# *Computational Methods*
CMSC/AMSC/MAPL 460

Least squares method: linear regression

Ramani Duraiswami,

Dept. of Computer Science

# Fitting data to a model

- Practical science involves lots of fitting of data to models
- Difference between fitting and interpolation?
  - Interpolation, the fit function passes through the point
  - Fitting, the fit function satisfies some norm based criterion
- Tasks arise commonly in science
  - Fit straight lines and curves to data
  - More generally fit data to a model
- Model contains parameters
  - Job of fitting is to estimate the parameters that “best” make the model fit the data
  - “best” → define best
- Simplest example of model fitting problem
  - Linear regression

# Models

- Have a certain model structure
  - E.g., "linear" "quadratic" "trigonometric" "Gaussian"
- Models have specifiable parameters
- e.g.
- Model Structure Data Parameters

| | Structure | Data | Parameters |
|---|---|---|---|
| Straight line: | $a\,x + by + c = 0$ | $(x_i, y_i)$ | $(a,b,c)$ |
| Polynomial: | $y=c_0+c_1\,x+\ \ldots+c_n\,x^n$ | $(x_i, y_i)$ | $(c_0, c_1, \ldots, c_n)$ |

- General model
- $y(t) \simeq \beta_1\,\phi_1\,(t) + \ldots + \beta_n\,\phi_n\,(t) +$
- Data $(y_i, t_i)$ and
- $y = \mathbf{\Phi}\beta$
- Solve $\beta = \mathbf{y} \backslash\ \mathbf{\Phi}$

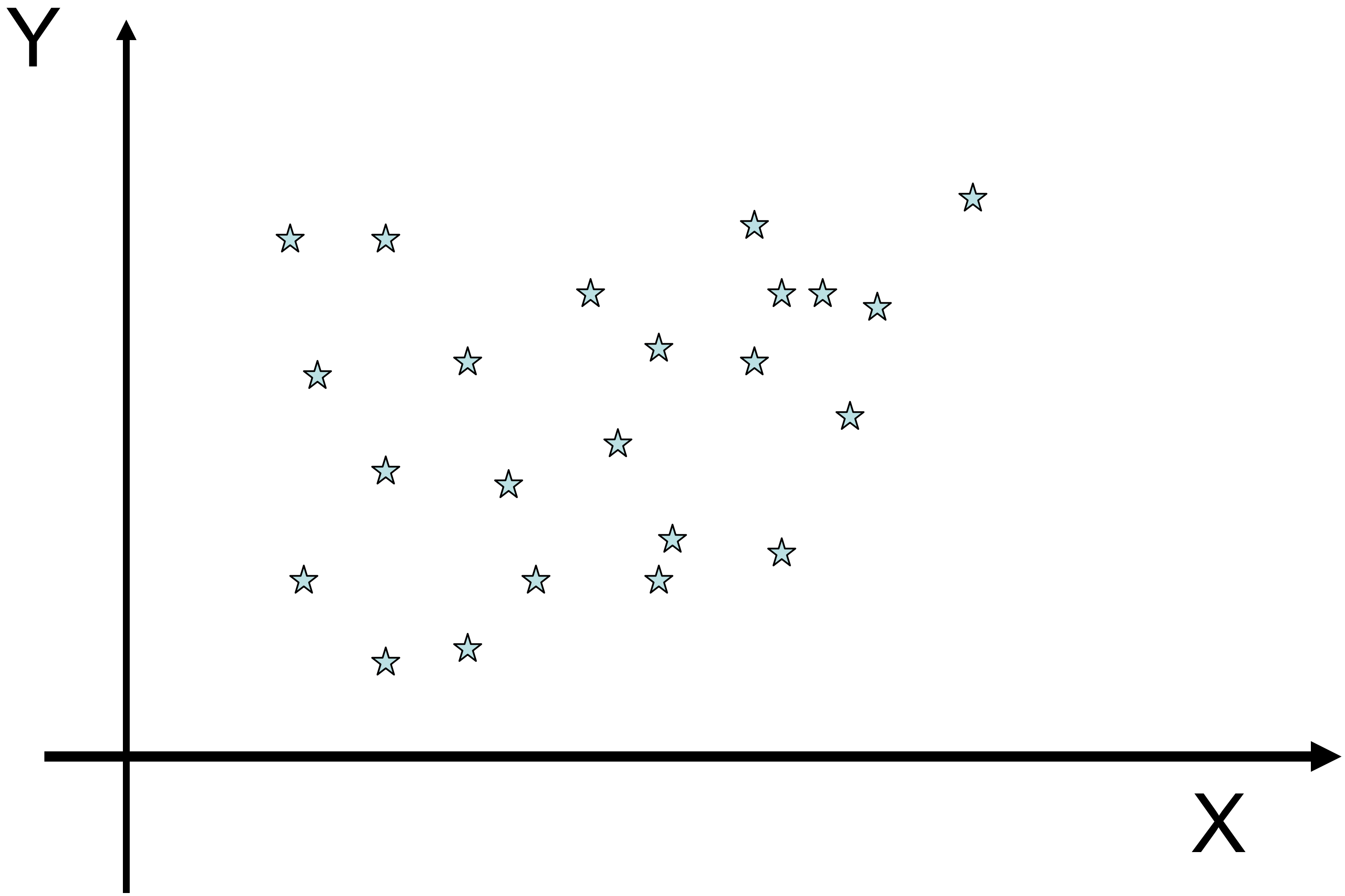
# Relationships among Variables

- In much science we seek relations between variables

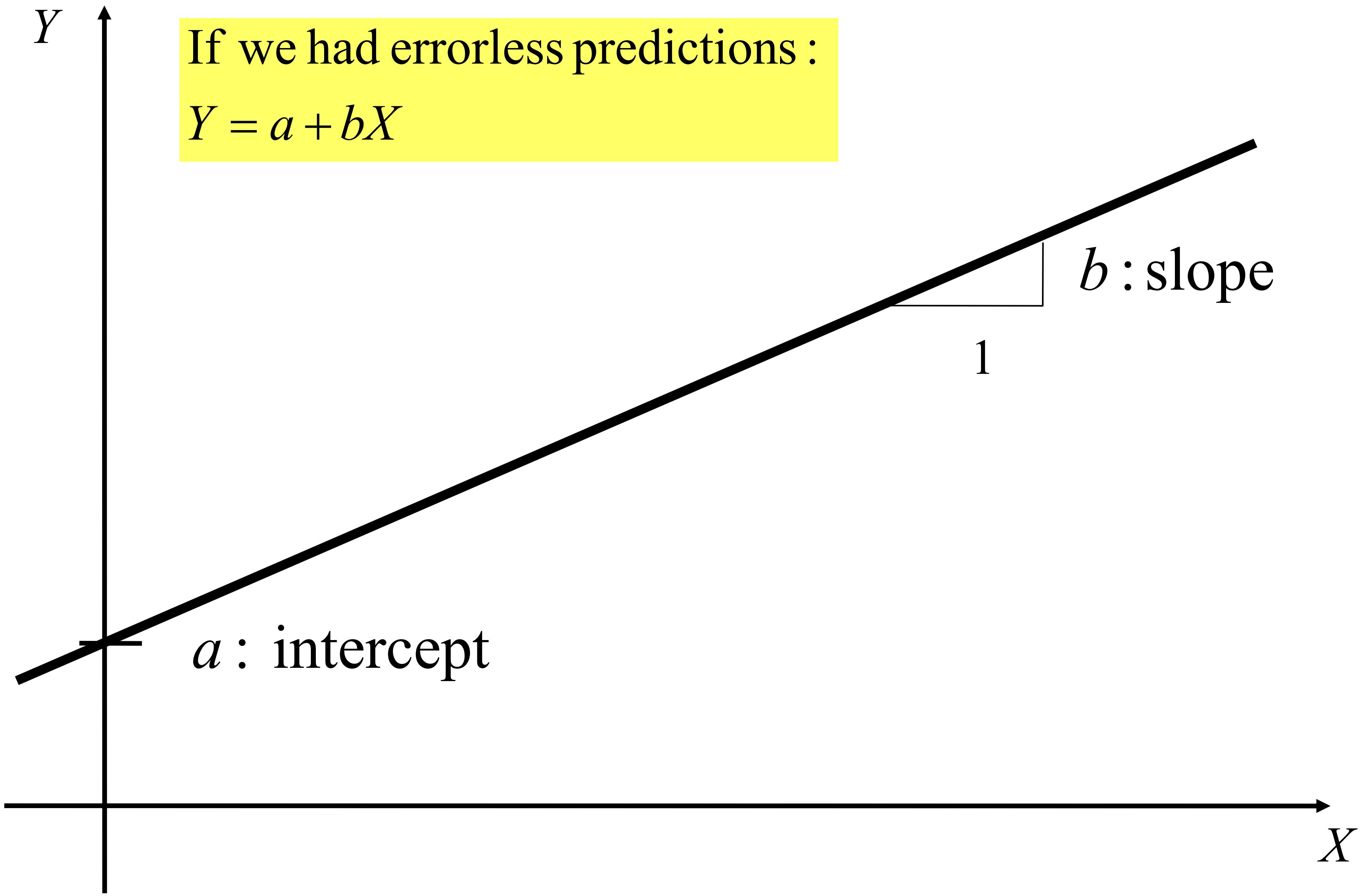
**One variable is used to “explain” another variable**

| X Variable | | Y Variable |
|---|---|---|
| Independent Variable<br>Explaining Variable<br>Exogenous Variable<br>Predictor Variable | → | Dependent Variable<br>Response Variable<br>Endogenous Variable<br>Criterion Variable |

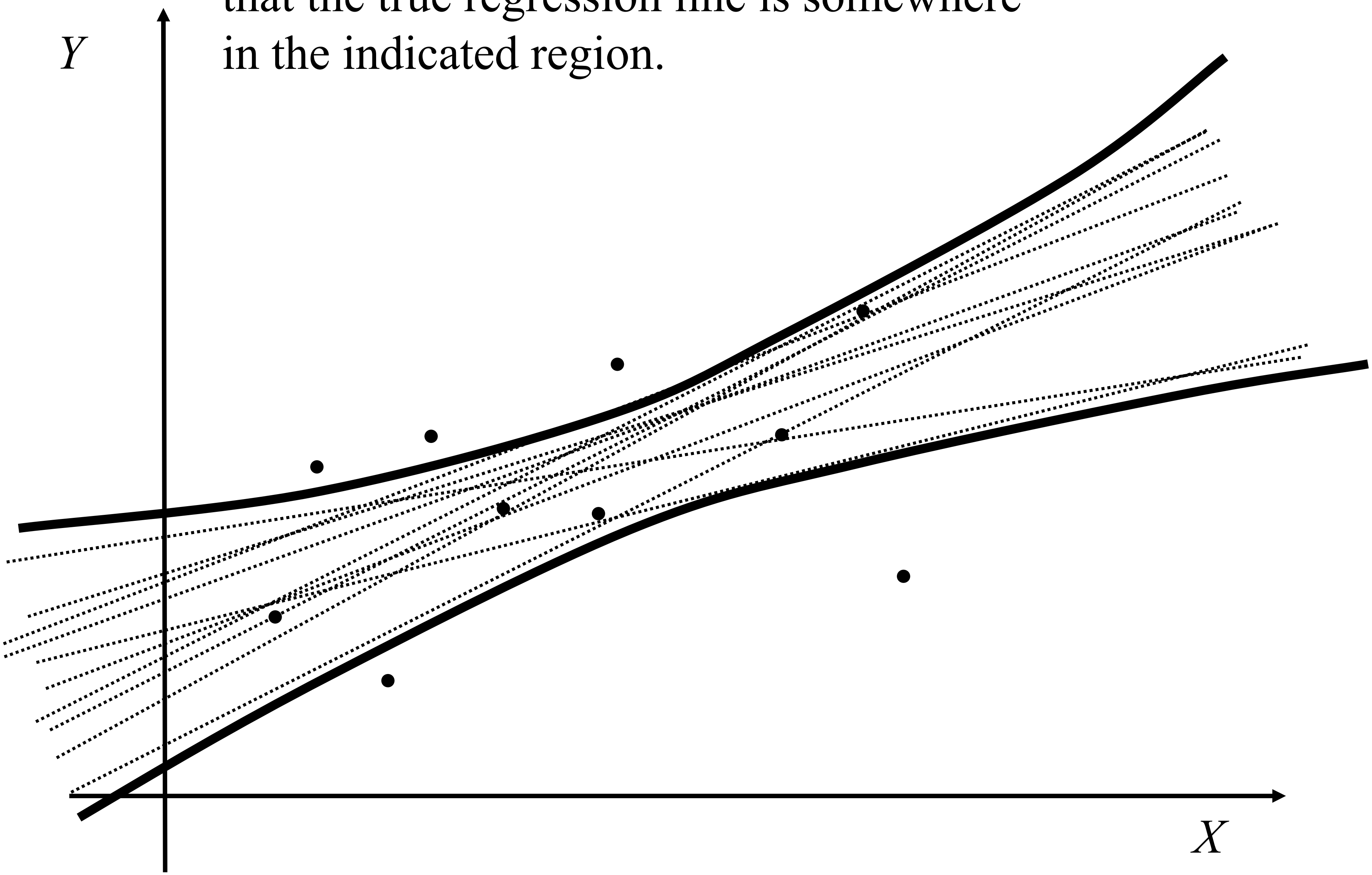
# Scatter Plots



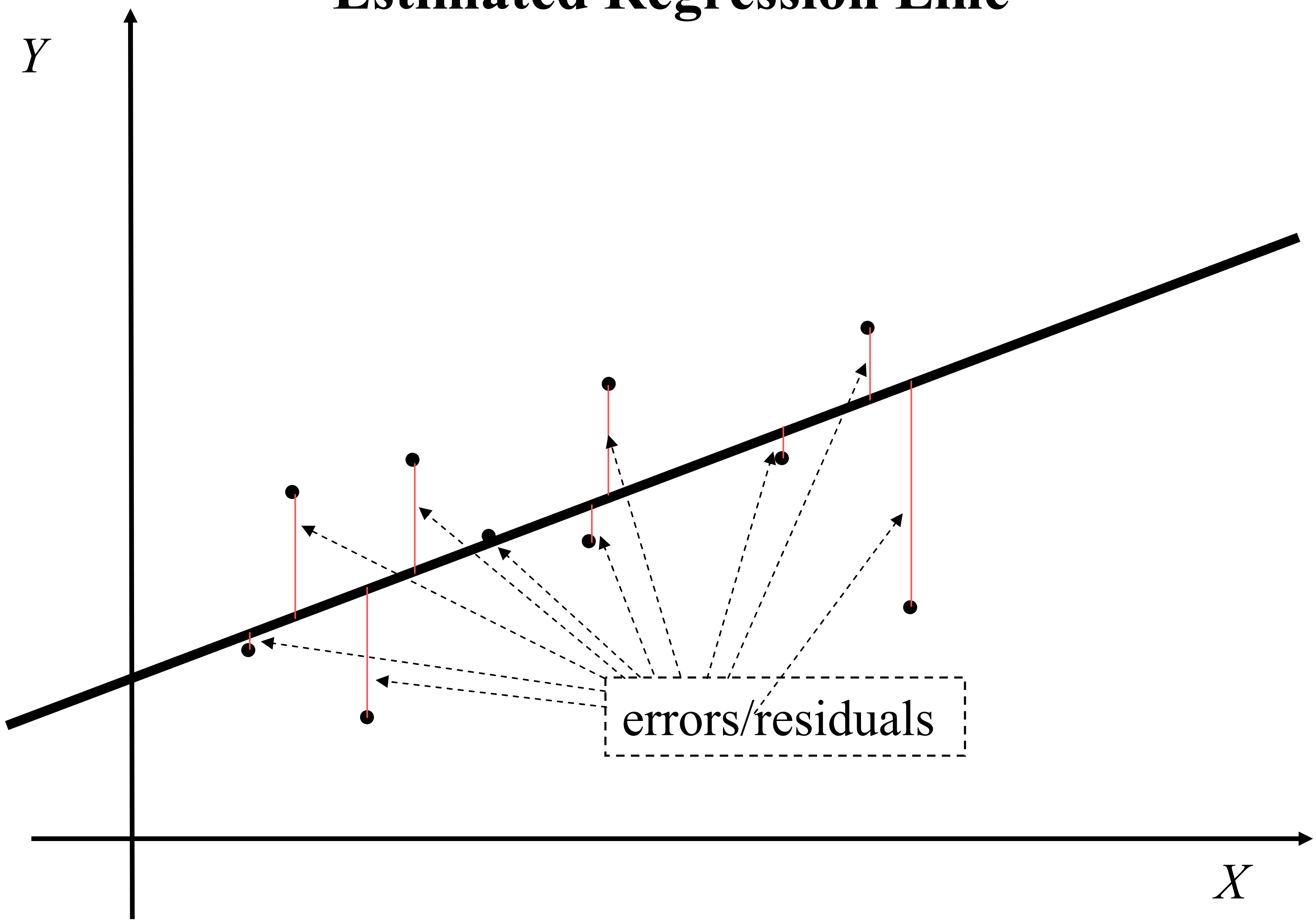
# Simple Least-Squares Regression



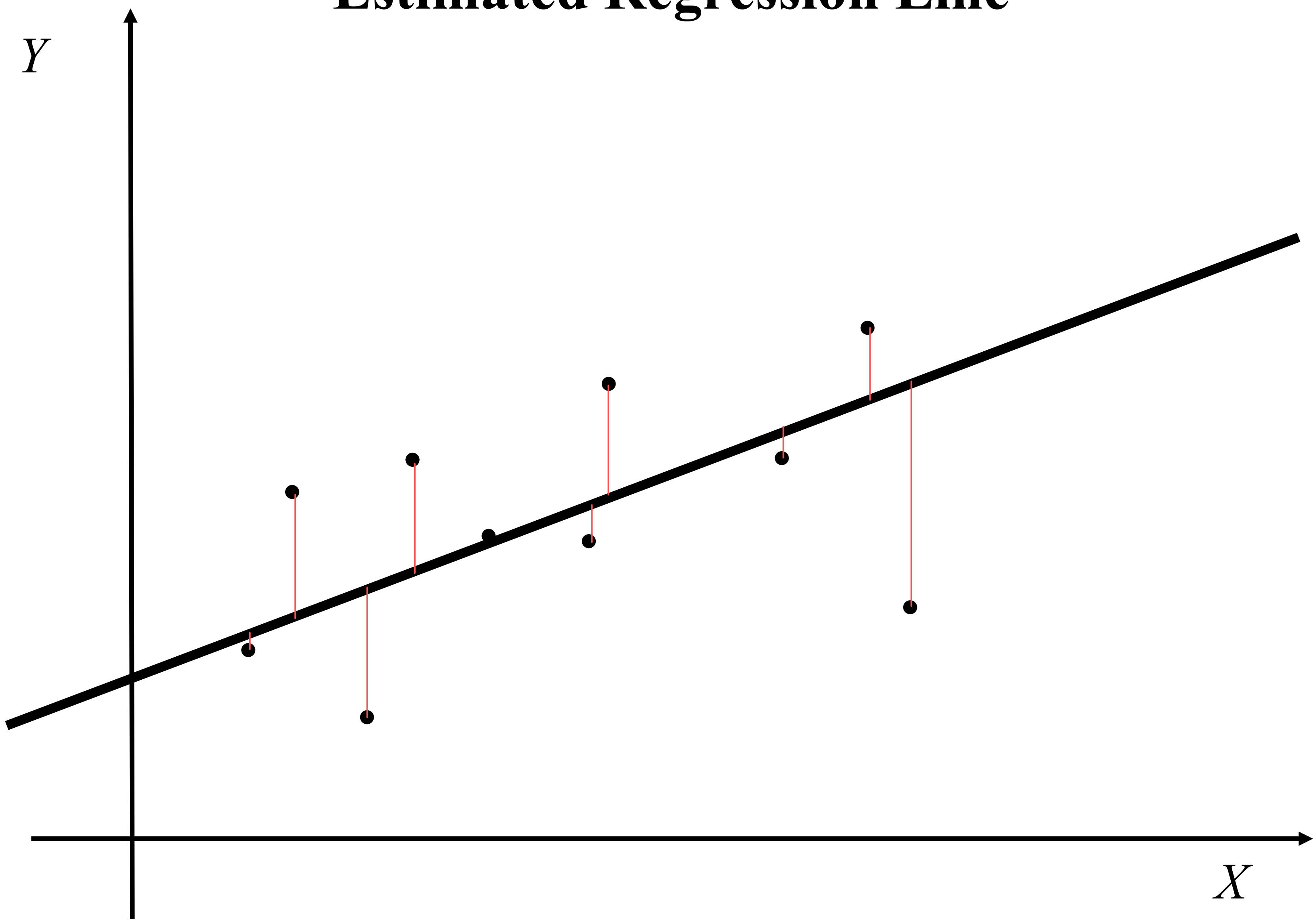
We will end up being reasonably confident that the true regression line is somewhere in the indicated region.



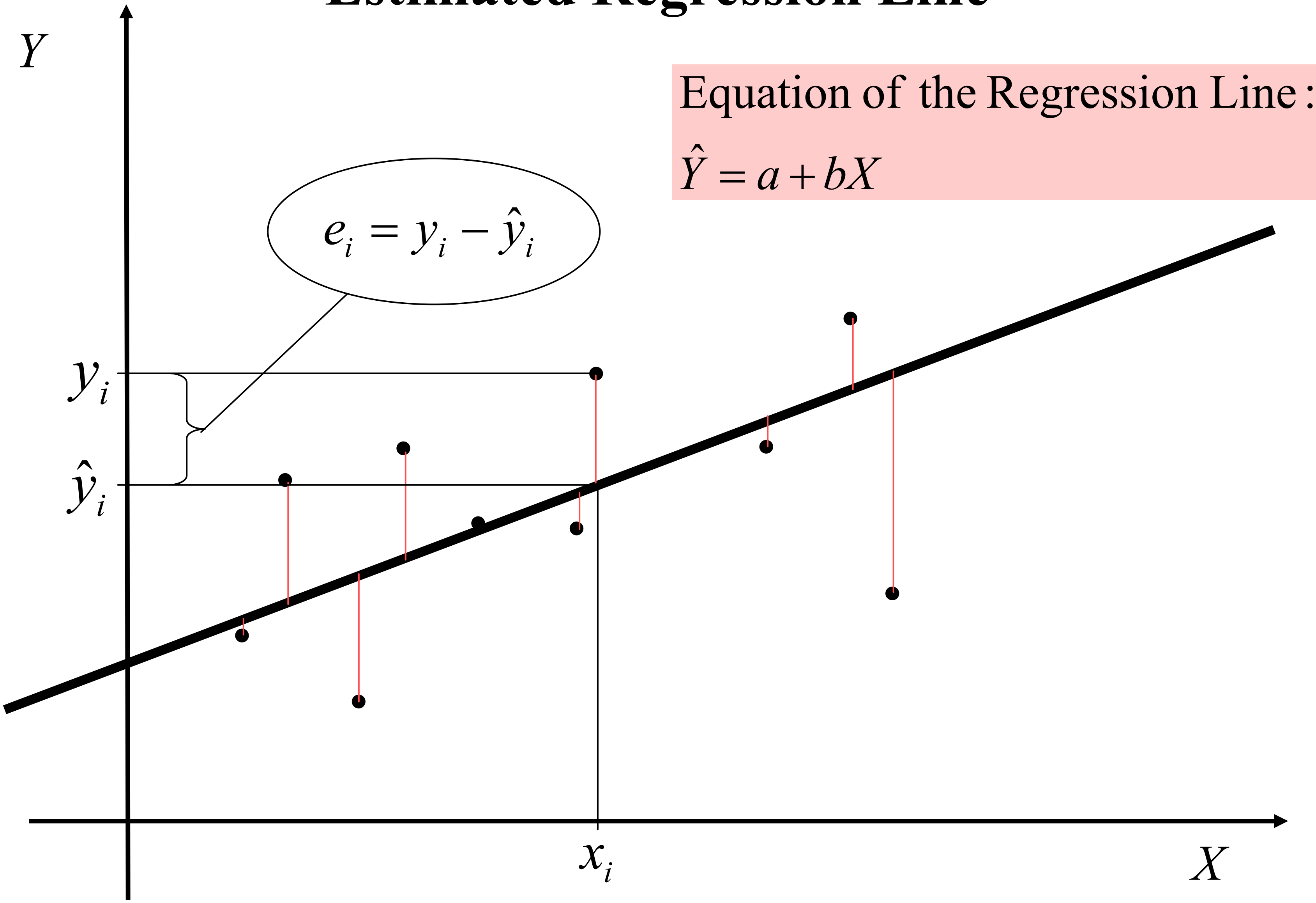
# Estimated Regression Line



# Estimated Regression Line

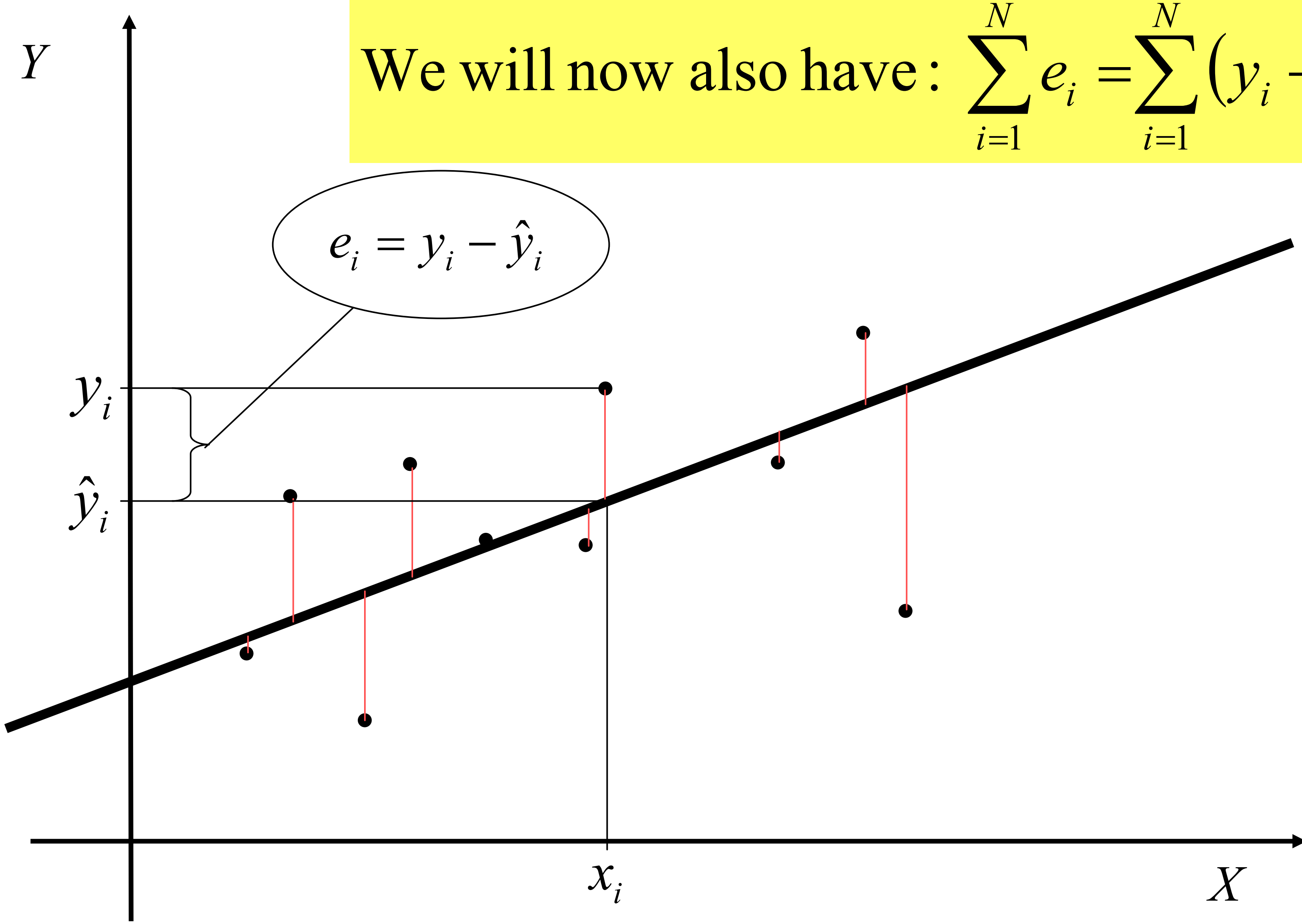


# Estimated Regression Line



$$\text{Remember: } \sum_{i=1}^{N} (y_i - \bar{y}) = 0$$

$$\text{We will now also have: } \sum_{i=1}^{N} e_i = \sum_{i=1}^{N} (y_i - \hat{y}_i) = 0$$



# How do we find a and b?

## In Least-Squares Regression:

Find $a, b$ to minimize the sum of squared errors/residuals

$$\sum_{i=1}^{N} (e_i)^2 = \sum_{i=1}^{N} (y_i - [bx_i + a])^2$$

# In Least-Squares Regression:

$$b = \frac{\sum_{i=1}^{N} (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{N} (X_i - \overline{X})^2}, \qquad a = \overline{Y} - b\overline{X}$$

Computational Formula

$$b = \frac{N\sum_{i=1}^{N} X_i Y_i - \left(\sum_{i=1}^{N} X_i\right)\left(\sum_{i=1}^{N} Y_i\right)}{N\sum_{i=1}^{N} X_i^2 - \left(\sum_{i=1}^{N} X_i\right)^2}$$

# Models

- Have a certain model structure
  - E.g., " straight line" "parabolic" "trigonometric" "Gaussian"
- Models have specifiable parameters
- e.g.
- Model Structure Data Parameters

| Model | Structure | Data | Parameters |
|---|---|---|---|
| Straight line: | $a\,x + by + c = 0$ | $(x_i, y_i)$ | $(a,b,c)$ |
| Polynomial: | $y = c_0 + c_1\,x + \ldots + c_n\,x^n$ | $(x_i, y_i)$ | $(c_0, c_1, \ldots, c_n)$ |
| Trig.: | $y = c_0 + c_1 \sin x + \ldots + c_n \sin nx$ | $(x_i, y_i)$ | $(c_0, c_1, \ldots, c_n)$ |
| Gaussian | $y = c_0 \exp(-(x-\mu)^2/\sigma^2)$ | $(x_i, y_i)$ | $(c_0, \sigma, \mu)$ |
| "Kernel/RBF" | $y = \sum_i a_i\, k(x, x_i)$ | $(x_i, y_i)$ | $(a_i)$ and $(x_i)$ |

- These models are all separable

# Linear models

- All models considered are "linear" or "separable"
- (does not mean straight lines)
- Means that we can separate the "structure" and "parameters" as a matrix-vector product
- "structure" forms a matrix and "parameters" a vector
- Goal of model fitting: find the parameters
- Question: Is the Gaussian model a linear one?
  - Can be made separable by taking logs

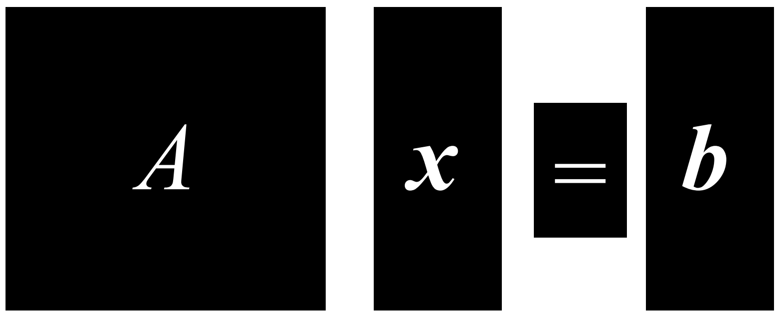
$$y(t) \approx K e^{\lambda t}$$

$$\log y \approx \beta_1 t + \beta_2, \text{ with } \beta_1 = \lambda, \ \beta_2 = \log K$$

  - However if there are sums of gaussians or exponentials the model is not separable

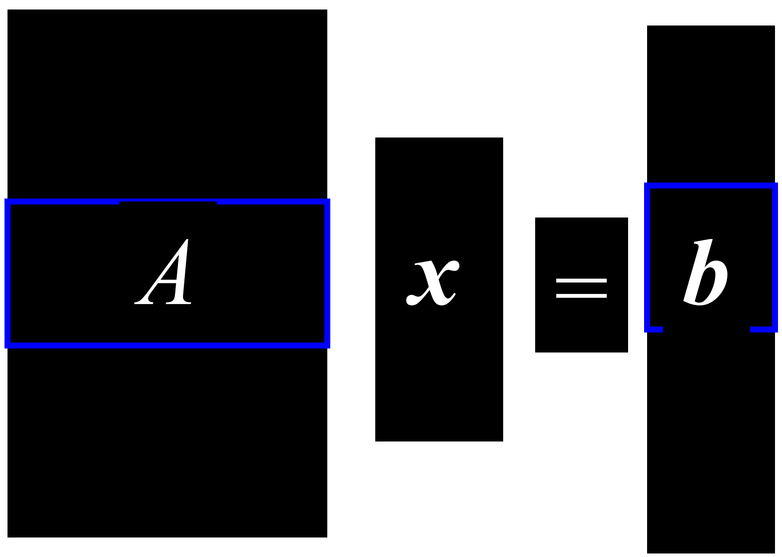
$$y(t) \approx \beta_1 e^{-\left(\frac{t-\mu_1}{\sigma_1}\right)^2} + \ldots \beta_n e^{-\left(\frac{t-\mu_n}{\sigma_n}\right)^2}$$

# Linear Systems




Square system:

- unique solution
- Gaussian elimination




Rectangular system ??

- underconstrained:
  infinity of solutions
- overconstrained:
  no solution

→ Minimize $|Ax-b|^2$

# Least Squares for more complex models

- Number of equations and unknowns may not match
- Data may have noise
- Look for solution by minimizing some cost function
- Simplest and most intuitive cost function: $\|\mathbf{Ax} - \mathbf{b}\|_2$
- Define for each data point $x_i$ a residual $r_i$
- *Minimize* $\sum_i r_i\, r_i$ with respect to $x_l$
- $r_i\, r_i = \sum_j (A_{ij}x_j - b_i) . \sum_k (A_{ik}x_k - b_i)$

$$\frac{\partial}{\partial x_l}(A_{ij}x_j - b_i)\bullet(A_{ik}x_k - b_i) = 0$$

$$(A_{ij}\delta_{jl})\bullet(A_{ik}x_k - b_i) + (A_{ij}x_j - b_i)\bullet(A_{ik}\delta_{kl}) = 0$$

$$A_{il}\bullet(A_{ik}x_k - b_i) + (A_{ij}x_j - b_i)\bullet A_{il} = 2\left(A_{il}A_{ik}x_k - A_{il}b_i\right) = 0$$

$$A_{il}A_{ik}x_k = A_{il}b_i$$

# Other Norms

- Here we fit using the "least-squares" or $L_2$ norm
- Could minimize the residual in other norms
- For example we may have more confidence in some data, and want to be sure that their residual is lower
  – Attach a weight to each residual

$$\|r\|_w^2 = \sum_1^m w_i r_i^2$$

- Or we may like the 1-norm or infinity norm better

$$\|r\|_1 = \sum_1^m |r_i| \qquad \|r\|_\infty = \max_i |r_i|$$

# Normal equations

- The system $A^t A x = A^t b$
  is called the Normal equations
- Can solve least squares problems using these
- For A size $m \times n$ and x of size n and b of size m what are the dimensions of the normal equations?
  - $n \times n$
- Solve via LU decomposition
- Is this a good idea?
  - Somewhat expensive as we have to form $A^t A$ which involves matrix multiplication and then solution
  - More importantly it is poorly conditioned
  - $\text{cond}(A^t A) = (\text{cond}(A))^2$

- Example

$$X = \begin{pmatrix} 1 & 1 \\ \delta & 0 \\ 0 & \delta \end{pmatrix}$$

$$X^T X = \begin{pmatrix} 1+\delta^2 & 1 \\ 1 & 1+\delta^2 \end{pmatrix}$$