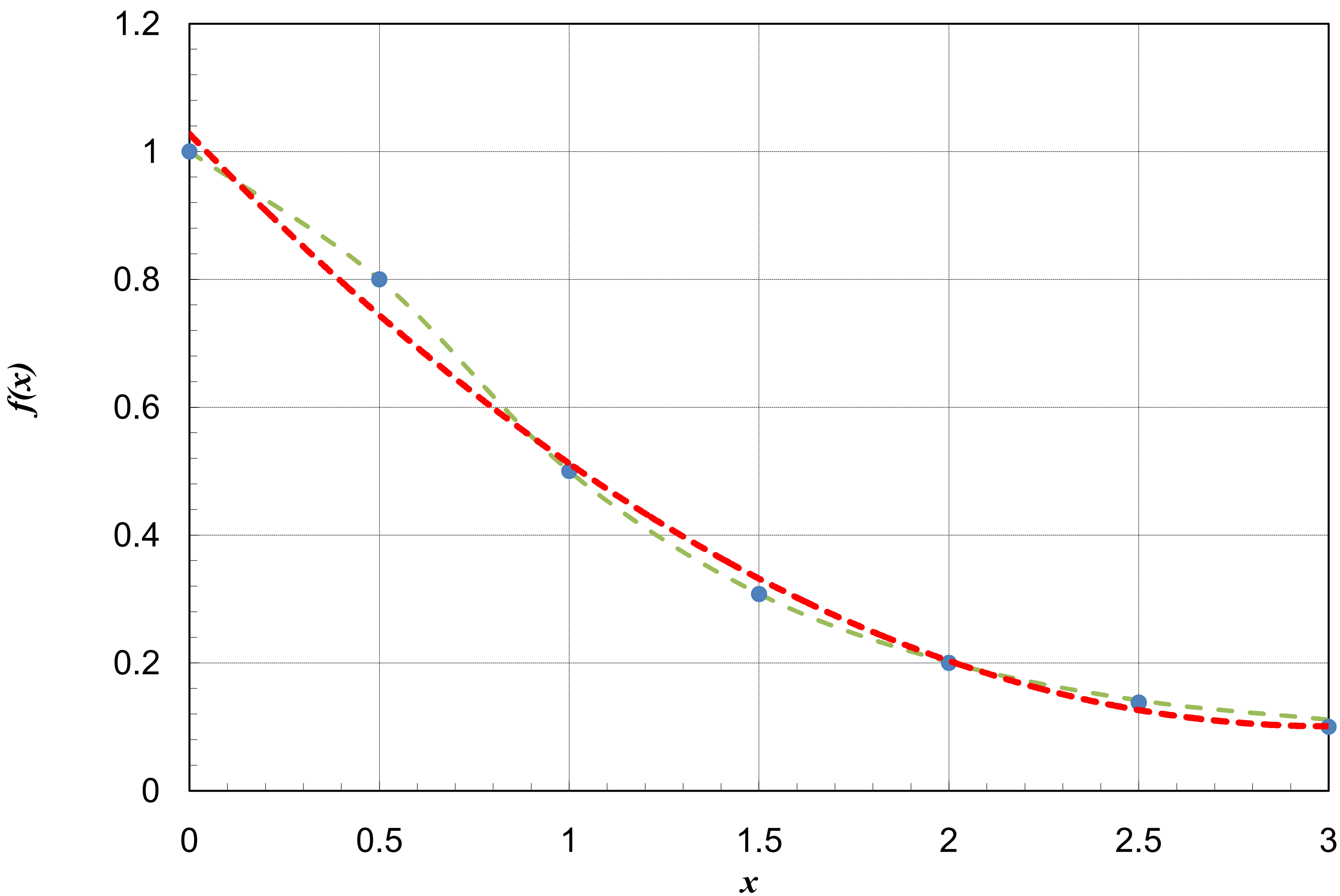


If the data, $f(x)$, may have uncertainty, we do not want the approximating function to pass through ALL data points. Regression minimizes the "error".

Regression

- Given $\left(x_k, f\left(x_k\right)\right) \quad k = 0,1,2,\ldots,n$
- Fit an approximating function such that it is "closest" to the data points
- Mostly polynomial, of degree m (m<n)
- Sometimes trigonometric functions
- As before, assume the approximation as

$$f_m(x) = \sum_{j=0}^{m} c_j \phi_j\left(x\right)$$

Regression: Least Squares

- Minimize the sum of squares of the difference between the function and the data:

$$\sum_{k=0}^{n}\left(f(x_k)-\sum_{j=0}^{m}c_j\phi_j(x_k)\right)^2$$

- Results in m+1 linear equations (that is why the term Linear Regression): $[A]\{c\}=\{b\}$. Called the *Normal Equations.*

$$a_{ij}=\sum_{k=0}^{n}\phi_i(x_k)\phi_j(x_k) \text{ and } b_i=\sum_{k=0}^{n}\phi_i(x_k)f(x_k)$$

Regression: Least Squares

- For example, using conventional form, $\phi_j = x^j$,

$$
\begin{bmatrix}
\sum_{k=0}^{n} 1 & \sum_{k=0}^{n} x_k & \sum_{k=0}^{n} x_k^2 & \dots & \sum_{k=0}^{n} x_k^m \\
\sum_{k=0}^{n} x_k & \sum_{k=0}^{n} x_k^2 & \sum_{k=0}^{n} x_k^3 & \dots & \sum_{k=0}^{n} x_k^{m+1} \\
\sum_{k=0}^{n} x_k^2 & \sum_{k=0}^{n} x_k^3 & \sum_{k=0}^{n} x_k^4 & \dots & \sum_{k=0}^{n} x_k^{m+2} \\
. & . & . & . & . \\
. & . & . & . & . \\
\sum_{k=0}^{n} x_k^m & \sum_{k=0}^{n} x_k^{m+1} & \sum_{k=0}^{n} x_k^{m+2} & . & \sum_{k=0}^{n} x_k^{2m}
\end{bmatrix}
\begin{Bmatrix} c_0 \\ c_1 \\ c_2 \\ . \\ . \\ c_m \end{Bmatrix}
=
\begin{Bmatrix}
\sum_{k=0}^{n} f(x_k) \\
\sum_{k=0}^{n} x_k f(x_k) \\
\sum_{k=0}^{n} x_k^2 f(x_k) \\
. \\
. \\
\sum_{k=0}^{n} x_k^m f(x_k)
\end{Bmatrix}
$$

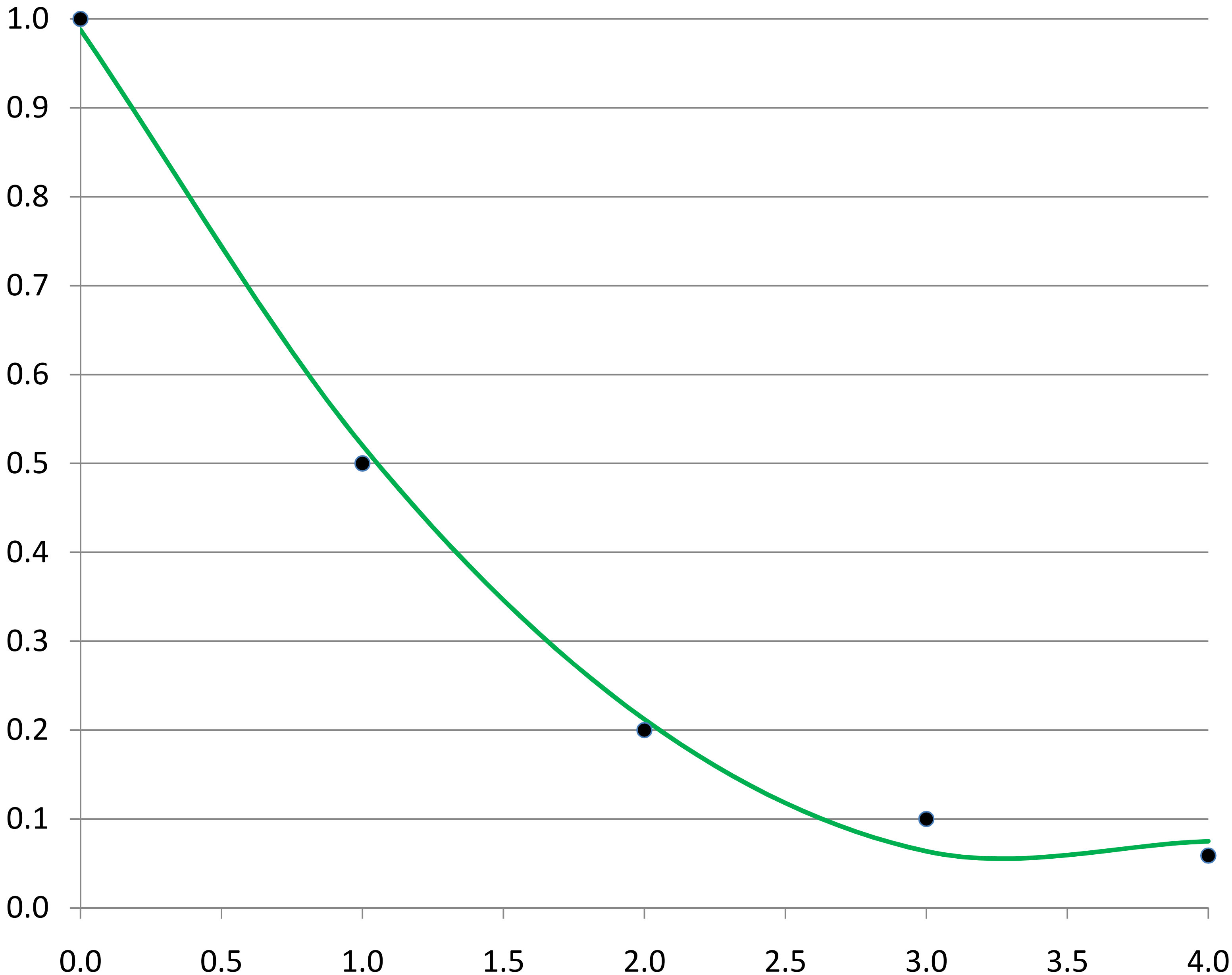
Least Squares: Example

- From the following data (n=4), estimate *f*(2.6), using regression with a quadratic polynomial (m=2):

x	0	1	2	3	4
f(x)	1	0.5	0.2	0.1	0.05882

$$\begin{bmatrix} \sum_{k=0}^{4} 1 & \sum_{k=0}^{4} x_k & \sum_{k=0}^{4} x_k^2 \\ \sum_{k=0}^{4} x_k & \sum_{k=0}^{4} x_k^2 & \sum_{k=0}^{4} x_k^3 \\ \sum_{k=0}^{4} x_k^2 & \sum_{k=0}^{4} x_k^3 & \sum_{k=0}^{4} x_k^4 \end{bmatrix} \begin{Bmatrix} c_0 \\ c_1 \\ c_2 \end{Bmatrix} = \begin{Bmatrix} \sum_{k=0}^{4} f(x_k) \\ \sum_{k=0}^{4} x_k f(x_k) \\ \sum_{k=0}^{4} x_k^2 f(x_k) \end{Bmatrix} \Rightarrow \begin{bmatrix} 5 & 10 & 30 \\ 10 & 30 & 100 \\ 30 & 100 & 354 \end{bmatrix} \begin{Bmatrix} c_0 \\ c_1 \\ c_2 \end{Bmatrix} = \begin{Bmatrix} 1.85882 \\ 1.43528 \\ 3.14112 \end{Bmatrix}$$

- Solution: 0.9879, -0.5476, 0.07983
- *f*(2.6)= 0.1039



1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0

Least Squares: Orthogonal polynomials

- Equidistant points $x_k; \quad k = 0,1,2,\ldots,n$
- Minimize: $\displaystyle\sum_{k=0}^{n}\left(f(x_k) - \sum_{j=0}^{m} c_j\phi_j(x_k)\right)^2$ => [*A*]{*c*}={*b*}

$$a_{ij} = \sum_{k=0}^{n}\phi_i(x_k)\phi_j(x_k) \text{ and } b_i = \sum_{k=0}^{n}\phi_i(x_k)f(x_k)$$

- Choose orthonormal basis functions: Known as Gram's polynomials, or discrete Tchebycheff polynomials -- denote by $G_i(x)$.
- Normalize the data range from −1 to 1.
- Implies that $x_i = -1 + \dfrac{2i}{n}$

Least Squares: Orthogonal polynomials

- $G_i(x)$ is a polynomial of degree i.

$$\sum_{k=0}^{n} G_0(x_k) G_0(x_k) = 1 \Rightarrow G_0(x) = \frac{1}{\sqrt{n+1}}$$

- Assume $G_1(x) = d_0 + d_1 x$

$$\sum_{k=0}^{n} \frac{1}{\sqrt{n+1}} \left(d_0 + d_1 x\right) = 0 \Rightarrow d_0 = 0 \quad \text{since} \sum x = 0$$

$$\sum_{k=0}^{n} \left(d_0 + d_1 x\right)^2 = 1 \Rightarrow d_1 = \frac{1}{\sqrt{\sum_{k=0}^{n} x^2}} = \frac{1}{\sqrt{\sum_{k=0}^{n} \left(-1 + \frac{2k}{n}\right)^2}}$$

Gram polynomials

- Therefore:

$$d_1 = \frac{1}{\sqrt{\sum_{k=0}^{n}\left(1+\frac{4k^2}{n^2}-\frac{4k}{n}\right)}} = \sqrt{\frac{3n}{(n+1)(n+2)}}$$

- Recursive relation:

$$G_{i+1}(x) = \alpha_i x G_i(x) - \frac{\alpha_i}{\alpha_{i-1}} G_{i-1}(x) \quad \text{for } i = 1,2,\ldots,n-1$$

$$G_0(x) = \frac{1}{\sqrt{n+1}};\ G_1(x) = x\sqrt{\frac{3n}{(n+1)(n+2)}};\ \alpha_i = \frac{n}{i+1}\sqrt{\frac{(2i+1)(2i+3)}{(n-i)(n+i+2)}}$$

Gram polynomials: Example

- From the following data (n=4), estimate *f*(2.6), using regression with a quadratic polynomial (m=2):

t	**0**	**1**	**2**	**3**	**4**
f(t)	1	0.5	0.2	0.1	0.05882

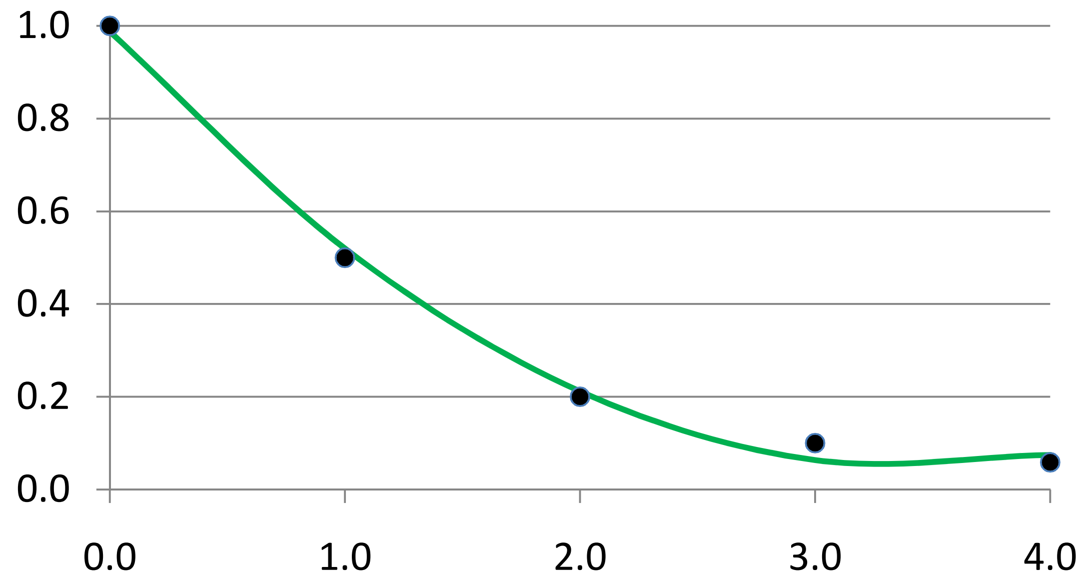
- Normalize: $x=t/2-1$
- For n=4, we get $G_0(x)=\frac{1}{\sqrt{5}};G_1(x)=x\sqrt{\frac{2}{5}};G_2(x)=\sqrt{\frac{2}{7}}\left(2x^2-1\right)$
- Normal Equations:

$$\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\begin{Bmatrix}c_0\\ c_1\\ c_2\end{Bmatrix}=\begin{Bmatrix}\sum_{k=0}^{4}G_0(x_k)f(x_k)\\ \sum_{k=0}^{4}G_1(x_k)f(x_k)\\ \sum_{k=0}^{4}G_2(x_k)f(x_k)\end{Bmatrix}=\begin{Bmatrix}0.831290\\ -0.721746\\ 0.298702\end{Bmatrix}$$

Gram polynomials: Example

$$f_2(x) = \frac{0.8313}{\sqrt{5}} - 0.7217\sqrt{\frac{2}{5}}x + 0.2987\sqrt{\frac{2}{7}}\left(2x^2 - 1\right)$$

- $f(t=2.6)=\mathrm{f}(x=0.3)=$ 0.1039
- Same as before
- How to estimate the closeness?
- Coefficient of determination



Regression: Coefficient of determination

- The "inherent spread" of the data may be represented by its deviation from mean as

$$S_t = \sum_{k=0}^{n} \left(f(x_k) - \bar{f} \right)^2$$

- $\bar{f}$ is the arithmetic mean of the function values

$$\bar{f} = \frac{\sum_{k=0}^{n} f(x_k)}{n+1}$$

- S_t is the sum of squares of the *total* deviations

Coefficient of determination

- Define a sum of *residual* deviation, from the fitted m[th]-degree polynomial as

$$S_r = \sum_{k=0}^{n} \left(f(x_k) - f_m(x_k) \right)^2$$

- Obviously, S_r should be as small as possible and, in the worst case, will be equal to S_t
- The *coefficient of determination* is defined as

$$r^2 = \frac{S_t - S_r}{S_t}$$

with its value ranging from 0 to 1.

Coefficient of determination

- A value of 0 for r^2 indicates that a constant value, equal to the mean, is the best-fit
- A value of 1 for r^2 indicates that the best-fit passes through ALL data points
- r is called the correlation coefficient
- r^2<0.3 is considered a poor fit, >0.8 is considered good
- The difference, S_t-S_r, may be thought of as the variability in the data *explained* by the regression.

Multiple Regression

- For a function of 2 (or more) variables

$$\left(x_k, y_k, f\left(x_k, y_k\right)\right) \quad k = 0,1,2,\ldots,n$$

- Minimize $\displaystyle \sum_{i=0}^{n}\left(f(x_i, y_i) - \sum_{k=0}^{m_2}\sum_{j=0}^{m_1} c_{j,k} x_i^j y_i^k \right)^2$

- Same as before: a set of (m_1+1)x(m_2+1) linear equations

Multiple Regression

- For example, with linear fit:

$$f_{11}(x,y) = c_{0,0} + c_{1,0}x + c_{0,1}y + c_{1,1}xy$$

$$\begin{bmatrix} n+1 & \sum x_i & \sum y_i & \sum x_i y_i \\ \sum x_i & \sum x_i^2 & \sum x_i y_i & \sum x_i^2 y_i \\ \sum y_i & \sum x_i y_i & \sum y_i^2 & \sum x_i y_i^2 \\ \sum x_i y_i & \sum x_i^2 y_i & \sum x_i y_i^2 & \sum x_i^2 y_i^2 \end{bmatrix} \begin{Bmatrix} c_{0,0} \\ c_{1,0} \\ c_{0,1} \\ c_{1,1} \end{Bmatrix} = \begin{Bmatrix} \sum f(x_i, y_i) \\ \sum x_i f(x_i, y_i) \\ \sum y_i f(x_i, y_i) \\ \sum x_i y_i f(x_i, y_i) \end{Bmatrix}$$

Nonlinear Regression

- Not all relationships between x and f could be expressed in linear form
- E.g., $$f(x) = c_0 e^{c_1 x} \text{ or } f(x) = \frac{c_0}{1 + c_1 e^{c_2 x}}$$
- The first one could be *linearized*

$$\ln f(x) = \ln c_0 + c_1 x$$

- But not the second one – nonlinear regression

Minimize $$\sum_{k=0}^{n} \left(f(x_k) - f_m(x, c_0, c_1, \ldots, c_m) \right)^2$$

Nonlinear Regression

- The normal equations are nonlinear
- May be solved using Newton method
- Start with an initial guess for the coefficients
- Use Taylor's series to form equations $A\Delta c=b$
- A and b comprise the derivatives of f wrt c
- *Jacobian matrix* is defined as before
- Residual r is f-f_m

$$
J = \begin{bmatrix}
\left.\dfrac{\partial f_m}{\partial c_0}\right|_{\left(x_0, c^{(k)}\right)} & \left.\dfrac{\partial f_m}{\partial c_1}\right|_{\left(x_0, c^{(k)}\right)} & \cdot & \left.\dfrac{\partial f_m}{\partial c_m}\right|_{\left(x_0, c^{(k)}\right)} \\
\left.\dfrac{\partial f_m}{\partial c_0}\right|_{\left(x_1, c^{(k)}\right)} & \left.\dfrac{\partial f_m}{\partial c_1}\right|_{\left(x_1, c^{(k)}\right)} & \cdot & \left.\dfrac{\partial f_m}{\partial c_m}\right|_{\left(x_1, c^{(k)}\right)} \\
\cdot & \cdot & \cdot & \cdot \\
\left.\dfrac{\partial f_m}{\partial c_0}\right|_{\left(x_n, c^{(k)}\right)} & \left.\dfrac{\partial f_m}{\partial c_1}\right|_{\left(x_n, c^{(k)}\right)} & \cdot & \left.\dfrac{\partial f_m}{\partial c_m}\right|_{\left(x_n, c^{(k)}\right)}
\end{bmatrix}_{(n+1)\times(m+1)}
$$

$$
[J]^T [J] \{\Delta c\} = [J]^T \{r\}
$$