Curve Fitting

CE 311 K - Introduction to Computer Methods

Daene C. McKinney

Curve Fitting

• Linear Regression – Normal Equations
• Polynomial Regression
• Nonlinear Transformations
Evaporation from Reservoir

\[ S_{t+1} = S_t + \dot{Q}_t - R_t - L_t \]

- \( L_t \): Losses from reservoir
- \( A \): Surface area of reservoir
- \( e_t \): Ave. evaporation rate

Curve Fitting

- Data at discrete points or times
- Want estimates at points between measurements
- Fit curve to data to estimate intermediate values
Interpolation

- Precise data – no error in $y$
- Force curve through each point

```
<table>
<thead>
<tr>
<th>Independent Variable, x</th>
<th>Dependent Variable, y</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>16</td>
<td>17</td>
</tr>
</tbody>
</table>
```

Regression

- Experimental Data
  - Noisy (contains errors or inaccuracies)
    - $x$ values are accurate, $y$ values are not
- Find general trend (relationship) between $x$ and $y = f(x)$
  - Without passing through any specific point

```
<table>
<thead>
<tr>
<th>Time (minutes)</th>
<th>Temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>14</td>
<td>7</td>
</tr>
<tr>
<td>16</td>
<td>8</td>
</tr>
</tbody>
</table>
```
Noisy Data From Experiment

<table>
<thead>
<tr>
<th>i</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>2.10</td>
<td>6.22</td>
<td>7.17</td>
<td>10.5</td>
<td>13.7</td>
</tr>
<tr>
<td>y</td>
<td>2.90</td>
<td>3.83</td>
<td>5.98</td>
<td>5.71</td>
<td>7.74</td>
</tr>
</tbody>
</table>

Least Squares Regression

- Minimize the residual between the data points and the line

Model: \( y = a_0 + a_1x \)

Estimate \( a_0 \) and \( a_1 \):

\[
y_i = a_0 + a_1 x_i \\
e_i = y_i - a_0 - a_1 x_i
\]

\[
S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2
\]

Find the values of \( a_0 \) and \( a_1 \) that minimize \( S_r \)
Least Squares Regression

- Minimize $S_r$ by taking derivatives WRT $a_0$ and $a_1$.

$$\frac{\partial S_r}{\partial a_0} = \frac{\partial}{\partial a_0}\left[\sum_{i=1}^{n}(y_i - a_0 - a_1 x_i)^2\right]$$

$$= \sum_{i=1}^{n}2(y_i - a_0 - a_1 x_i)(-1)$$

$$= 0$$

$$\frac{\partial S_r}{\partial a_1} = \frac{\partial}{\partial a_1}\left[\sum_{i=1}^{n}(y_i - a_0 - a_1 x_i)^2\right]$$

$$= \sum_{i=1}^{n}2(y_i - a_0 - a_1 x_i)(-x_i)$$

$$= 0$$

$$n a_0 + \left[\sum_{j=1}^{n} x_j\right] a_1 = \sum_{j=1}^{n} y_j$$

Normal Equations - Solution

$$a_0 = \frac{1}{n} \sum_{i=1}^{n} y_i - \frac{1}{n} \left(\sum_{i=1}^{n} x_i\right) \left(\sum_{i=1}^{n} x_i y_i\right)$$

$$a_1 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} x_i\right)^2$$
Example

\[\begin{array}{c|ccccc}
 i & 1 & 2 & 3 & 4 & 5 \\
 x_i & 2.10 & 6.22 & 7.17 & 10.5 & 13.7 \\
 y_i & 2.90 & 3.83 & 5.98 & 5.71 & 7.74 \\
\end{array}\]

\[
\sum_{i=1}^{5} x_i = 39.69 \\
\sum_{i=1}^{5} x_i^2 = 392.3201 \\
\sum_{i=1}^{5} y_i = 26.16 \\
\sum_{i=1}^{5} x_i y_i = 238.7416
\]

\[
a_0 = \frac{1}{5} (26.16)(392.3) - \frac{1}{5} (39.69)(238.7) - \frac{392.3 - \frac{1}{5} (39.69)^2}{4} = 2.038
\]

\[
a_1 = \frac{238.7 - \frac{1}{5} (39.69)(26.16)}{392.3 - \frac{1}{5} (39.69)^2} = 0.4023
\]

\[y = 2.038 + 0.4023x\]