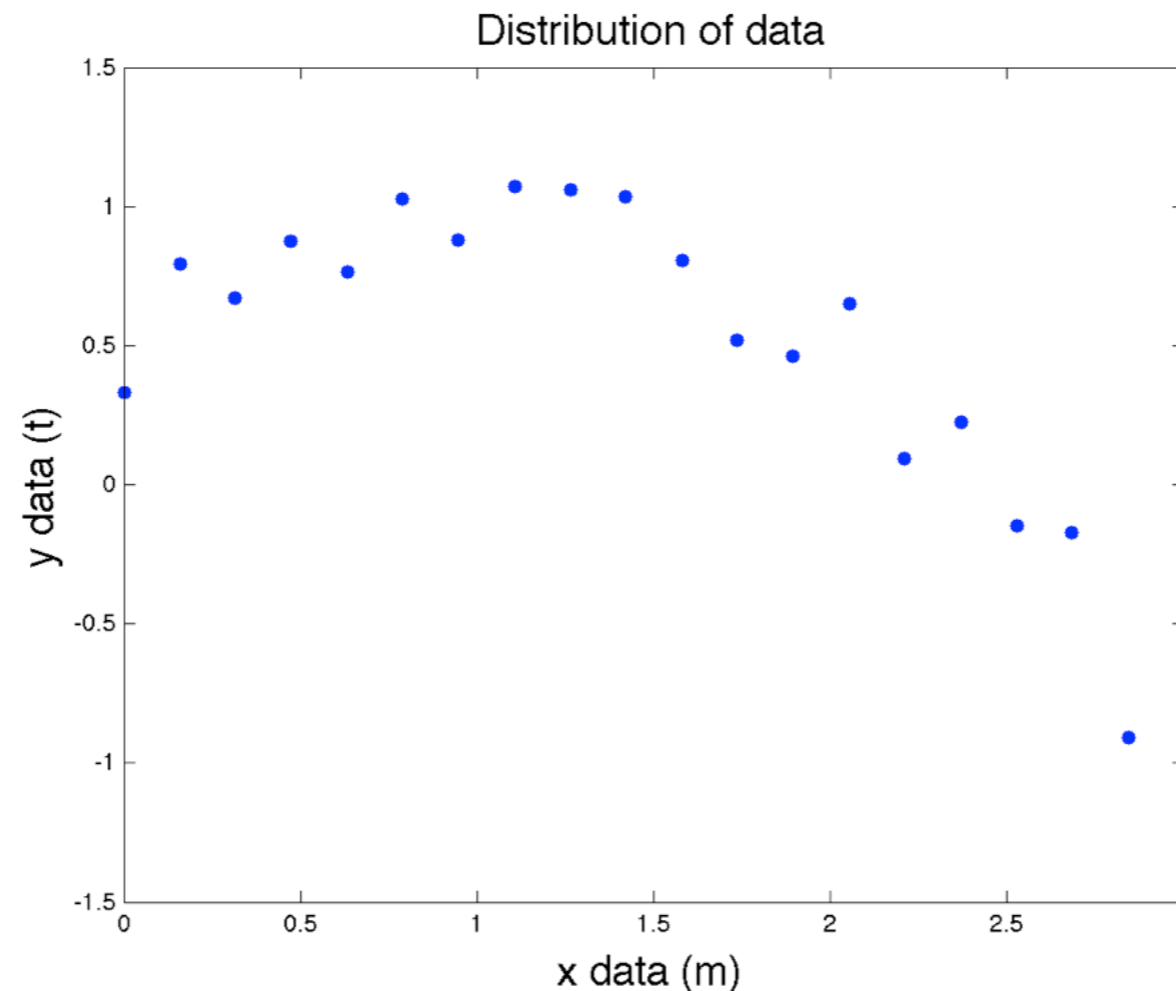


# Curve fitting

*Fitting parabolic and exponential curves*

# Linear least squares

What if our data looks like this?



A linear equation may no longer be a good model of the underlying physical process that generated the data.

# Parabolic model

A better fit might be

$$y = ax^2 + bx + c$$

**Question :**

How do you know what is a good model?

**Answer :**

Depends on what you are trying to do. In many cases, you have an understanding of the physical processes that produced the data, and so you can develop a model based on the physical assumptions.

# Fitting a parabola

Again, we are seeking model parameters. In this model, the unknown coefficients are  $a$ ,  $b$ , and  $c$ .

$$y = ax^2 + bx + c$$

Just as in the linear case, we write down an expression for each data point, assuming that the data point “solves” the model :

Given data points  $(x_i, y_i)$ , we assume

$$y_i = ax_i^2 + bx_i + c, \quad i = 1, 2, \dots, N$$

Notice that the system is still *linear* in  $a, b$ , and  $c$ .

# Fitting a parabola

We can express this as a linear system in  $a$ ,  $b$ , and  $c$  :

$$\begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ x_3^2 & x_3 & 1 \\ \vdots & \vdots & \vdots \\ x_N^2 & x_N & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_N \end{bmatrix}$$

Again, the system does not have a solution in general. We can however, find a best fit solution, again using the normal equations.

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}$$

# Solving the normal equations

We can construct the linear system as

```
% Set up linear system  
N = length(x);  
A = [x.^2 x ones(N,1)];  
b = y;
```

One could solve the normal equations directly :

**Expensive!**

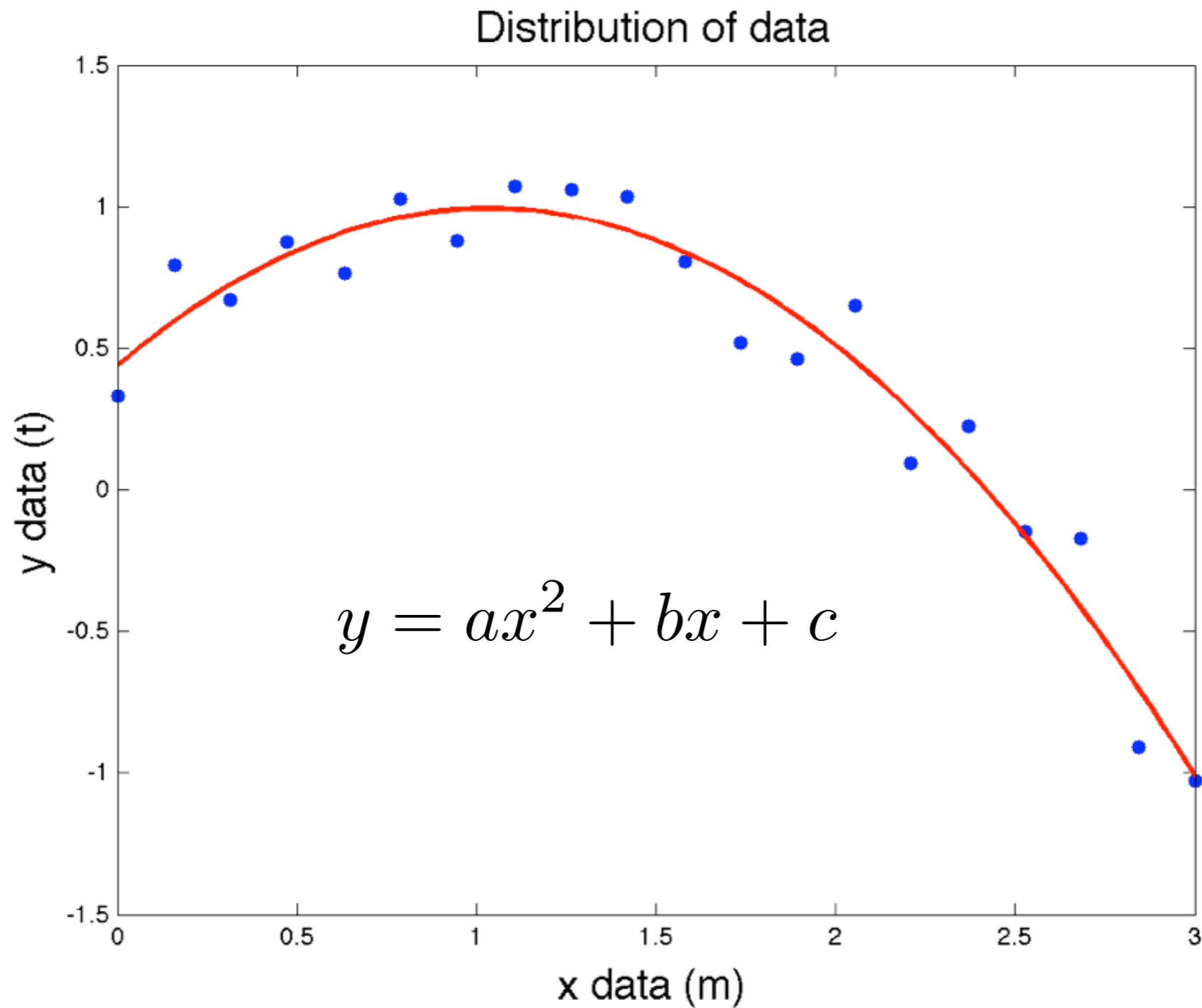
~~`coefs = inv(A'*A)*A'*b;`~~

But it is much better (and easier to remember!) to do this:

**Matlab uses a QR factorization**

`coefs = A\b;`

# Best fit parabola



a = -0.519047; b = 1.074011; c = 0.438339

# Fitting general polynomials

General polynomials can be fit using linear least squares.

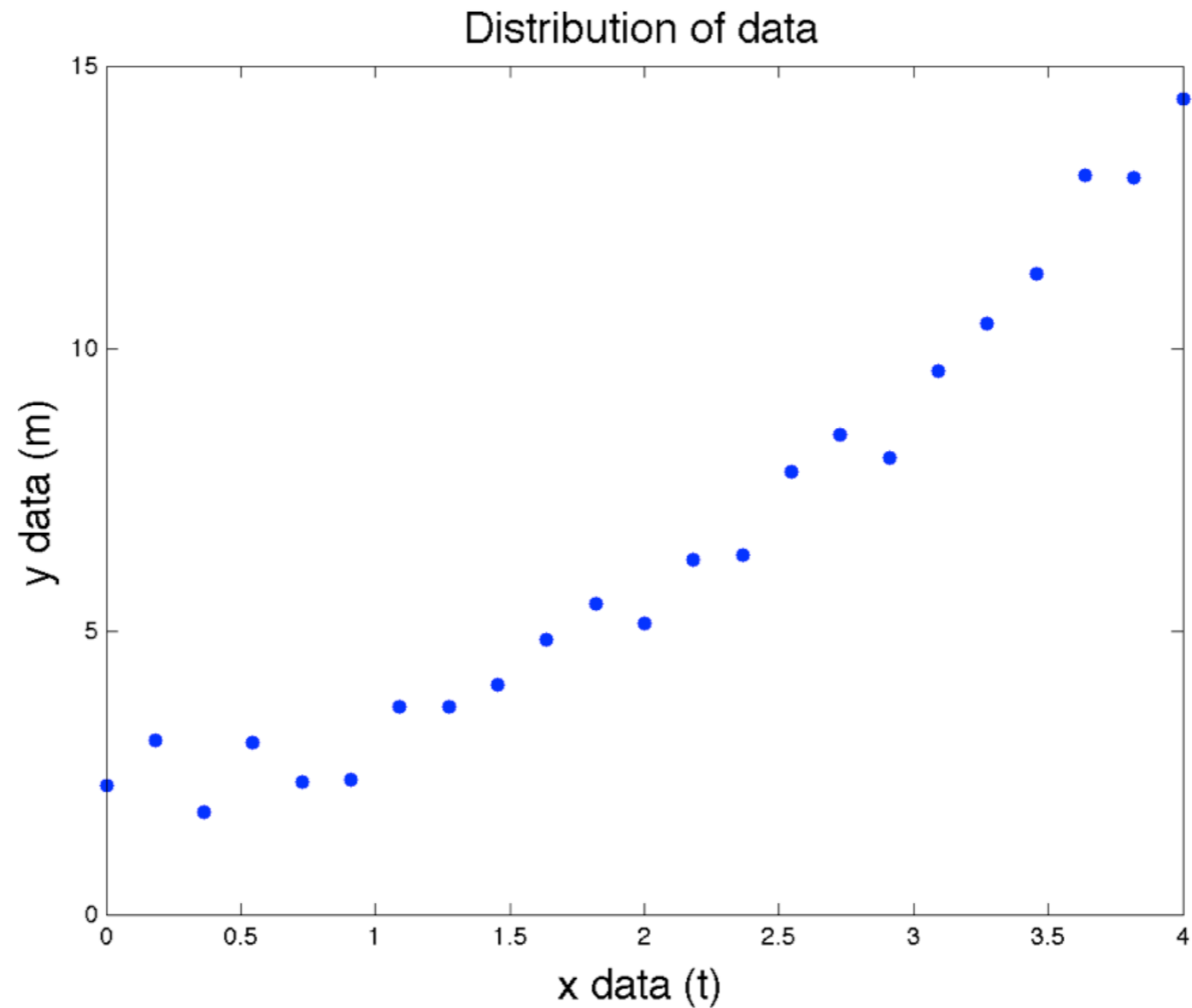
$$y = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$$

where it is assumed that the number of data points  $N$  is at least  $n + 1$ .

One must be careful, however, as the linear systems will become increasingly ill-conditioned as the  $n$  increases.



# Fitting an exponential curve



Many distributions of data may be better fit by an exponential curve. Can we still use linear least squares?

# Fitting an exponential curve

An exponential model is :

$$y = ae^{bx}$$

But the model is not linear in the parameters  $a$  and  $b$ .  
How are we going to use linear least squares?

The trick is to take the natural logarithm of both sides  
to get :

$$\ln(y) = \ln(a) + bx$$

The model is now linear in  $b$  and  $\ln(a)$ .

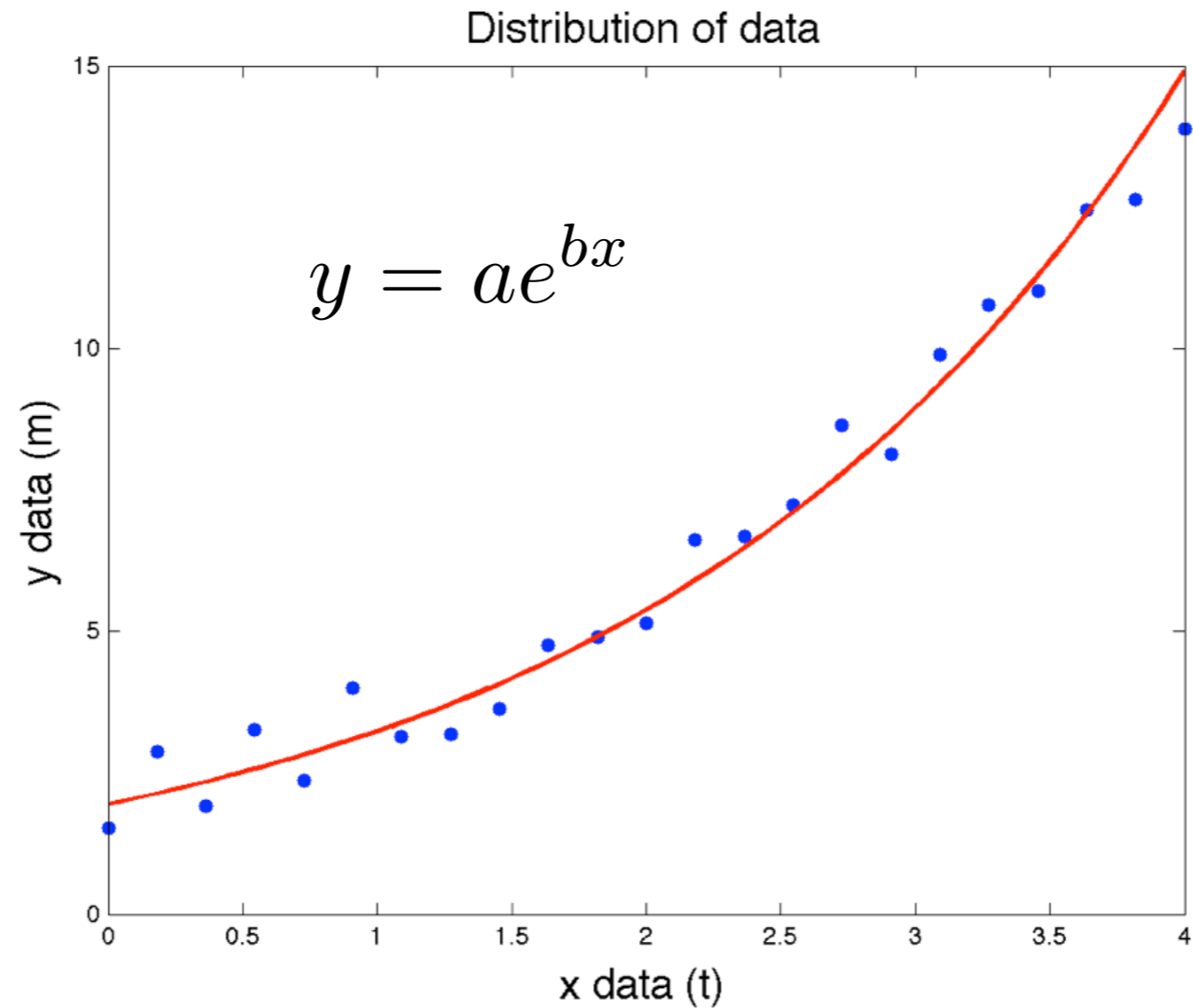
# Fitting an exponential curve

Again, we can set up a *linear* system :

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ \vdots & \vdots \\ x_N & 1 \end{bmatrix} \begin{bmatrix} b \\ \tilde{a} \end{bmatrix} = \begin{bmatrix} \ln(y_1) \\ \ln(y_2) \\ \ln(y_3) \\ \vdots \\ \ln(y_N) \end{bmatrix}, \quad \tilde{a} = \ln(a)$$

We can then compute  $a = e^{\tilde{a}}$ .

# Fitting an exponential curve



Coefficients : a = 1.944548; b = 0.509077

# Example

The temperature dependence of the reaction rate coefficient of a chemical reaction is often modeled by the Arrhenius equation

$$k = A \exp(-E_a/RT)$$

where  $k$  is the reaction rate,  $A$  is the *preexponential factor*,  $E_a$  is the activation energy,  $R$  is the universal gas constant, and  $T$  is the absolute temperature. Experimental data for a particular reaction yield the following results.

$T(K)$	773	786	797	810	810	820	834
$k$	1.63	2.95	4.19	8.13	8.19	14.9	22.2

Use a least-squares fit of this data to obtain values for  $A$  and  $E_a$  for the reaction. Take  $R = 8314 J/kg/K$ .