18-660: Numerical Methods for Engineering Design and Optimization

Xin Li
Department of ECE
Carnegie Mellon University
Pittsburgh, PA 15213
Overview

- Linear Regression
  - Ordinary least-squares regression
  - Minimax optimization
  - Design of experiments
Linear regression (also referred to as response surface modeling) is widely used for many engineering problems:

- We do not know the analytical form of $f(x)$
- But we can generate a set of sampling points for $f(x)$
- Fit an approximate function for $f(x)$ from these sampling points

$$f(x) \approx \alpha_1 \cdot b_1(x) + \alpha_2 \cdot b_2(x) + \cdots$$

$f(x)$ is approximated as the linear combination of multiple basis functions.
Linear Regression

- Major steps of linear regression
  - Select a model template (e.g., polynomial function)
  - Generate a number of sampling points
  - Compute performance values at these sampling points
  - Create a set of linear equations to solve model coefficients

- A simple example
  - \( f(x) = \exp(x), \ x \in [-1, 1] \)
  - We will use this simple example to show how we can generally build a regression model from sampling data
Linear Regression Example

- **Step 1:** select a model template

\[ f(x) \approx bx + c \]

- **Step 2:** generate a number of sampling points

<table>
<thead>
<tr>
<th>Samples</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>-1</td>
<td>-0.5</td>
<td>0</td>
<td>0.5</td>
<td>1</td>
</tr>
</tbody>
</table>

- **Step 3:** compute performance values at these sampling points

<table>
<thead>
<tr>
<th>Samples</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>f(x)</td>
<td>0.3679</td>
<td>0.6065</td>
<td>1.0000</td>
<td>1.6487</td>
<td>2.7183</td>
</tr>
</tbody>
</table>
Step 4: create linear equations for model coefficients

\[ f(x) \approx bx + c \]

<table>
<thead>
<tr>
<th>Samples</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>-1</td>
<td>-0.5</td>
<td>0</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>f(x)</td>
<td>0.3679</td>
<td>0.6065</td>
<td>1.0000</td>
<td>1.6487</td>
<td>2.7183</td>
</tr>
</tbody>
</table>

\[
\begin{bmatrix}
-1 & 1 \\
-0.5 & 1 \\
0 & 1 \\
0.5 & 1 \\
1 & 1
\end{bmatrix}
\begin{bmatrix}
b \\
c
\end{bmatrix}
= 
\begin{bmatrix}
0.3679 \\
0.6065 \\
1.0000 \\
1.6487 \\
2.7183
\end{bmatrix}
\]

i-th sampling point

x values

f(x) values
Step 5: solve over-determined linear equations

- # of equations is greater than # of coefficients – over-determined
- No exact solution exists to satisfy all equations, but we can find the least-squares solution:

\[ A \cdot \alpha = B \]

\[ \min_{\alpha} ||A \cdot \alpha - B||^2 \]

Ordinary least-squares (OLS) regression

For a vector \( \varepsilon \in \mathbb{R}^M \), \( ||\varepsilon||_2 \) is defined as:

\[ ||\varepsilon||_2 = \sqrt{\sum_{i=1}^{M} \varepsilon_i^2} \]
Linear Regression Example

\[ A \cdot \alpha = B \]

\[
\begin{bmatrix}
A & \cdot & \alpha - B
\end{bmatrix}
\begin{bmatrix}
\varepsilon_1 \\
\varepsilon_2 \\
\vdots \\
\varepsilon_M
\end{bmatrix}
\]

Error at the i-th sampling point

\[ \min_{\alpha} \left\| A \cdot \alpha - B \right\|^2_2 \]

\[ \min_{\alpha} \sum_{i=1}^{M} \varepsilon_i^2(\alpha) \]
There are several possible ways to solve over-determined linear equations for linear regression.

- We will explain these algorithms in detail in future lectures.
- For now, you can simply use “α = A\B” in MATLAB.
Linear Regression Example

- Step 5: solve over-determined linear equations

\[
\begin{bmatrix}
-1 & 1 \\
-0.5 & 1 \\
0 & 1 \\
0.5 & 1 \\
1 & 1
\end{bmatrix}
\begin{bmatrix}
b \\
c
\end{bmatrix}
= \begin{bmatrix}
0.3679 \\
0.6065 \\
1.0000 \\
1.6487 \\
2.7183
\end{bmatrix}
\]

\[b = 1.1486\]

\[c = 1.2683\]

Linear model results in large error
What if we build a quadratic model for \( y = \exp(x) \)?

- Select a model template

\[
f(x) \approx ax^2 + bx + c
\]

- Generate a number of sampling points

<table>
<thead>
<tr>
<th>Samples</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x)</td>
<td>-1</td>
<td>-0.5</td>
<td>0</td>
<td>0.5</td>
<td>1</td>
</tr>
</tbody>
</table>

- Compute performance values at these sampling points

<table>
<thead>
<tr>
<th>Samples</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>(f(x))</td>
<td>0.3679</td>
<td>0.6065</td>
<td>1.0000</td>
<td>1.6487</td>
<td>2.7183</td>
</tr>
</tbody>
</table>
Quadratic Model Example

Create a set of linear equations to solve model coefficients

\[ f(x) \approx ax^2 + bx + c \]

<table>
<thead>
<tr>
<th>Samples</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>-1</td>
<td>-0.5</td>
<td>0</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>f(x)</td>
<td>0.3679</td>
<td>0.6065</td>
<td>1.0000</td>
<td>1.6487</td>
<td>2.7183</td>
</tr>
</tbody>
</table>

\[
\begin{bmatrix}
1 & -1 & 1 \\
0.25 & -0.5 & 1 \\
0 & 0 & 1 \\
0.25 & 0.5 & 1 \\
1 & 1 & 1
\end{bmatrix}
\begin{bmatrix}
a \\
b \\
c
\end{bmatrix}
= 
\begin{bmatrix}
0.3679 \\
0.6065 \\
1.0000 \\
1.6487 \\
2.7183
\end{bmatrix}
\]

\[ x^2 \quad x \quad \text{f(x) values} \]
Quadratic Model Example

Build quadratic model for $y = \exp(x)$

$$
\begin{bmatrix}
1 & -1 & 1 \\
0.25 & -0.5 & 1 \\
0 & 0 & 1 \\
0.25 & 0.5 & 1 \\
1 & 1 & 1
\end{bmatrix}
\begin{bmatrix}
a \\
b \\
c
\end{bmatrix}
= 
\begin{bmatrix}
0.3679 \\
0.6065 \\
1.0000 \\
1.6487 \\
2.7183
\end{bmatrix}
$$

\[
a = 0.5477 \\
b = 1.1486 \\
c = 0.9944
\]

Quadratic model results in much better accuracy in this example
Linear Model vs. Quadratic Model

**Linear RSM**  \[ \exp(x) \approx 1.1486x + 1.2683 \]

**Quadratic RSM**  \[ \exp(x) \approx 0.5477x^2 + 1.1486x + 0.9944 \]

- **Regression model is different from direct Taylor expansion**
  - E.g., different constant terms in linear and quadratic models – they are selected to minimize the least-squares error.
Minimax Optimization

We can also solve over-determined linear equations to satisfy other optimality criteria (i.e., not ordinary least-squares)

\[ A \cdot \alpha = B \]

\[
\begin{bmatrix}
A(i,:) \cdot \alpha - B_i
\end{bmatrix}
\]

Minimize the maximal absolute error

Error at the i-th sampling point
Minimax Optimization

- Other optimality criteria can be similarly formulated

\[ A \cdot \alpha = B \]

- Minimize the maximal relative error

\[ \min_{\alpha} \max_i \left| \frac{A(i,:) \cdot \alpha - B_i}{B_i} \right| \]

These formulations are minimax optimization problems
General minimax problems are difficult to solve
- Cost function does not have continuous derivative
Minimax Optimization

- However, our minimax problem for regression modeling can be re-formulated into a special form.

- Consider the example of absolute error minimization:
  \[
  \min_{\alpha} \max_{i} |A(i,:) \cdot \alpha - B_i| 
  \]

  Introduce a slack variable \( t \)

  \[
  \min_{\alpha, t} t \\
  \text{S.T.} \\
  \left\{ \begin{array}{l}
  |A(1,:) \cdot \alpha - B_1| \leq t \\
  |A(2,:) \cdot \alpha - B_2| \leq t \\
  \vdots \\
  |A(M,:) \cdot \alpha - B_M| \leq t 
  \end{array} \right. 
  \]

  Subject to

  Cost function

  Constraints
Minimax Optimization

\[
\begin{align*}
\min_{\alpha, t} & \quad t \\
\text{S.T.} & \quad |A(1,:) \cdot \alpha - B_1| \leq t \\
& \quad |A(2,:) \cdot \alpha - B_2| \leq t \\
& \quad \vdots \\
& \quad |A(M,:) \cdot \alpha - B_M| \leq t
\end{align*}
\]

\[
\begin{align*}
\min_{\alpha, t} & \quad t \\
\text{S.T.} & \quad -t \leq A(1,:) \cdot \alpha - B_1 \leq t \\
& \quad -t \leq A(2,:) \cdot \alpha - B_2 \leq t \\
& \quad \vdots \\
& \quad -t \leq A(M,:) \cdot \alpha - B_M \leq t
\end{align*}
\]

- Re-written as a **linear programming (LP)** problem
  - Both cost function and constraints are linear
  - No closed-form solution exists for LP
  - Can be numerically solved by an efficient (i.e., low complexity) and robust (i.e., global convergence) algorithm
Design of Experiments (DOE)

- We already know the basics for linear regression

- Open problem:
  - How can we select few samples to achieve good accuracy?

- A bad linear model example: \( f(x_1, x_2) = a \cdot x_1 + b \cdot x_2 + c \)

Sampling points for linear model

\[
\begin{align*}
(x_1 = -1, \quad x_2 = 0, \quad f_1) \\
(x_1 = 0, \quad x_2 = 0, \quad f_2) \\
(x_1 = 1, \quad x_2 = 0, \quad f_3)
\end{align*}
\]
Design of Experiments (DOE)

Linear model example (continued)

\[ f(x_1, x_2) = a \cdot x_1 + b \cdot x_2 + c \]

\[
\begin{bmatrix}
-1 & 0 & 1 \\
0 & 0 & 1 \\
1 & 0 & 1 \\
\end{bmatrix}
\begin{bmatrix}
a \\
b \\
c \\
\end{bmatrix}
= 
\begin{bmatrix}
f_1 \\
f_2 \\
f_3 \\
\end{bmatrix}
\]

\[
\begin{bmatrix}
-1 & 0 & 1 \\
0 & 0 & 1 \\
1 & 0 & 1 \\
\end{bmatrix}
\begin{bmatrix}
(1 \ -1) \\
(0 \ 0) \\
(1 \ 1) \\
\end{bmatrix}
= 
\begin{bmatrix}
f_1 \\
f_2 \\
f_3 \\
\end{bmatrix}
\]

Singular matrix (cannot solve the coefficient b)
Design of Experiments (DOE)

- Linear model example (continued)

\[
\begin{array}{ccc}
-1 & 0 & 1 \\
\end{array}
\]

No variation is applied to \( x_2 \)

Add additional sampling points for \( x_2 \)
Design of Experiments (DOE)

A bad quadratic model example: \( f(x_1, x_2) = a_{11} \cdot x_1^2 + a_{12} \cdot x_1 \cdot x_2 + a_{22} \cdot x_2^2 + b_1 \cdot x_1 + b_2 \cdot x_2 + c \)

Sampling points for quadratic model:

- \((x_1 = 0, x_2 = 0, f_1)\)
- \((x_1 = 0, x_2 = -1, f_2)\)
- \((x_1 = 0, x_2 = 1, f_3)\)
- \((x_1 = -1, x_2 = 0, f_4)\)
- \((x_1 = 1, x_2 = 0, f_5)\)
Design of Experiments (DOE)

- Quadratic model example (continued)

\[
f(x_1, x_2) = a_{11} \cdot x_1^2 + a_{12} \cdot x_1 \cdot x_2 + a_{22} \cdot x_2^2 + b_1 \cdot x_1 + b_2 \cdot x_2 + c
\]

\[
\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & -1 & 1 & 1 \\
0 & 0 & 1 & 0 & 1 & 1 & 1 \\
1 & 0 & 0 & -1 & 0 & 1 & 1 \\
1 & 0 & 0 & 1 & 0 & 1 & 1
\end{bmatrix}
\begin{bmatrix}
a_{11} \\
a_{12} \\
a_{22} \\
b_1 \\
b_2 \\
c
\end{bmatrix}
= \begin{bmatrix}
f_1 \\
f_2 \\
f_3 \\
f_4 \\
f_5
\end{bmatrix}
\]

Singular matrix (cannot solve the coefficient \(a_{12}\))
Design of Experiments (DOE)

- Quadratic model example (continued)

Cross-product terms cannot be captured

Add additional sampling points for $x_1x_2$
Design of Experiments (DOE)

- Design of experiments (DOE) is a research area that studies how to optimally select sampling points for modeling.

- Given a model template (e.g., linear or quadratic function), optimize sampling points for certain optimal criterion.
  - E.g., maximize modeling accuracy.

- Numerical optimization may be required to find the optimal sampling scheme.

D. Montgomery, Design and Analysis of Experiments, John Wiley & Sons, 2004
Summary

- Linear regression
  - Ordinary least-squares regression
  - Minimax optimization
  - Design of experiments