2 + x_2^2 = 1\}$, which is not a convex set. Hence it is not a convex optimization problem.

- The problem

$$\begin{aligned} \min \quad & f(x) \\ s.t. \quad & h_j(x) = 0 \qquad j = 1, 2, \ldots, p \end{aligned}$$

  is not a convex function if any h_j is nonlinear.

- The above problem will be a convex function if $h_j(x) = {a^j}^T x + b_j$ for all j.

- The problem

$$\begin{aligned} \min \quad & f(x) \\ s.t. \quad & h_j(x) = 0 \qquad j = 1, 2, \ldots, p \end{aligned}$$

is not a convex function if any h_j is nonlinear.

- The above problem will be a convex function if $h_j(x) = {a^j}^T x + b_j$ for all j.
- A convex optimization problem with equality constraints can be written as

$$\begin{aligned} \min \quad & f(x) \\ s.t. \quad & Ax = b \end{aligned}$$

where $A \in M_{p\times n}$, $b \in \mathbb{R}^p$.

- Consider the problem

$$\begin{aligned} \min\ & (x_1-2)^2+(x_2-2)^2 \\ s.t.\ & x_1^2+x_2^2 \le 1 \end{aligned}$$

- Consider the problem

$$\min \ (x_1 - 2)^2 + (x_2 - 2)^2$$
$$s.t. \ x_1^2 + x_2^2 \leq 1$$

- Set of feasible points is $X = \{x \in \mathbb{R}^2 | x_1^2 + x_2^2 \leq 1\}$ is a convex set. Hence it is a convex optimization problem.

- Consider the problem

$$\begin{aligned} \min \ & (x_1-2)^2+(x_2-2)^2 \\ s.t. \ & x_1^2+x_2^2 \le 1 \end{aligned}$$

- Set of feasible points is $X=\{x\in\mathbb{R}^2|x_1^2+x_2^2\le 1\}$ is a convex set. Hence it is a convex optimization problem.
- The problem

$$\begin{aligned} \min \ & f(x) \\ s.t. \ & g_i(x)\le 0 \qquad i=1,2,\ldots,m \end{aligned}$$

will be a convex optimization problem if g_i is convex for all i.

- Consider the problem

$$\min \ (x_1 - 2)^2 + (x_2 - 2)^2$$
$$s.t. \ x_1^2 + x_2^2 \leq 1$$

- Set of feasible points is $X = \{x \in \mathbb{R}^2 | x_1^2 + x_2^2 \leq 1\}$ is a convex set. Hence it is a convex optimization problem.
- The problem

$$\min \ f(x)$$
$$s.t. \ g_i(x) \leq 0 \qquad i = 1, 2, \ldots, m$$

will be a convex optimization problem if g_i is convex for all i.

- The problem

$$\begin{aligned} \min \quad & f(x) \\ s.t. \quad & g_i(x) \leq 0 \qquad i = 1, 2, \ldots, m \\ & h_j(x) = 0 \qquad j = 1, 2, \ldots, p \end{aligned}$$

  will be a convex optimization problem if
  - f is a convex function
  - g_i is convex for all i
  - h_j is affine ($h_j(x) = {a^j}^T x - b_j$)

Global/local minima

- Consider the optimization problem

$$(P): \quad \min_{x \in X \subseteq \mathbb{R}^n} f(x)$$

Global/local minima

- Consider the optimization problem

$$(P): \quad \min_{x \in X \subseteq \mathbb{R}^n} \; f(x)$$

- $x^* \in X$ is said to be a:
  - global minima of f on X if $f(x^*) \leq f(x)$ for all $x \in X$.
  - strict global minima of f on X if $f(x^*) < f(x)$ for all $x \in X \setminus \{x^*\}$.
  - local minima of f if there exists a neighbourhood of x^* ($N(x^*, \epsilon), \epsilon > 0$) such that $f(x^*) \leq f(x)$ for all $x \in X \cap N(x^*, \epsilon)$
  - strict local minima of f if there exists a neighbourhood of x^* ($N(x^*, \epsilon), \epsilon > 0$) such that $f(x^*) < f(x)$ for all $x \in X \cap N(x^*, \epsilon) \setminus \{x^*\}$

Lemma 1

Every local minima of a convex optimization problem is a global minima.

Lemma 1

Every local minima of a convex optimization problem is a global minima.

Proof:

- Suppose x^* is a local minima of $\min_{x \in X} \; f(x)$.
- Then there exists $N(x^*, \epsilon)$ such that $f(x^*) \leq f(x)$ for all $x \in X \cap N(x^*, \epsilon)$.

Lemma 1

Every local minima of a convex optimization problem is a global minima.

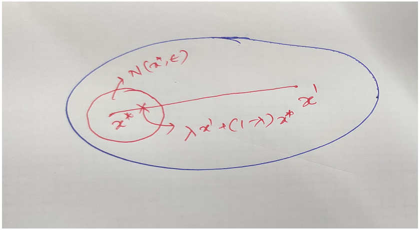
Proof:

- Suppose x^* is a local minima of $\min_{x \in X} f(x)$.
- Then there exists $N(x^*, \epsilon)$ such that $f(x^*) \leq f(x)$ for all $x \in X \cap N(x^*, \epsilon)$.
- If possible suppose x^* is not a global minima. Then there exists x^1 in X such that $f(x^1) < f(x^*)$.

Lemma 1

Every local minima of a convex optimization problem is a global minima.

Proof:

- Suppose x^* is a local minima of $\min_{x \in X} \; f(x)$.
- Then there exists $N(x^*, \epsilon)$ such that $f(x^*) \leq f(x)$ for all $x \in X \cap N(x^*, \epsilon)$.
- If possible suppose x^* is not a global minima. Then there exists x^1 in X such that $f(x^1) < f(x^*)$.
- 

- Since X is a convex set, $\lambda x^1 + (1-\lambda)x^* \in N(x^*, \epsilon)$ for sufficiently small $\lambda > 0$.

- Since X is a convex set, $\lambda x^1 + (1-\lambda)x^* \in N(x^*, \epsilon)$ for sufficiently small $\lambda > 0$.
- Since $\lambda x^1 + (1-\lambda)x^* \in N(x^*, \epsilon)$

$$f(x^*) \leq f\left(\lambda x^1 + (1-\lambda)x^*\right) \tag{1}$$

- Since f is a convex function

$$f\left(\lambda x^1 + (1-\lambda)x^*\right) \leq \lambda f(x^1) + (1-\lambda)f(x^*)$$

- Since X is a convex set, $\lambda x^1 + (1-\lambda)x^* \in N(x^*, \epsilon)$ for sufficiently small $\lambda > 0$.
- Since $\lambda x^1 + (1-\lambda)x^* \in N(x^*, \epsilon)$

$$f(x^*) \leq f\left(\lambda x^1 + (1-\lambda)x^*\right) \tag{1}$$

- Since f is a convex function

$$f\left(\lambda x^1 + (1-\lambda)x^*\right) \leq \lambda f(x^1) + (1-\lambda)f(x^*)$$

- Then from (1),

$$f(x^*) \leq \lambda f(x^1) + (1-\lambda)f(x^*)$$

This implies $f(x^*) \leq f(x^1)$, a contradiction.

Lemma 2

If f is a strictly convex function then $\min_{x \in X} f(x)$ has unique solution.

Lemma 2

If f is a strictly convex function then $\min_{x \in X} f(x)$ *has unique solution.*

Proof:

- If possible, suppose $\min_{x \in X} f(x)$ has two optimal solution $x^{1,*}, x^{2,*}$ such that $f(x^{1,*} = f(x^{2,*}) = f^*$.
- Then

$$f(1/2x^{1,*} + 1/2x^{2,*} <$$

a contradiction.

- Consider the unconstrained optimization problem $\min_{x \in \mathbb{R}^n} f(x)$ where f is a (strictly) convex function.

- Consider the unconstrained optimization problem $\min_{x \in \mathbb{R}^n} f(x)$ where f is a (strictly) convex function.
- Suppose $\nabla f(x^*) = 0$ for some $x^* \in \mathbb{R}^n$.

- Consider the unconstrained optimization problem $\min_{x \in \mathbb{R}^n} f(x)$ where f is a (strictly) convex function.
- Suppose $\nabla f(x^*) = 0$ for some $x^* \in \mathbb{R}^n$.
- Then for any y,

$$f(y)(>) \geq$$

- Consider the unconstrained optimization problem $\min_{x \in \mathbb{R}^n} f(x)$ where f is a (strictly) convex function.
- Suppose $\nabla f(x^*) = 0$ for some $x^* \in \mathbb{R}^n$.
- Then for any y,

$$f(y)(>) \geq$$

- Hence x^* is (unique) global minimizer.

Least square problem

- Consider the least square problem

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2}\|Ax - y\|^2 \tag{2}$$

Least square problem

- Consider the least square problem

$$\min_{x \in \mathbb{R}^n} \; f(x) = \frac{1}{2}\|Ax - y\|^2 \qquad (2)$$

- Clearly $\nabla f(x) = A^T Ax - A^T y$ and $\nabla^2 f(x) = A^T A$.

Least square problem

- Consider the least square problem

$$\min_{x \in \mathbb{R}^n} \; f(x) = \frac{1}{2}\|Ax - y\|^2 \tag{2}$$

- Clearly $\nabla f(x) = A^T Ax - A^T y$ and $\nabla^2 f(x) = A^T A$.
- For any $x \in \mathbb{R}^n$, $x^T(A^T A)x = (Ax)^T(Ax) = \|Ax\|^2 \geq 0$. This implies f is a convex function.

Least square problem

- Consider the least square problem

$$\min_{x\in\mathbb{R}^n} f(x) = \frac{1}{2}\|Ax - y\|^2 \quad (2)$$

- Clearly $\nabla f(x) = A^T Ax - A^T y$ and $\nabla^2 f(x) = A^T A$.
- For any $x \in \mathbb{R}^n$, $x^T(A^T A)x = (Ax)^T(Ax) = \|Ax\|^2 \geq 0$. This implies f is a convex function.
- **Lemma:** f will be strictly convex if column vectors of A are linearly independent.

Least square problem

- Consider the least square problem

$$\min_{x\in\mathbb{R}^n} f(x) = \frac{1}{2}\|Ax - y\|^2 \tag{2}$$

- Clearly $\nabla f(x) = A^T Ax - A^T y$ and $\nabla^2 f(x) = A^T A$.
- For any $x \in \mathbb{R}^n$, $x^T(A^T A)x = (Ax)^T(Ax) = \|Ax\|^2 \geq 0$. This implies f is a convex function.
- **Lemma:** f will be strictly convex if column vectors of A are linearly independent.
- **Proof:** Suppose $A = [a^1, a^2, \dots, a^n]$, then $Ax = a^1 x_1 + a^2 x_2 + \cdots + a^n x_n$. Now if $\{a^1, a^2, \dots, a^n\}$ is linearly independent then $Ax = 0$ if and only if $x = 0$. Then for any nonzero x, $x^T(A^T A)x = \|Ax\|^2 > 0$. Hence $\nabla^2 f$ is positive definite. So f is strictly convex.

- Since f is convex x^* will be a minima iff $\nabla f(x^*) = 0$. This implies

$$x^* = (A^T A)^{-1}(A^T y).$$

Since $A^T A$ is positive definite $det(A^T A) > 0$, so $(A^T A)^{-1}$ exists.

- For any $x \in \mathbb{R}^n$,

$$\begin{aligned}
\|Ax - y\|^2 &= \|Ax - Ax^* + Ax^* - y\|^2 \\
&= \|Ax - Ax^*\|^2 + \|Ax^* - y\|^2 + 2(Ax - Ax^*)^T(Ax^* - y) \\
&= \|Ax - Ax^*\|^2 + \|Ax^* - y\|^2 + (x - x^*)^T(A^T Ax^* - A^T y) \\
&\geq \|Ax^* - y\|^2.
\end{aligned}$$

The last inequality follows since $x^* = (A^T A)^{-1} A^T y$ implies $(A^T Ax^* - A^T y) = 0$.

- For any $x \in \mathbb{R}^n$,

$$\begin{aligned}\|Ax - y\|^2 &= \|Ax - Ax^* + Ax^* - y\|^2 \\ &= \|Ax - Ax^*\|^2 + \|Ax^* - y\|^2 + 2(Ax - Ax^*)^T(Ax^* - y) \\ &= \|Ax - Ax^*\|^2 + \|Ax^* - y\|^2 + (x - x^*)^T(A^TAx^* - A^Ty) \\ &\geq \|Ax^* - y\|^2.\end{aligned}$$

  The last inequality follows since $x^* = (A^TA)^{-1}A^Ty$ implies $(A^TAx^* - A^Ty) = 0$.

- This shows x^* is the minima (arg min) of the optimization problem (2).

- For any $x \in \mathbb{R}^n$,

$$\begin{aligned} \|Ax - y\|^2 &= \|Ax - Ax^* + Ax^* - y\|^2 \\ &= \|Ax - Ax^*\|^2 + \|Ax^* - y\|^2 + 2(Ax - Ax^*)^T(Ax^* - y) \\ &= \|Ax - Ax^*\|^2 + \|Ax^* - y\|^2 + (x - x^*)^T(A^TAx^* - A^Ty) \\ &\geq \|Ax^* - y\|^2. \end{aligned}$$

  The last inequality follows since $x^* = (A^TA)^{-1}A^Ty$ implies $(A^TAx^* - A^Ty) = 0$.
- This shows x^* is the minima (arg min) of the optimization problem (2).
- Suppose $A = QR$ be the QR-factorization, then we can show that least square solution can be written as $x^* = R^{-1}Q^Ty$ (prove it).

Application of least square problem

- Consider the data set

Memory (GB)	Ram (Ram)	Price (thousand rs)
32	2	7
32	3	8
64	4	10
128	8	19.5
256	8	25

Table 1: Data set based on mobile quality and price

- From Table 1, we can construct least square problem with

$$A = \begin{bmatrix} 32 & 2 \\ 32 & 3 \\ 64 & 4 \\ 128 & 8 \\ 256 & 8 \end{bmatrix} \text{ and } y = (7, 8, 10, 19.5, 25)^T.$$

- From Table 1, we can construct least square problem with $A = \begin{bmatrix} 32 & 2 \\ 32 & 3 \\ 64 & 4 \\ 128 & 8 \\ 256 & 8 \end{bmatrix}$ and $y = (7, 8, 10, 19.5, 25)^T$.
- Hence the solution of the least square problem is $x^* = (A^T A)^{-1} A^T y = (0.0354, 1.9784)^T$.

- From Table 1, we can construct least square problem with
$$A = \begin{bmatrix} 32 & 2 \\ 32 & 3 \\ 64 & 4 \\ 128 & 8 \\ 256 & 8 \end{bmatrix} \text{ and } y = (7, 8, 10, 19.5, 25)^T.$$
- Hence the solution of the least square problem is $x^* = (A^T A)^{-1} A^T y = (0.0354, 1.9784)^T$.
- For a new mobile with memory 512 GB and Ram 12 GB, the price will be $\hat{y} = \hat{a}^T x^* = 41.85$, where $\hat{a} = (512, 12)^T$.

Straight-line fitting (linear model) using least square problem

- Suppose $\mathcal{D} = \{(x_i, y_i); 1 = 1, 2, \ldots, N\}$ is a given we set of points. We need to find a straight line $y = mx + c$, such that $y_i \approx x_i$ for all i.

Straight-line fitting (linear model) using least square problem

- Suppose $\mathcal{D} = \{(x_i, y_i); 1 = 1, 2, \ldots, N\}$ is a given we set of points. We need to find a straight line $y = mx + c$, such that $y_i \approx x_i$ for all i.
- We have to find a vector $\bar{m}^T = (m, c)^T$ such that

$$\frac{1}{2N}\sum_{i=1}^{N}|mx_i + c - y_i|^2 = \frac{1}{2N}\sum_{i=1}^{N}|(x_i, 1)\bar{m} - y_i|^2$$

is minimum.

Straight-line fitting (linear model) using least square problem

- Suppose $\mathcal{D} = \{(x_i, y_i); 1 = 1, 2, \ldots, N\}$ is a given we set of points. We need to find a straight line $y = mx + c$, such that $y_i \approx x_i$ for all i.
- We have to find a vector $\bar{m}^T = (m, c)^T$ such that

$$\frac{1}{2N}\sum_{i=1}^{N}|mx_i + c - y_i|^2 = \frac{1}{2N}\sum_{i=1}^{N}|(x_i, 1)\bar{m} - y_i|^2$$

is minimum.

- Define $A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \cdots & \\ x_N & 1 \end{bmatrix}$ and $y = (y_1, y_2, \ldots, y_N)^T$. Then the least square problem become

$$\min_{\bar{m}} \quad \frac{1}{2N}\|A\bar{m} - y\|^2.$$

- Using least square technique we can find the solution of the above optimization problem as $\bar{m}^* = (m^*, c^*)^T = (A^TA)^{-1}A^Ty$.

Straight-line fitting (linear model) using least square problem

- Suppose $\mathcal{D} = \{(x_i, y_i); 1 = 1, 2, \ldots, N\}$ is a given we set of points. We need to find a straight line $y = mx + c$, such that $y_i \approx x_i$ for all i.
- We have to find a vector $\bar{m}^T = (m, c)^T$ such that

$$\frac{1}{2N}\sum_{i=1}^{N} |mx_i + c - y_i|^2 = \frac{1}{2N}\sum_{i=1}^{N} |(x_i, 1)\bar{m} - y_i|^2$$

  is minimum.
- Define $A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \ldots & \\ x_N & 1 \end{bmatrix}$ and $y = (y_1, y_2, \ldots, y_N)^T$. Then the least square problem become

$$\min_{\bar{m}} \quad \frac{1}{2N}\|A\bar{m} - y\|^2.$$

- Using least square technique we can find the solution of the above optimization problem as $\bar{m}^* = (m^*, c^*)^T = (A^T A)^{-1} A^T y$.
- So the best fitting straight line is $y = m^* x + c^*$.

Polynomial fitting (nonlinear model) using least square problem

- Suppose $\mathcal{D} = \{(x_i, y_i); 1 = 1, 2, \dots, N\}$ is a given we set of points. We need to find a p degree polynomial line

$$y = \beta_b x^p + \beta_{p-1} x^{p-1} + \cdots + \beta_1 x + \beta_0$$

such that $y_i \approx x_i$ for all i.

Polynomial fitting (nonlinear model) using least square problem

- Suppose $\mathcal{D} = \{(x_i, y_i); 1 = 1, 2, \ldots, N\}$ is a given we set of points. We need to find a p degree polynomial line

$$y = \beta_b x^p + \beta_{p-1} x^{p-1} + \cdots + \beta_1 x + \beta_0$$

such that $y_i \approx x_i$ for all i.

- We have to find a vector $\beta = (\beta_p, \beta_{p-1}, \ldots, \beta_1, \beta_0)^T$ such that

$$\frac{1}{2N} \sum_{i=1}^{N} |\beta_b x^p + \beta_{p-1} x^{p-1} + \cdots + \beta_1 x + \beta_0 - y_i|^2$$
$$= \sum_{i=1}^{N} \left|(x_i^p, x_i^{p-1}, \ldots, x_i, 1)\beta - y_i\right|^2$$

is minimum.

Polynomial fitting (nonlinear model) using least square problem contd...

- Define $A = \begin{bmatrix} x_1^p & x_1^{p-1} & \dots & x_1 & 1 \\ x_2^2 & x_2^{p-1} & \dots & x_2 & 1 \\ \dots \\ x_N^p & x_N^{p-1} & \dots & x_N & 1 \end{bmatrix}$ and $y = (y_1, y_2, \dots, y_N)^T$. Then the least square problem become

$$\min_{\beta} \quad \frac{1}{2N} \|A\beta - y\|^2.$$

Polynomial fitting (nonlinear model) using least square problem contd...

- Define $A = \begin{bmatrix} x_1^p & x_1^{p-1} & \dots & x_1 & 1 \\ x_2^2 & x_2^{p-1} & \dots & x_2 & 1 \\ \dots \\ x_N^p & x_N^{p-1} & \dots & x_N & 1 \end{bmatrix}$ and $y = (y_1, y_2, \dots, y_N)^T$. Then the least square problem become

$$\min_{\beta} \ \frac{1}{2N} \|A\beta - y\|^2.$$

- Using least square technique we can find the solution of the above optimization problem as $\beta^* = (A^T A)^{-1} A^T y$.

Polynomial fitting (nonlinear model) using least square problem contd...

- Define $A = \begin{bmatrix} x_1^p & x_1^{p-1} & \dots & x_1 & 1 \\ x_2^2 & x_2^{p-1} & \dots & x_2 & 1 \\ \dots & & & & \\ x_N^p & x_N^{p-1} & \dots & x_N & 1 \end{bmatrix}$ and $y = (y_1, y_2, \dots, y_N)^T$. Then the least square problem become

$$\min_{\beta} \quad \frac{1}{2N} \|A\beta - y\|^2.$$

- Using least square technique we can find the solution of the above optimization problem as $\beta^* = (A^T A)^{-1} A^T y$.
- So the best fitting curve is

$$y = {\beta^*}_p x^p + {\beta^*}_{p-1} x^{p-1} + \dots + {\beta^*}_1 x + {\beta^*}_0.$$

Exponential curve fitting using least square problem

- Suppose $\mathcal{D} = \{(x_i, y_i); 1 = 1, 2, \ldots, N\}$ is a given we set of points. We need to find a curve
$$y = \beta_0 e^{\beta_1 x}$$
, such that $y_i \approx x_i$ for all i

Exponential curve fitting using least square problem

- Suppose $\mathcal{D} = \{(x_i, y_i); 1 = 1, 2, \ldots, N\}$ is a given we set of points. We need to find a curve
$$y = \beta_0 e^{\beta_1 x}$$
, such that $y_i \approx x_i$ for all i.
- Taking log of both sides of $y = \beta_0 e^{\beta_1 x}$, we have $\log(y) = \log(\beta_0) + \beta_1 x$.

Exponential curve fitting using least square problem

- Suppose $\mathcal{D} = \{(x_i, y_i); 1 = 1, 2, \ldots, N\}$ is a given we set of points. We need to find a curve
$$y = \beta_0 e^{\beta_1 x}$$
, such that $y_i \approx x_i$ for all i.
- Taking log of both sides of $y = \beta_0 e^{\beta_1 x}$, we have $\log(y) = \log(\beta_0) + \beta_1 x$.
- Define $\log(y) = z$ and $\log(\beta_0) = \bar{\beta}_0$. So the above equation becomes $z = \beta_1 x + \bar{\beta}_0$.

Exponential curve fitting using least square problem

- Suppose $\mathcal{D} = \{(x_i, y_i); 1 = 1, 2, \ldots, N\}$ is a given we set of points. We need to find a curve
$$y = \beta_0 e^{\beta_1 x}$$
, such that $y_i \approx x_i$ for all i.
- Taking log of both sides of $y = \beta_0 e^{\beta_1 x}$, we have $\log(y) = \log(\beta_0) + \beta_1 x$.
- Define $\log(y) = z$ and $\log(\beta_0) = \bar{\beta}_0$. So the above equation becomes $z = \beta_1 x + \bar{\beta}_0$.
- We have to find a vector $b\bar{e}ta^T = (\beta_1, \bar{\beta}_0)^T$ such that
$$\frac{1}{2N}\sum_{i=1}^{N} |\beta_1 x_i + \bar{\beta}_0 - z_i|^2 = \sum_{i=1}^{N} \left|(x_i, 1)b\bar{e}ta - z_i\right|^2$$
is minimum.

Exponential curve fitting using least square problem

- Define $A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \dots & \\ x_N & 1 \end{bmatrix}$ and $z = (z_1, z_2, \dots, z_N)^T$. Then the least square problem become

$$\min_{\bar{\beta}} \quad \frac{1}{2N} \|A\bar{\beta} - z\|^2.$$

Exponential curve fitting using least square problem

- Define $A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \dots & \\ x_N & 1 \end{bmatrix}$ and $z = (z_1, z_2, \dots, z_N)^T$. Then the least square problem become

$$\min_{\bar{\beta}} \quad \frac{1}{2N} \|A\bar{\beta} - z\|^2.$$

- Using least square technique we can find the solution of the above optimization problem as $\bar{\beta}^* = (A^T A)^{-1} A^T z$.

Exponential curve fitting using least square problem

- Define $A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \dots & \\ x_N & 1 \end{bmatrix}$ and $z = (z_1, z_2, \dots, z_N)^T$. Then the least square problem become
$$\min_{\bar{\beta}} \quad \frac{1}{2N} \|A\bar{\beta} - z\|^2.$$
- Using least square technique we can find the solution of the above optimization problem as $\bar{\beta}^* = (A^T A)^{-1} A^T z$.
- So the equation of best fitting exponential curve is
$$y = e^{\bar{\beta}_1^* x + \bar{\beta}^*{}_0}$$

Exponential curve fitting using least square problem

- Define $A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \dots & \\ x_N & 1 \end{bmatrix}$ and $z = (z_1, z_2, \dots, z_N)^T$. Then the least square problem become

$$\min_{\bar{\beta}} \quad \frac{1}{2N} \|A\bar{\beta} - z\|^2.$$

- Using least square technique we can find the solution of the above optimization problem as $\bar{\beta}^* = (A^T A)^{-1} A^T z$.
- So the equation of best fitting exponential curve is

$$y = e^{\bar{\beta}_1^* x + \bar{\beta}^*{}_0}$$

Python code for cubic curve fitting

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
dt=pd.read_excel(r'C:\Teaching & Research\Courses\2022-23'\
    r'\Optimization for ML\Lab assignments\excell sheets\xy.xl
B=dt.values
x,y=B[:,0],B[:,1]
A=np.empty((0,4),dtype=float)
n=len(x)
A=np.column_stack((np.column_stack((np.column_stack((pow(x.T,3
      pow(x.T,2))),x.T)),np.ones((len(x),1),dtype=float)))
#print(A)
beta=np.dot(np.linalg.inv(np.dot(A.T,A)),np.dot(A.T,y.T))
print(beta)
plt.figure(figsize = (10,12))
plt.plot(x, y, 'b.')
x1=np.linspace(0,5.2,num=201)
plt.plot(x1, np.polyval(beta,x1), 'r')
plt.xlabel('x')
plt.ylabel('y')
plt.show()
```

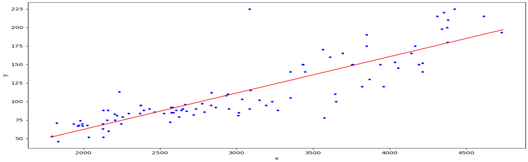


Figure 1: Fitting the data set by a straight line (linear curve)

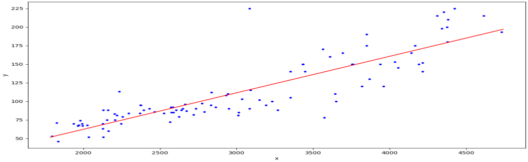


Figure 1: Fitting the data set by a straight line (linear curve)

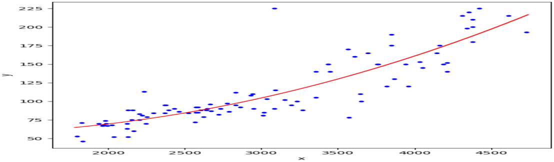


Figure 2: Fitting the data set by a quadratic polynomial curve

Data fitting...

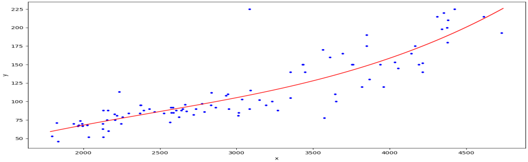


Figure 3: Fitting the data set by a cubic polynomial

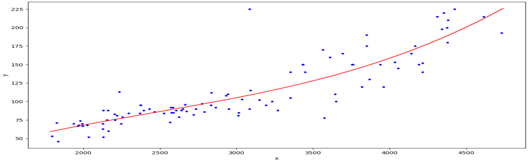


Figure 3: Fitting the data set by a cubic polynomial

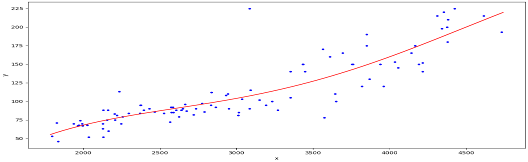


Figure 4: Fitting the data set by a bi-quadratic (4 degree) polynomial curve

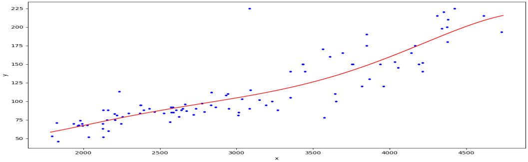


Figure 5: Fitting the data set by a cubic polynomial

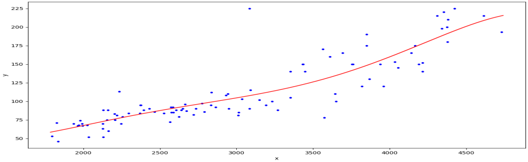


Figure 5: Fitting the data set by a cubic polynomial

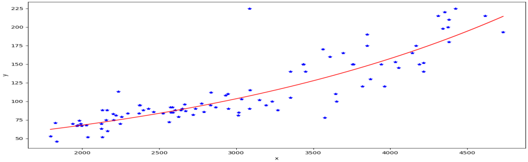


Figure 6: Fitting the data set by an exponential curve

- Boyd, S. and Vandenberghe, L.: Convex optimization. Cambridge university press, 2004.
- Boyd, S. and Vandenberghe L.: Introduction to applied linear algebra: vectors, matrices, and least squares. Cambridge university press, 2018.