LECTURE 15: Linear models with normal noise

$$X_i = \sum_{j=1}^{m} a_{ij}\Theta_j + W_i$$

W_i, Θ_j: independent, normal

- Very common and convenient model
- Bayes' rule: normal posteriors
- MAP and LMS estimates coincide
  - simple formulas (linear in the observations)
- Many nice properties
- Trajectory estimation example

Recognizing normal PDFs

$X \sim N(\mu, \sigma^2)$

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

$c \cdot e^{-8(x-3)^2}$

$f_X(x) = c \cdot e^{-(\alpha x^2 + \beta x + \gamma)}$ $\quad \alpha > 0 \quad$ Normal with mean $-\beta/2\alpha$ and variance $1/2\alpha$

Estimating a normal random variable in the presence of additive normal noise

$X = \Theta + W$ $\Theta, W : N(0,1)$, independent

$$f_{\Theta|X}(\theta \mid x) = \frac{f_\Theta(\theta)\, f_{X|\Theta}(x \mid \theta)}{f_X(x)}$$

$$f_X(x) = \int f_\Theta(\theta) f_{X|\Theta}(x \mid \theta)\, d\theta$$

$f_{X|\Theta}(x \mid \theta)$:

$f_{\Theta|X}(\theta \mid x) =$

$\hat{\theta}_{\text{MAP}} =$ $\hat{\theta}_{\text{LMS}} = \mathbf{E}[\Theta \mid X = x] =$

$\widehat{\Theta}_{\text{MAP}} = \mathbf{E}[\Theta \mid X] =$

Estimating a normal random variable in the presence of additive normal noise

$$f_{\Theta|X}(\theta \mid x) = \frac{f_\Theta(\theta)\, f_{X|\Theta}(x \mid \theta)}{f_X(x)}$$

$$f_X(x) = \int f_\Theta(\theta) f_{X|\Theta}(x \mid \theta)\, d\theta$$

$X = \Theta + W \qquad \Theta, W : N(0,1), \quad$ independent

$$\widehat{\Theta}_{\text{MAP}} = \widehat{\Theta}_{\text{LMS}} = \mathbf{E}[\Theta \mid X] = \frac{X}{2}$$

- Even with general means and variances:
  - posterior is normal
  - LMS and MAP estimators coincide
  - these estimators are "linear," of the form $\widehat{\Theta} = aX + b$

The case of multiple observations

$$f_{\Theta|X}(\theta \mid x) = \frac{f_\Theta(\theta)\, f_{X|\Theta}(x \mid \theta)}{f_X(x)}$$

$$f_X(x) = \int f_\Theta(\theta) f_{X|\Theta}(x \mid \theta)\, d\theta$$

$$X_1 = \Theta + W_1 \qquad \Theta \sim N(x_0, \sigma_0^2) \qquad W_i \sim N(0, \sigma_i^2)$$

$$\vdots$$

$$X_n = \Theta + W_n \qquad \Theta, W_1, \ldots, W_n \text{ independent}$$

$$f_{X_i|\Theta}(x_i \,|\, \theta) =$$

$$f_{X|\Theta}(x \,|\, \theta) =$$

$$f_{\Theta|X}(\theta \,|\, x) =$$

The case of multiple observations

$$f_{\Theta|X}(\theta \,|\, x) = c \cdot \exp\Big\{ -\,\mathsf{quad}(\theta) \Big\} \qquad \mathsf{quad}(\theta) = \frac{(\theta - x_0)^2}{2\sigma_0^2} + \frac{(\theta - x_1)^2}{2\sigma_1^2} + \cdots + \frac{(\theta - x_n)^2}{2\sigma_n^2}$$

$$\widehat{\theta}_{\mathsf{MAP}} = \widehat{\theta}_{\mathsf{LMS}} = \mathbf{E}[\Theta \,|\, X = x] = \frac{\displaystyle\sum_{i=0}^{n} \frac{x_i}{\sigma_i^2}}{\displaystyle\sum_{i=0}^{n} \frac{1}{\sigma_i^2}}$$

The case of multiple observations

- Key conclusions:
  - posterior is normal
  - LMS and MAP estimates coincide
  - these estimates are "linear," of the form $\widehat{\theta} = a_0 + a_1 x_1 + \cdots + a_n x_n$
- Interpretations:
  - estimate $\widehat{\theta}$: weighted average of x_0 (prior mean) and x_i (observations)
  - weights determined by variances

$$\widehat{\theta}_{\text{MAP}} = \widehat{\theta}_{\text{LMS}} = \mathbf{E}[\Theta \mid X = x] = \frac{\displaystyle\sum_{i=0}^{n} \frac{x_i}{\sigma_i^2}}{\displaystyle\sum_{i=0}^{n} \frac{1}{\sigma_i^2}}$$

The mean squared error

$$f_{\Theta|X}(\theta \,|\, x) = c \cdot \exp\Big\{ - \text{quad}(\theta) \Big\}$$

$$\hat{\theta} = \frac{\displaystyle\sum_{i=0}^{n} \frac{x_i}{\sigma_i^2}}{\displaystyle\sum_{i=0}^{n} \frac{1}{\sigma_i^2}}$$

$$\text{quad}(\theta) = \frac{(\theta - x_0)^2}{2\sigma_0^2} + \frac{(\theta - x_1)^2}{2\sigma_1^2} + \cdots + \frac{(\theta - x_n)^2}{2\sigma_n^2}$$

- Performance measures:

$$\mathbf{E}\Big[(\Theta - \widehat{\Theta})^2 \mid X = x\Big] = \mathbf{E}\Big[(\Theta - \hat{\theta})^2 \mid X = x\Big] = \text{var}(\Theta \mid X = x) = 1 \Big/ \sum_{i=0}^{n} \frac{1}{\sigma_i^2}$$

$$\mathbf{E}\Big[(\Theta - \widehat{\Theta})^2\Big] =$$

$f_X(x) = c \cdot e^{-(\alpha x^2 + \beta x + \gamma)}$ $\quad \alpha > 0 \quad$ Normal with mean $-\beta/2\alpha$ and variance $1/2\alpha$

The mean squared error

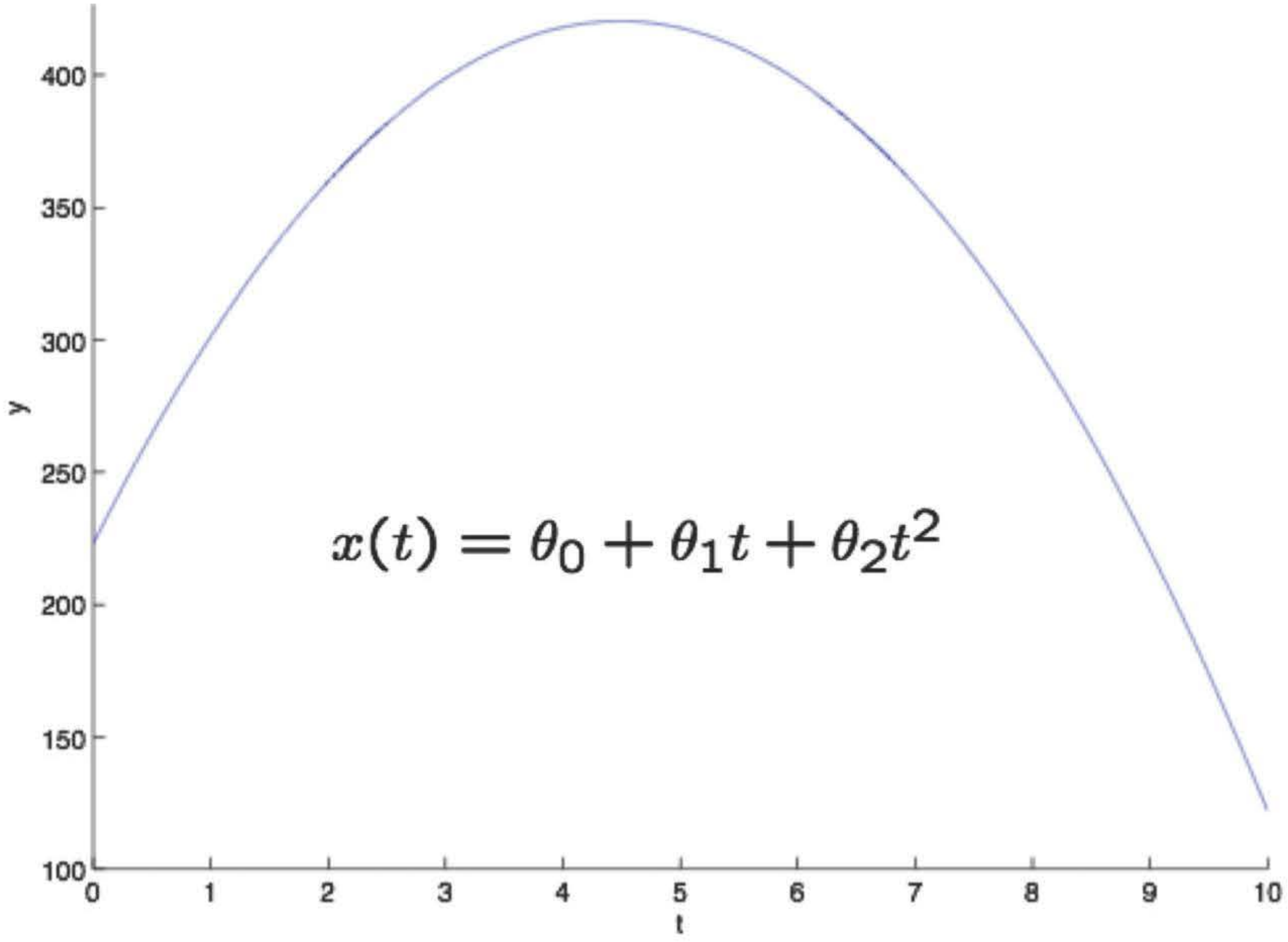
$$\mathbf{E}\big[(\Theta - \widehat{\Theta})^2 \mid X = x\big] = \mathbf{E}\big[(\Theta - \widehat{\Theta})^2\big] = 1\Big/\sum_{i=0}^{n} \frac{1}{\sigma_i^2}$$

$$\hat{\theta} = \frac{\sum_{i=0}^{n} \frac{x_i}{\sigma_i^2}}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}}$$

- Example: $\sigma_0^2 = \sigma_1^2 = \cdots = \sigma_n^2 = \sigma^2$
- conditional mean squared error same for all x
- Example: $X = \Theta + W$ $\quad \Theta \sim N(0,1), \quad W \sim N(0,1)$

  independent Θ, W $\quad \widehat{\Theta} = X/2 \quad \mathbf{E}\big[(\Theta - \widehat{\Theta})^2 \mid X = x\big] =$

The case of multiple parameters: trajectory estimation



- Random variables $\Theta_0, \Theta_1, \Theta_2$

  independent; priors f_{Θ_j}

- Measurements at times $t_1, \ldots, t_n$

  $X_i = \Theta_0 + \Theta_1 t_i + \Theta_2 t_i^2 + W_i$

  noise model: f_{W_i}

  independent W_i; independent from Θ_j

A model with normality assumptions

$$X_i = \Theta_0 + \Theta_1 t_i + \Theta_2 t_i^2 + W_i \qquad i = 1, \ldots, n$$

$$f_{\Theta|X}(\theta \mid x) = \frac{f_\Theta(\theta)\, f_{X|\Theta}(x \mid \theta)}{f_X(x)}$$

$$f_X(x) = \int f_\Theta(\theta) f_{X|\Theta}(x \mid \theta)\, d\theta$$

- assume $\Theta_j \sim N(0, \sigma_j^2)$, $W_i \sim N(0, \sigma^2)$; independent
- Given $\Theta = \theta = (\theta_0, \theta_1, \theta_2)$, X_i is:

$$f_{X_i|\Theta}(x_i \mid \theta) = c \cdot \exp\left\{ -(x_i - \theta_0 - \theta_1 t_i - \theta_2 t_i^2)^2 / 2\sigma^2 \right\}$$

- posterior: $f_{\Theta|X}(\theta \mid x) =$

$$c(x) \exp\left\{ -\frac{1}{2}\left(\frac{\theta_0^2}{\sigma_0^2} + \frac{\theta_1^2}{\sigma_1^2} + \frac{\theta_2^2}{\sigma_2^2}\right) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \theta_0 - \theta_1 t_i - \theta_2 t_i^2)^2 \right\}$$

A model with normality assumptions

$$f_{\Theta|X}(\theta \,|\, x) = \; c(x) \exp\left\{ -\frac{1}{2}\left(\frac{\theta_0^2}{\sigma_0^2} + \frac{\theta_1^2}{\sigma_1^2} + \frac{\theta_2^2}{\sigma_2^2}\right) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \theta_0 - \theta_1 t_i - \theta_2 t_i^2)^2 \right\}$$

- MAP estimate: maximize over $(\theta_0, \theta_1, \theta_2)$;
  (minimize quadratic function)

Linear normal models

- Θ_j and X_i and are linear functions of independent normal random variables

- $f_{\Theta|X}(\theta \mid x) = c(x) \exp\Big\{ - \text{quadratic}(\theta_1, \ldots, \theta_m)\Big\}$

- MAP estimate: maximize over $(\theta_1, \ldots \theta_m)$;
  (minimize quadratic function)

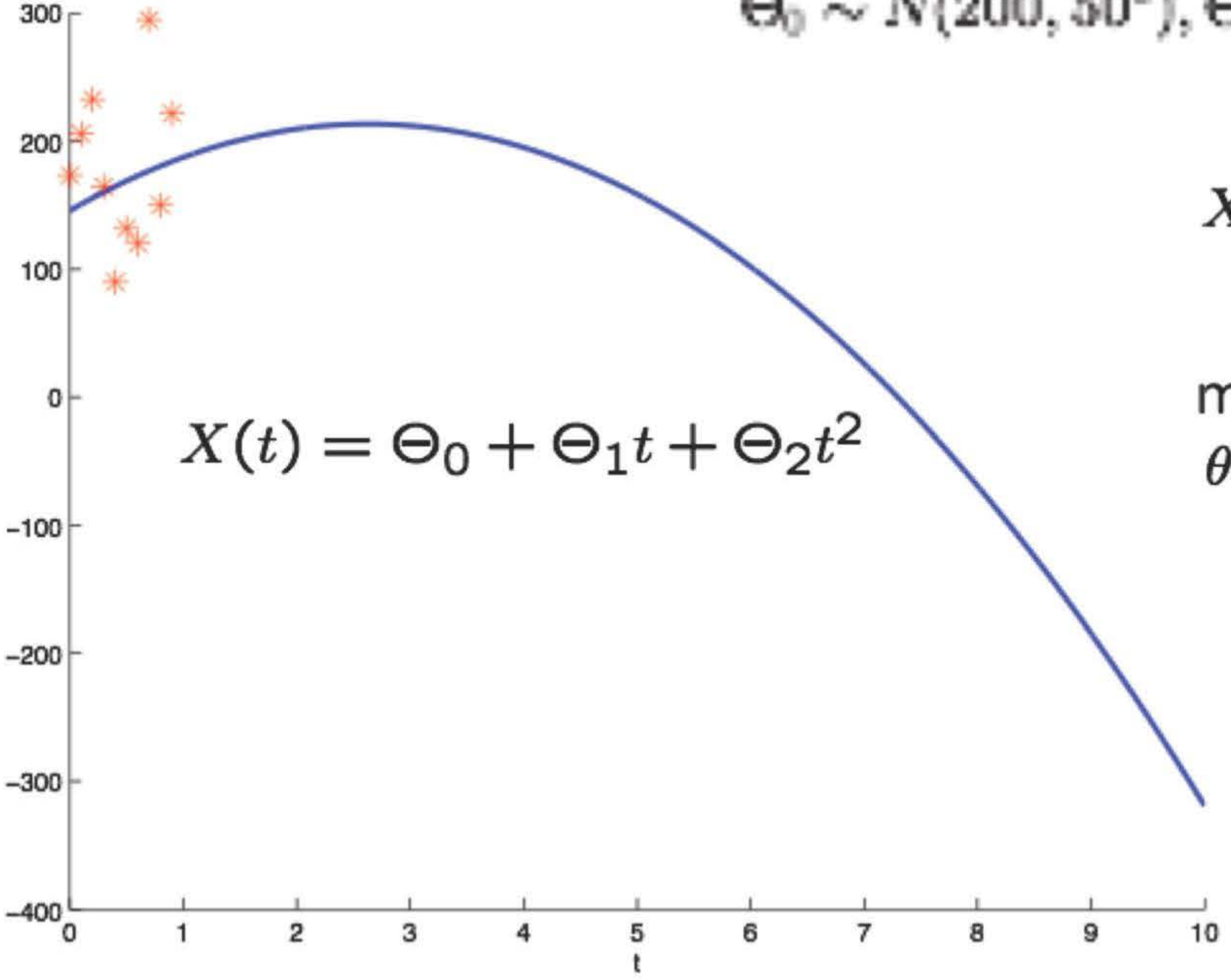
  $\widehat{\Theta}_{\text{MAP},j}$: linear function of $X = (X_1, \ldots, X_n)$

- Facts:
  - $\widehat{\Theta}_{\text{MAP},j} = \mathbf{E}[\Theta_j \mid X]$
  - marginal posterior PDF of Θ_j: $f_{\Theta_j|X}(\theta_j \mid x)$, is normal
  - MAP estimate based on the joint posterior PDF:
    same as MAP estimate based on the marginal posterior PDF
  - $\mathbf{E}\Big[(\widehat{\Theta}_{i,\text{MAP}} - \Theta_i)^2 \mid X = x\Big]$: same for all x

An illustration

Estimating the trajectory of a free-falling object

$$\Theta_0 \sim N(200, 50^2), \Theta_1 \sim N(50, 50^2), \Theta_2 = -9.81, W_i \sim N(0, 50^2)$$



$$X(t) = \Theta_0 + \Theta_1 t + \Theta_2 t^2$$

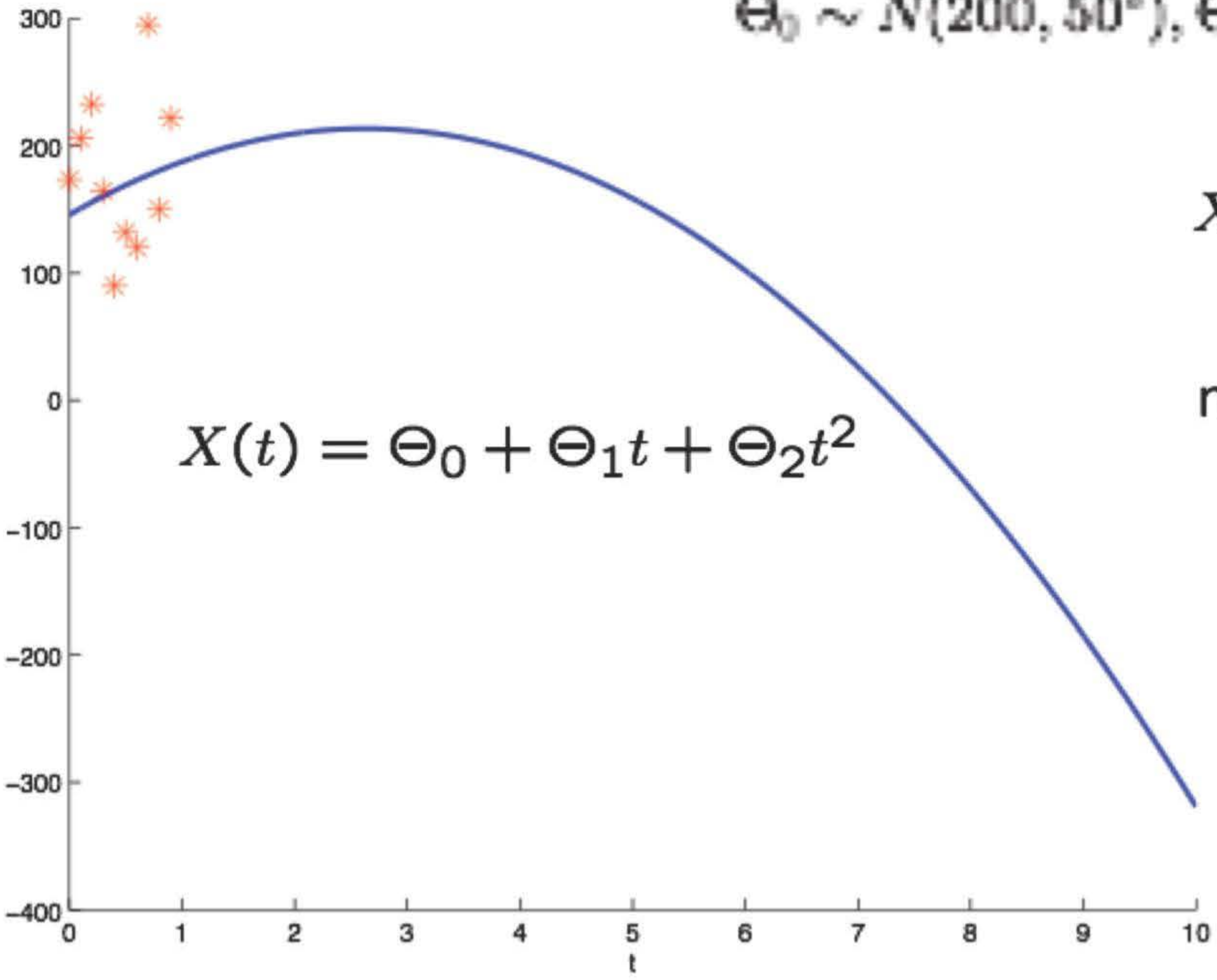
$$X_i = \Theta_0 + \Theta_1 t_i + \Theta_2 t_i^2 + W_i$$

minimize $\theta_0, \theta_1, \theta_2$

$$\frac{1}{2}\left(\frac{\theta_0^2}{\sigma_0^2} + \frac{\theta_1^2}{\sigma_1^2} + \frac{\theta_2^2}{\sigma_2^2}\right) + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \theta_0 - \theta_1 t_i - \theta_2 t_i^2)^2$$

An illustration

Estimating the trajectory of a free-falling object



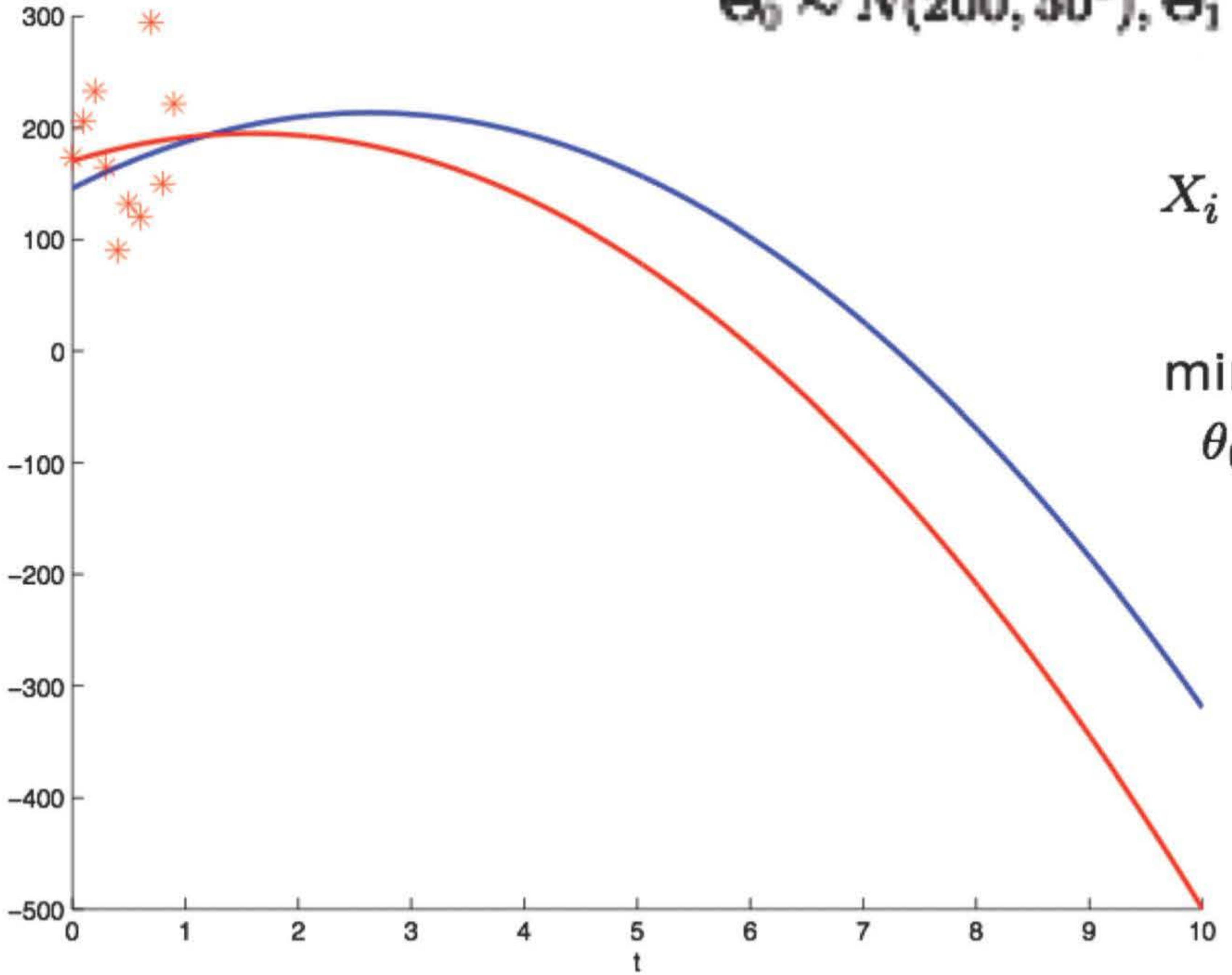
$$\Theta_0 \sim N(200, 50^2), \Theta_1 \sim N(50, 50^2), \Theta_2 = -9.81, W_i \sim N(0, 50^2)$$

$$X_i = \Theta_0 + \Theta_1 t_i + \Theta_2 t_i^2 + W_i$$

$$\underset{\theta_0, \theta_1}{\text{minimize}} \quad (\theta_0 - 200)^2 + (\theta_1 - 50)^2 + \sum_{i=1}^{n} (x_i - \theta_0 - \theta_1 t_i + 9.81 t_i^2)^2$$

An illustration

Estimating the trajectory of a free-falling object



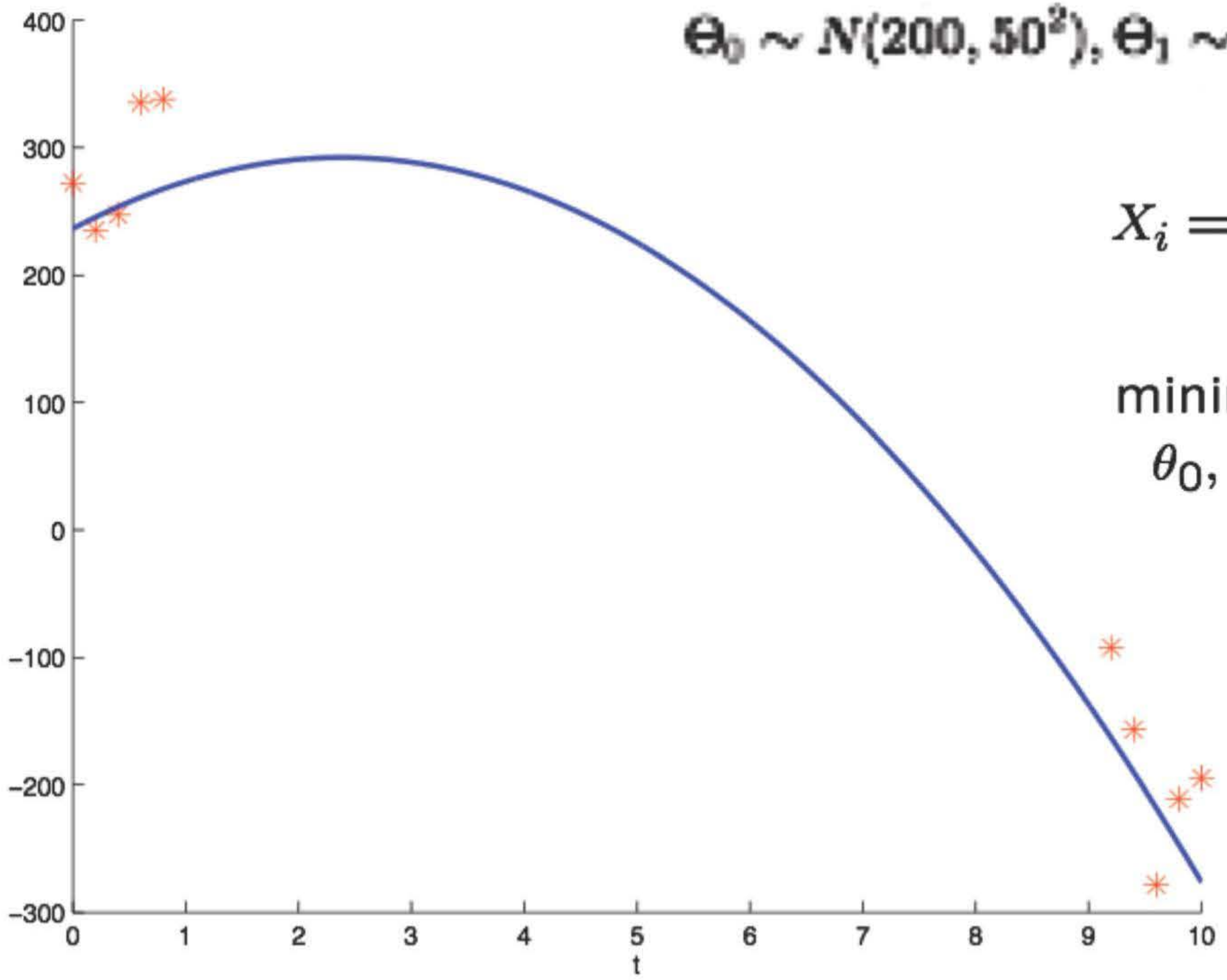
$$\Theta_0 \sim N(200, 50^2), \Theta_1 \sim N(50, 50^2), \Theta_2 = -9.81, W_i \sim N(0, 50^2)$$

$$X_i = \Theta_0 + \Theta_1 t_i + \Theta_2 t_i^2 + W_i$$

$$\underset{\theta_0,\, \theta_1}{\text{minimize}} \quad (\theta_0 - 200)^2 + (\theta_1 - 50)^2 + \sum_{i=1}^{n} (x_i - \theta_0 - \theta_1 t_i + 9.81 t_i^2)^2$$

An illustration

Estimating the trajectory of a free-falling object



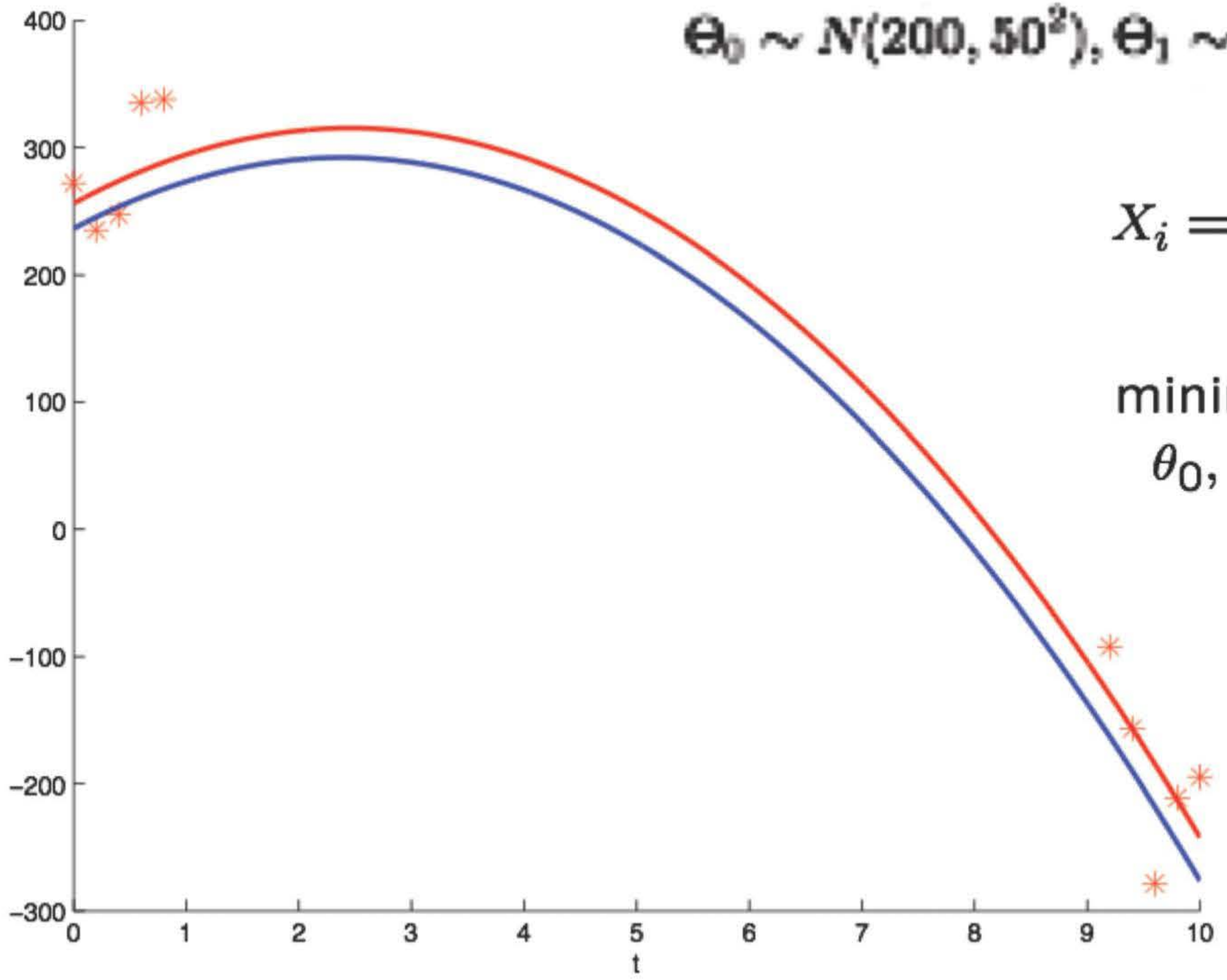
$$\Theta_0 \sim N(200, 50^2), \Theta_1 \sim N(50, 50^2), \Theta_2 = -9.81, W_i \sim N(0, 50^2)$$

$$X_i = \Theta_0 + \Theta_1 t_i + \Theta_2 t_i^2 + W_i$$

$$\underset{\theta_0,\, \theta_1}{\text{minimize}} \quad (\theta_0 - 200)^2 + (\theta_1 - 50)^2 + \sum_{i=1}^{n} (x_i - \theta_0 - \theta_1 t_i + 9.81 t_i^2)^2$$

An illustration

Estimating the trajectory of a free-falling object



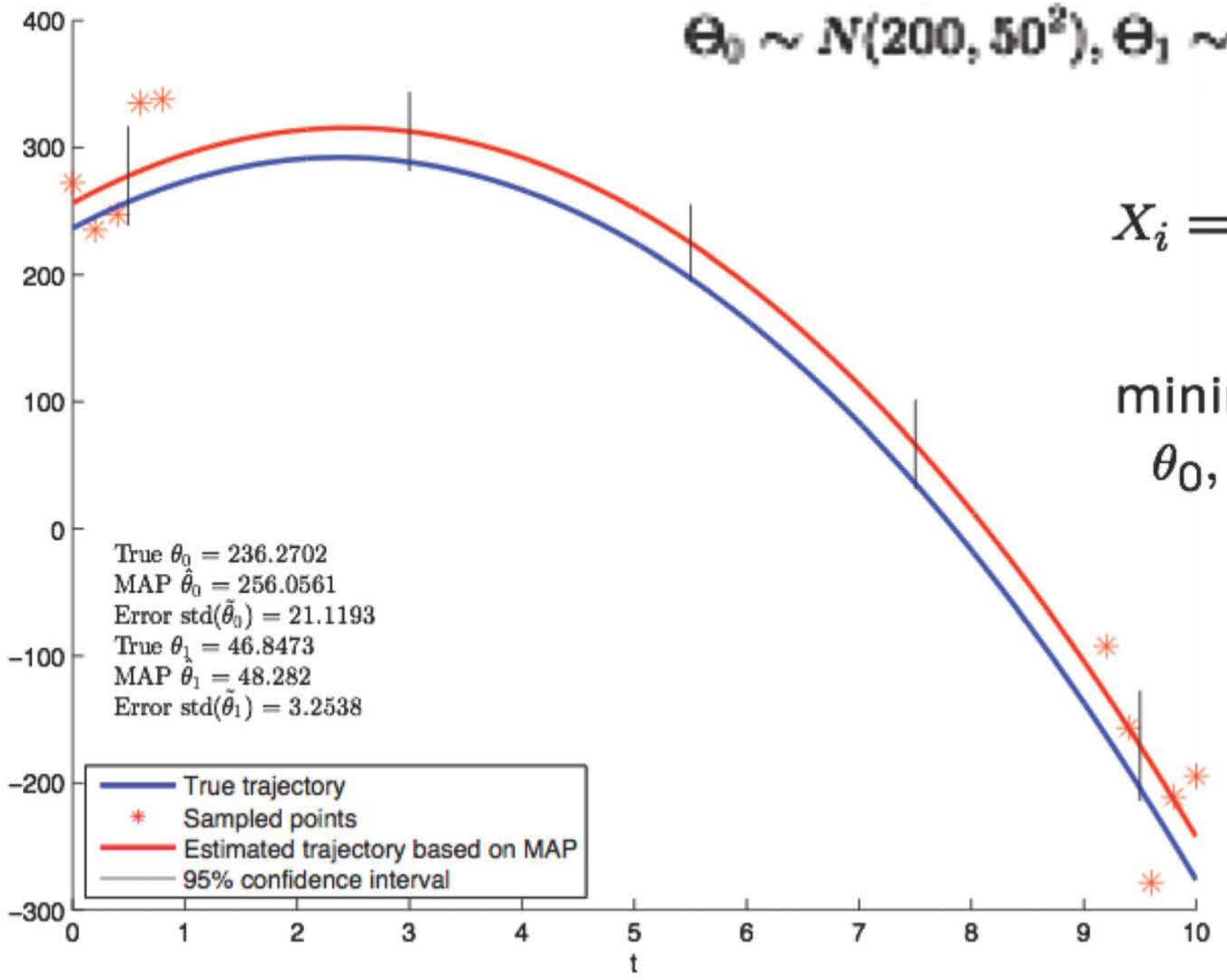
$$\Theta_0 \sim N(200, 50^2), \Theta_1 \sim N(50, 50^2), \Theta_2 = -9.81, W_i \sim N(0, 50^2)$$

$$X_i = \Theta_0 + \Theta_1 t_i + \Theta_2 t_i^2 + W_i$$

$$\underset{\theta_0,\, \theta_1}{\text{minimize}} \quad (\theta_0 - 200)^2 + (\theta_1 - 50)^2 + \sum_{i=1}^{n} (x_i - \theta_0 - \theta_1 t_i + 9.81 t_i^2)^2$$

An illustration

Estimating the trajectory of a free-falling object



$$\Theta_0 \sim N(200, 50^2), \Theta_1 \sim N(50, 50^2), \Theta_2 = -9.81, W_i \sim N(0, 50^2)$$

$$X_i = \Theta_0 + \Theta_1 t_i + \Theta_2 t_i^2 + W_i$$

$$\underset{\theta_0,\, \theta_1}{\text{minimize}} \quad (\theta_0 - 200)^2 + (\theta_1 - 50)^2 + \sum_{i=1}^{n} (x_i - \theta_0 - \theta_1 t_i + 9.81 t_i^2)^2$$

MIT OpenCourseWare
https://ocw.mit.edu

Resource: Introduction to Probability
John Tsitsiklis and Patrick Jaillet

The following may not correspond to a particular course on MIT OpenCourseWare, but has been provided by the author as an individual learning resource.

For information about citing these materials or our Terms of Use, visit: https://ocw.mit.edu/terms.