Least-squares Problems

1. Generalized linear systems
2. In-class exercise
3. Linear regression
4. Nonlinear regression

Least-squares Problems

Generalized Linear Systems

E. H. Moore
1920

Generalized Linear Systems

- Linear algebraic system

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
&\vdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m
\end{aligned}
$$

- Completely defined system: $m = n$
  - » Equal number of equations and unknowns
  - » Unique solution exists if det(A) is non-zero
- Overdetermined system: $m > n$
  - » More equations than unknowns
  - » No solution generally exists
- Underdetermined system: $m < n$
  - » More unknowns than equations
  - » Infinite number of solutions

Overdetermined Systems

- Problem: **Ax** = **b**
  - » Not enough unknowns to satisfy all the equations
  - » Find **Ax** "closest" to **b**
  - » Common problem in linear and polynomial regression
- Minimize least-squares error measure

$$\boldsymbol{\varepsilon} = \mathbf{A}\mathbf{x} - \mathbf{b}$$

$$\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} = \varepsilon_1^2 + \varepsilon_2^2 + \cdots + \varepsilon_n^2 = \|\boldsymbol{\varepsilon}\|_2^2$$

$$\min_{\mathbf{x}} \Phi = \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} = (\mathbf{A}\mathbf{x} - \mathbf{b})^T(\mathbf{A}\mathbf{x} - \mathbf{b})$$

- Minimization

$$\frac{d\Phi}{d\mathbf{x}} = 0 = \frac{d}{d\mathbf{x}}(\mathbf{A}\mathbf{x} - \mathbf{b})^T(\mathbf{A}\mathbf{x} - \mathbf{b})$$

Overdetermined Systems

- Matrix differentiation formula

$$\frac{d(\mathbf{Ax}-\mathbf{b})^T(\mathbf{Ax}-\mathbf{b})}{d\mathbf{x}} = 2\mathbf{A}^T(\mathbf{Ax}-\mathbf{b}) \quad \Rightarrow \quad \frac{d\Phi}{d\mathbf{x}} = 2\mathbf{A}^T(\mathbf{Ax}-\mathbf{b}) = 0$$

- Normal equations: $(\mathbf{A}^T\mathbf{A})\mathbf{x} = \mathbf{A}^T\mathbf{b}$
- Solution: $\mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$
  - » $\mathbf{A}^T\mathbf{A}$ must be nonsingular
  - » $(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ is called the left inverse matrix
  - » The solution **x** minimizes the least-squares difference from zero over all the equations

Overdetermined Systems Example

- Linear system with 3 equations and 2 unknowns

$$\begin{bmatrix} 1 & 2 \\ 2 & -1 \\ -2 & -1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 25 \\ -25 \\ -25 \end{bmatrix} \quad \mathbf{A}^T\mathbf{A} = \begin{bmatrix} 1 & 2 & -2 \\ 2 & -1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & -1 \\ -2 & -1 \end{bmatrix} = \begin{bmatrix} 9 & 2 \\ 2 & 6 \end{bmatrix}$$

$$\mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} = \begin{bmatrix} 9 & 2 \\ 2 & 6 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 2 & -2 \\ 2 & -1 & -1 \end{bmatrix} \begin{bmatrix} 25 \\ -25 \\ -25 \end{bmatrix} = \begin{bmatrix} -1 \\ 17 \end{bmatrix}$$

- The solution **x** minimizes the least-squares error

$$\varepsilon_1 = x_1 + 2x_2 - 25 = 8 \qquad \varepsilon_2 = 2x_1 - x_2 + 25 = 8$$

$$\varepsilon_3 = -2x_1 - x_2 + 25 = 10 \quad \|\boldsymbol{\varepsilon}\|_2^2 = \varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2 = 200$$

- No other **x** can make $||\boldsymbol{\varepsilon}||_2$ smaller

Underdetermined Systems

- Problem: $\mathbf{Ax} = \mathbf{b}$
  - » Too many unknowns
  - » Find the "smallest" $\mathbf{x}$ that satisfies the equations
- Minimize least-squares error measure

$$\min_{\mathbf{x}} \Phi = \frac{1}{2}\mathbf{x}^T\mathbf{x} = \frac{1}{2}\|\mathbf{x}\|_2^2$$
$$\text{subject to}: \mathbf{Ax} = \mathbf{b}$$

- Minimization problem
  - » Convert constrained problem into unconstrained problem
  - » Utilize method of Lagrange multipliers
- Solution: $\mathbf{x} = \mathbf{A}^{\mathrm{T}}(\mathbf{AA}^{\mathrm{T}})^{-1}\mathbf{b}$
  - » $\mathbf{AA}^{\mathrm{T}}$ must be nonsingular
  - » $\mathbf{A}^{\mathrm{T}}(\mathbf{AA}^{\mathrm{T}})^{-1}$ is called the right inverse
  - » Called the minimum norm solution

Underdetermined System Example

- Linear system with 2 equations and 3 unknowns

$$\begin{bmatrix} 1 & 2 & -2 \\ 2 & -1 & -1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 25 \\ -25 \end{bmatrix} \quad \mathbf{A}\mathbf{A}^T = \begin{bmatrix} 1 & 2 & -2 \\ 2 & -1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & -1 \\ -2 & -1 \end{bmatrix} = \begin{bmatrix} 9 & 2 \\ 2 & 6 \end{bmatrix}$$

$$\mathbf{x} = \mathbf{A}^T (\mathbf{A}\mathbf{A}^T)^{-1} \mathbf{b} = \begin{bmatrix} 1 & 2 \\ 2 & -1 \\ -2 & -1 \end{bmatrix} \begin{bmatrix} 9 & 2 \\ 2 & 6 \end{bmatrix}^{-1} \begin{bmatrix} 25 \\ -25 \end{bmatrix} = \begin{bmatrix} -7 \\ 13.5 \\ -2.5 \end{bmatrix}$$

- The least-squares solution **x** satisfies both equations exactly
- No **x** with smaller $\|\mathbf{x}\|_2$ can satisfy the equations

Least-squares Problems

In-class Exercise

Least-squares Problems

Linear Regression

Roger Penrose
1956

Linear Regression

- Linear algebraic equation: $y = \alpha_1 u + \alpha_0$
- Perform N experiments: $u_i \rightarrow y_i$, $\{u_i, y_i\}$
- Linear regression equation: $\hat{y}_i = \alpha_1 u_i + \alpha_0$
- 2 unknowns (α_1, α_0) and N equations ($N > 2$)
- Formulate overdetermined least-squares problem:

$$\begin{bmatrix} u_1 & 1 \\ u_2 & 1 \\ \vdots & \vdots \\ u_N & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_0 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \quad \Rightarrow \quad \mathbf{Ax} = \mathbf{b}$$

Linear Regression Example

- Effect of monomer concentration (u) on the polymer production rate (y)

Expt	**1**	**2**	**3**	**4**	**5**
u	0	2	4	6	8
y	0	30	78	143	225

- Formulate **A** matrix and **b** vector

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 2 & 1 \\ 4 & 1 \\ 6 & 1 \\ 8 & 1 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} 0 \\ 30 \\ 78 \\ 143 \\ 225 \end{bmatrix}$$

Linear Regression Example

- Form and invert $\mathbf{A}^{\mathrm{T}}\mathbf{A}$

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 0 & 2 & 4 & 6 & 8 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 2 & 1 \\ 4 & 1 \\ 6 & 1 \\ 8 & 1 \end{bmatrix} = \begin{bmatrix} 120 & 20 \\ 20 & 5 \end{bmatrix}$$

$$(\mathbf{A}^T\mathbf{A})^{-1} = \frac{1}{200} \begin{bmatrix} 5 & -20 \\ -20 & 120 \end{bmatrix} = \begin{bmatrix} 0.025 & -0.1 \\ -0.1 & 0.6 \end{bmatrix}$$

- Compute parameter estimates

$$\mathbf{x} = \begin{bmatrix} \alpha_1 \\ \alpha_0 \end{bmatrix} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} = \begin{bmatrix} 0.025 & -0.1 \\ -0.1 & 0.6 \end{bmatrix} \begin{bmatrix} 0 & 2 & 4 & 6 & 8 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 30 \\ 78 \\ 143 \\ 225 \end{bmatrix} = \begin{bmatrix} 28.15 \\ -17.40 \end{bmatrix}$$

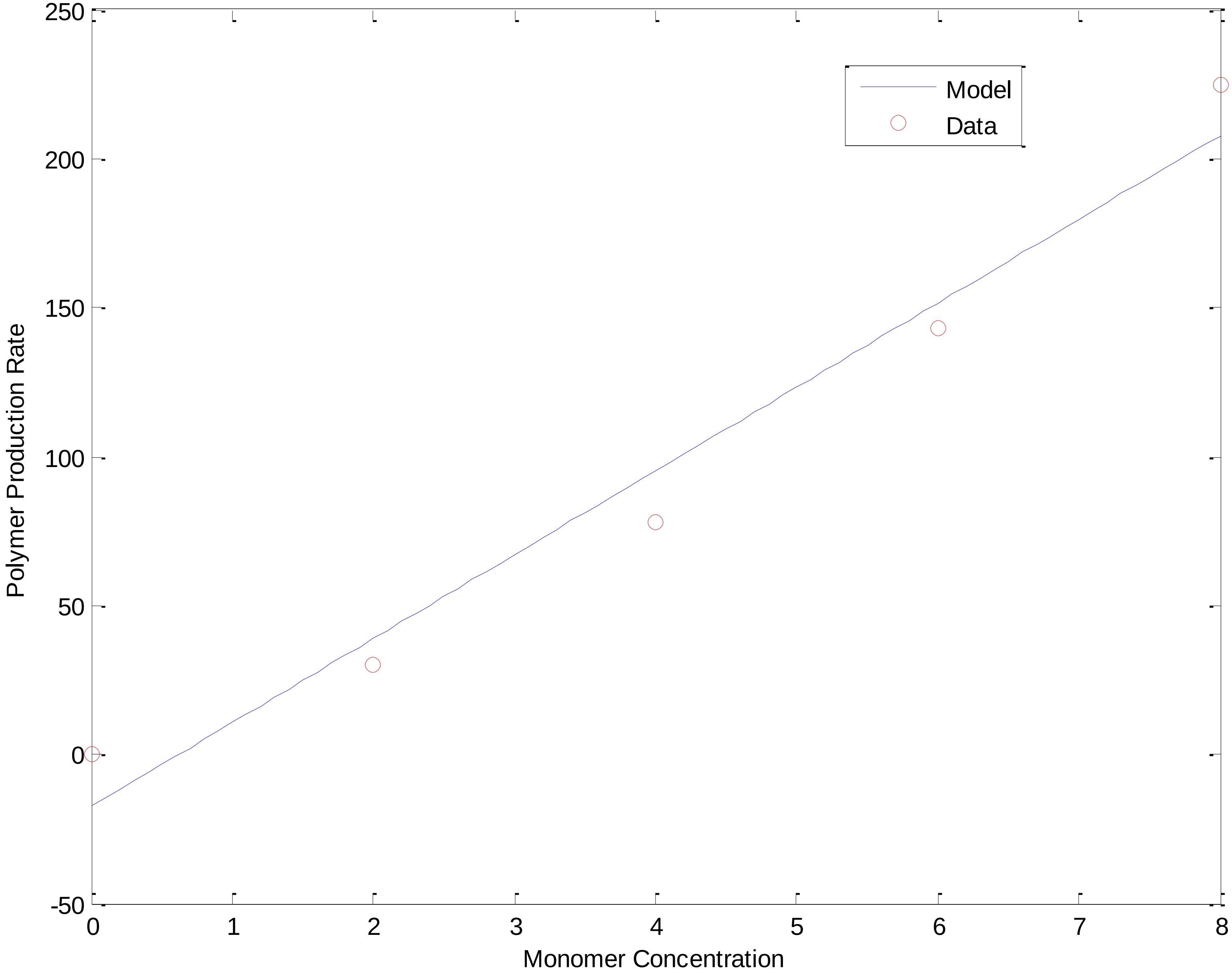
Linear Regression Example

- Compute sum of squared errors

$$\boldsymbol{\varepsilon} = \mathbf{Ax} - \mathbf{b} = \begin{bmatrix} \hat{y}_1 - y_1 \\ \hat{y}_2 - y_2 \\ \hat{y}_3 - y_3 \\ \hat{y}_4 - y_4 \\ \hat{y}_5 - y_5 \end{bmatrix} = \begin{bmatrix} \alpha_1 u_1 + \alpha_0 - y_1 \\ \alpha_1 u_2 + \alpha_0 - y_1 \\ \alpha_1 u_3 + \alpha_0 - y_3 \\ \alpha_1 u_4 + \alpha_0 - y_4 \\ \alpha_1 u_5 + \alpha_0 - y_5 \end{bmatrix} = \begin{bmatrix} 17.4 \\ -8.9 \\ -17.2 \\ 8.5 \\ 17.2 \end{bmatrix}$$

$$\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = \begin{bmatrix} 17.4 & -8.9 & -17.2 & 8.5 & 17.2 \end{bmatrix} \begin{bmatrix} 17.4 \\ -8.9 \\ -17.2 \\ 8.5 \\ 17.2 \end{bmatrix} = 1046$$

Linear Regression Example



Least-squares Problems

Polynomial Regression

Polynomial Regression

- Polynomial equation: $y = \alpha_n u^n + \alpha_{n-1} u^{n-1} + \cdots + \alpha_1 u + \alpha_0$
- Perform N experiments: $u_i \rightarrow y_i$, $\{u_i, y_i\}$
- Regression equation:

$$\hat{y}_i = \alpha_n u_i^n + \alpha_{n-1} u_i^{n-1} + \cdots + \alpha_1 u_i + \alpha_0$$

- n+1 unknowns ($\alpha_n, \ldots, \alpha_0$) and N equations ($N > n+1$)
- Formulate overdetermined least-squares problem:

$$\begin{bmatrix} u_1^n & u_1^{n-1} & \cdots & 1 \\ u_2^n & u_2^{n-1} & \cdots & 1 \\ \vdots & \vdots & \ddots & 1 \\ u_N^n & u_N^{n-1} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \alpha_n \\ \alpha_{n-1} \\ \vdots \\ \alpha_0 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \Rightarrow \quad \mathbf{A}\mathbf{x} = \mathbf{b}$$

Quadratic Regression Example

- Effect of monomer concentration (u) on polymer production rate (y)

Expt	1	2	3	4	5
u	0	2	4	6	8
y	0	30	78	143	225

- Formulate **A** matrix and **b** vector

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 \\ 4 & 2 & 1 \\ 16 & 4 & 1 \\ 36 & 6 & 1 \\ 64 & 8 & 1 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} 0 \\ 30 \\ 78 \\ 143 \\ 225 \end{bmatrix}$$

Quadratic Regression Example

- Compute parameter estimates

$$\mathbf{x} = \begin{bmatrix} \alpha_2 \\ \alpha_1 \\ \alpha_0 \end{bmatrix} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} = \begin{bmatrix} 5664 & 800 & 120 \\ 800 & 120 & 120 \\ 20 & 120 & 5 \end{bmatrix} \begin{bmatrix} 0 & 4 & 16 & 36 & 64 \\ 0 & 2 & 4 & 6 & 8 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 30 \\ 78 \\ 143 \\ 225 \end{bmatrix} = \begin{bmatrix} 2.16 \\ 10.86 \\ -0.11 \end{bmatrix}$$

- Compute sum of squared errors

$$\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} = \begin{bmatrix} -0.11 & 0.26 & -0.09 & -0.14 & 0.09 \end{bmatrix} \begin{bmatrix} -0.11 \\ 0.26 \\ -0.09 \\ -0.14 \\ 0.09 \end{bmatrix} = 0.11$$

Quadratic Regression Example