MATH 380

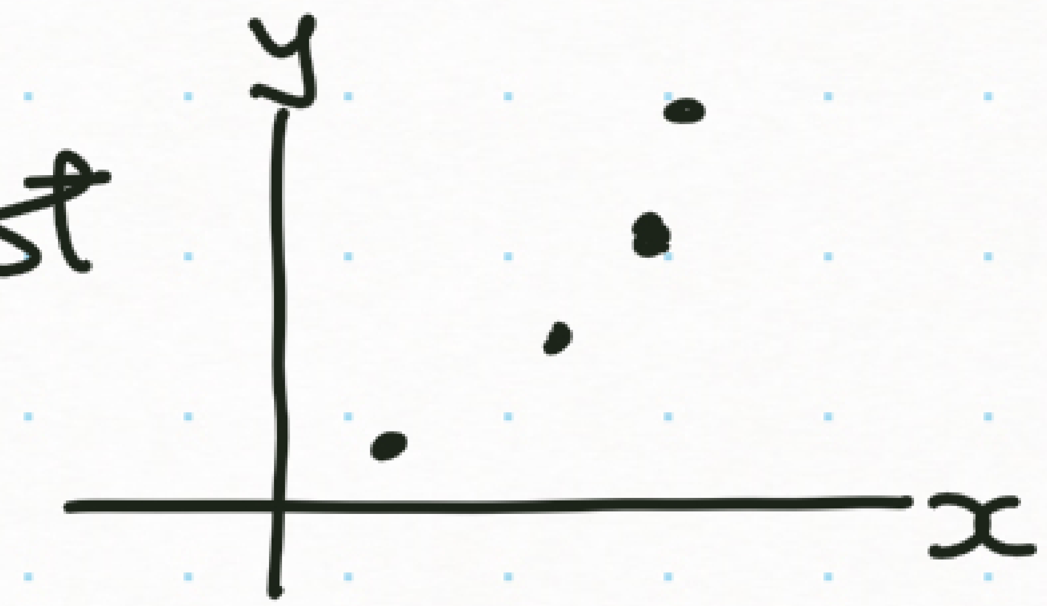
Hemanshu Kaul

kaul@iit.edu

Given data

x	1	2	3	4
y	8.1	22.1	60.1	165

Plot

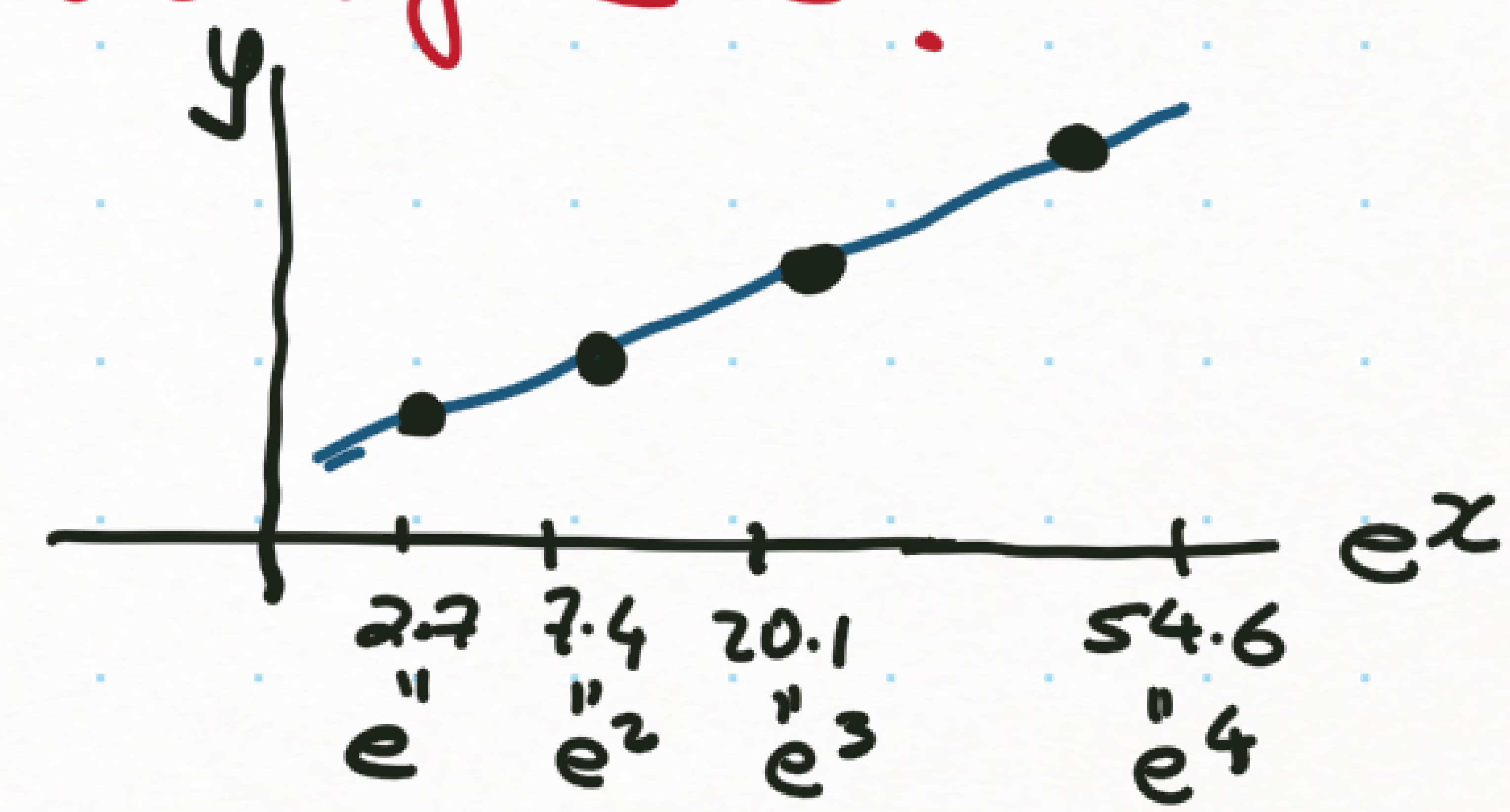


Assuming model $y \propto e^x$, i.e. $y = Ce^x$

How to find C?

• Plot e^x vs. y

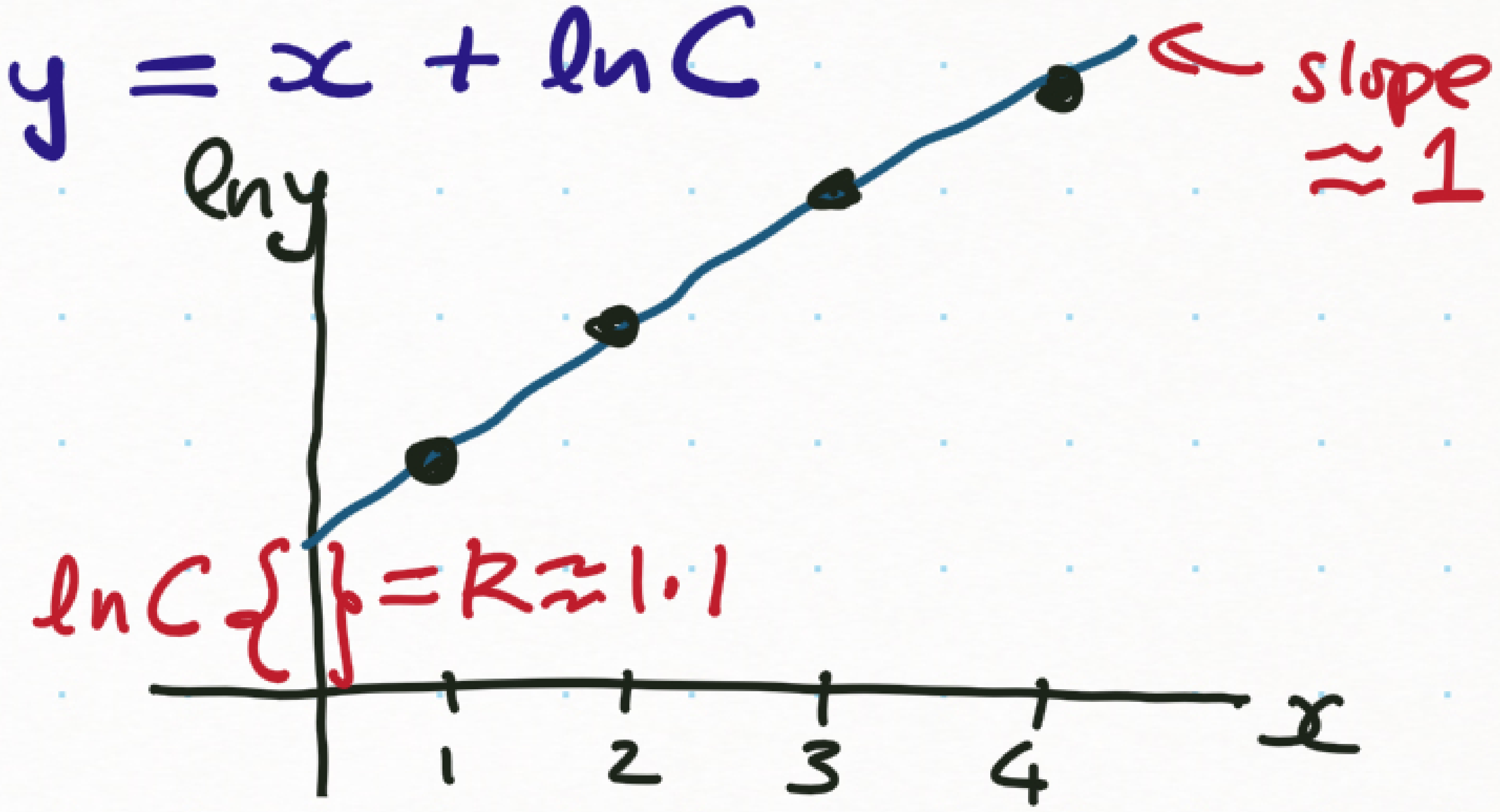
e^x	e^1	e^2	e^3	e^4
y	8.1	22.1	60.1	165



C can be approximated as the slope,

$$C \approx \frac{165 - 60.1}{e^4 - e^3} = \frac{165 - 60.1}{54.6 - 20.1} \approx 3.0$$

• $y = Ce^x \iff \ln y = \ln(Ce^x) \iff \ln y = x + \ln C$



Plot x vs. $\ln y$

x	1	2	3	4
$\ln y$	2.1	3.1	4.1	5.1

$\ln C = k \Rightarrow C = e^k \Rightarrow C = e^{1.1} \approx 3.0$

Original data (x vs. y) $\longrightarrow$ Transformed data (x vs $f(y)$) $\xrightarrow{\text{"Fit"}}$ Transformed model ($f(y) \propto x$) $\longrightarrow$ Original model ($y \propto g(x)$)

Goal Given m data points (x_i, y_i), $i = 1, 2, \dots, m$
Assuming a model (relationship) $y = f(x)$,
in particular, $f(x) = ax + b$

We want to find values for a & b which
"$\underbrace{\text{best fit}}_{?}$" the given model to the datapoints.

Pick a & b s.t. "$\underbrace{\text{overall error}}_{?}$" is minimized.

Goal Given m data points (x_i, y_i), $i = 1, 2, \ldots, m$
Assuming a model (relationship) $y = f(x)$,
in particular, $f(x) = ax + b$

We want to find values for a & b which "best fit"? the given model to the datapoints.

Pick a & b s.t. "overall error"? is minimized.

We understand error between a single observation (x_1, y_1) and the model predicted value $(x_1, f(x_1))$
i.e, $|y_1 - f(x_1)|$

But what is the error between a set of observations $(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$ and the corresponding set of predicted values $(x_1, f(x_1)), (x_2, f(x_2)), \ldots, (x_m, f(x_m))$?

We want to minimize the "distance" between two vectors of points:

$(y_1, y_2, \ldots, y_m)$ and $(f(x_1), f(x_2), \ldots, f(x_m))$

"observations at $(x_1, \ldots, x_m)$" "predicted values at $(x_1, \ldots, x_m)$"

minimize distance between two vectors in $\mathbb{R}^m$.

We want to minimize the "distance" between two vectors of points:

$(y_1, y_2, \dots, y_m)$ and $(f(x_1), f(x_2), \dots, f(x_m))$

"observations at $(x_1, \dots, x_m)$" "predicted values at $(x_1, \dots, x_m)$"

l_p-distance in $\mathbb{R}^m$

For $p \geq 1$, l_p-distance between $(y_1, y_2, \dots, y_m)$ & $(f(x_1), \dots, f(x_m))$ is defined to be $\left(\sum_{i=1}^{m} |y_i - f(x_i)|^p\right)^{1/p}$

Pay attention to $p=1$ and $p=2$

We want to minimize the "distance" between two vectors of points:

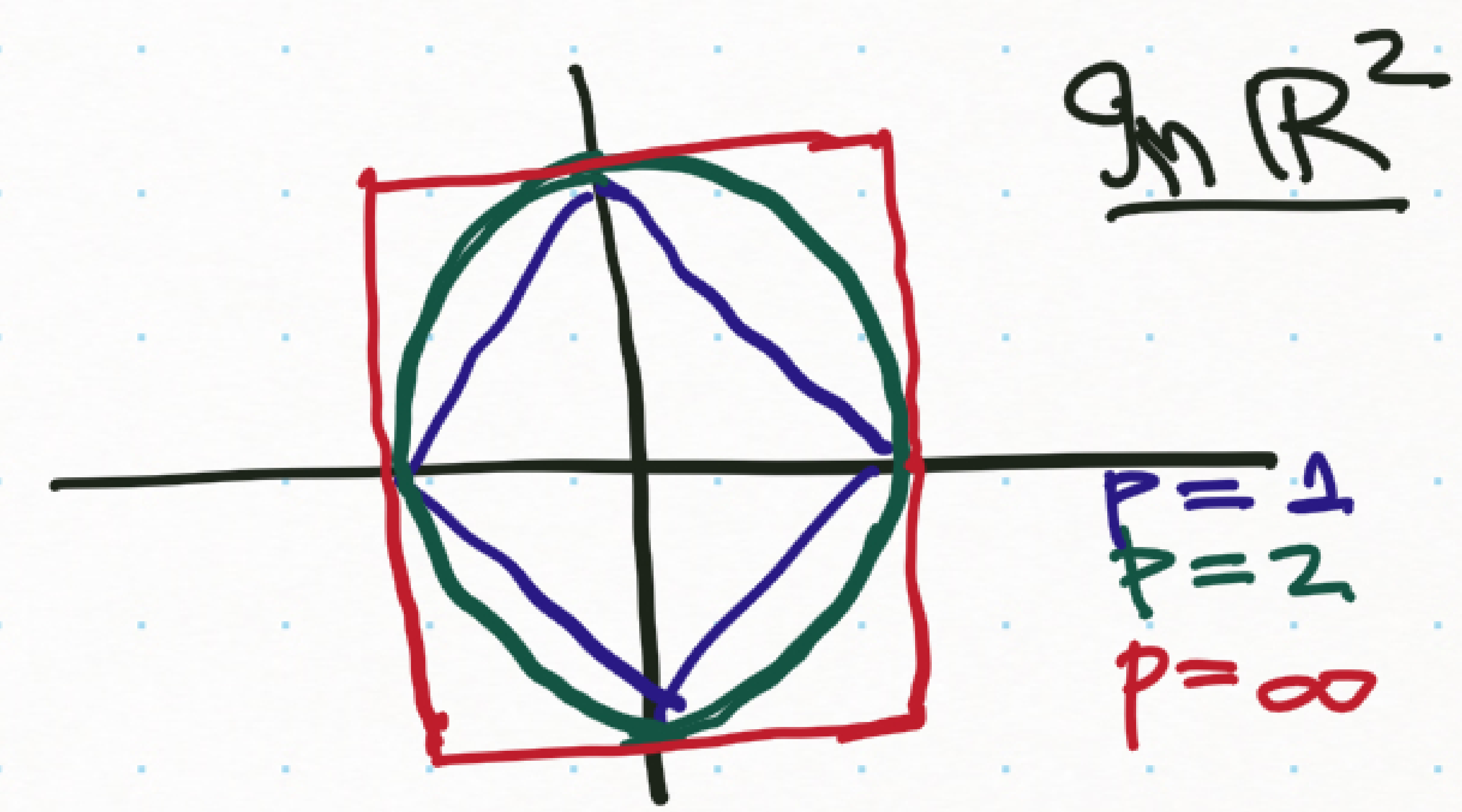
$(y_1, y_2, \ldots, y_m)$ and $(f(x_1), f(x_2), \ldots, f(x_m))$

"observations at $(x_1, \ldots, x_m)$" "predicted values at $(x_1, \ldots, x_m)$"

l_p - distance in $\mathbb{R}^m$

For $p \geq 1$, l_p-distance between $(y_1, y_2, \ldots, y_m)$ & $(f(x_1), \ldots, f(x_m))$ is defined to be $\left(\sum_{i=1}^{m} |y_i - f(x_i)|^p\right)^{1/p}$

For $p = \infty$, l_∞-distance between $(y_1, \ldots, y_m)$ & $(f(x_1), \ldots, f(x_m))$ is defined to be $\left(\max_{i=1,\ldots,m} |y_i - f(x_i)|\right)$



We want to minimize the "distance" between two vectors of points:

$(y_1, y_2, \ldots, y_m)$ and $(f(x_1), f(x_2), \ldots, f(x_m))$

"observations at $(x_1, \ldots, x_m)$" "predicted values at $(x_1, \ldots, x_m)$"

l_p-distance in $\mathbb{R}^m$

For $p \geq 1$, l_p-distance between $(y_1, y_2, \ldots, y_m)$ & $(f(x_1), \ldots, f(x_m))$ is defined to be $\left(\sum_{i=1}^{m} |y_i - f(x_i)|^p\right)^{1/p}$

For $p = \infty$, l_∞-distance between $(y_1, \ldots, y_m)$ & $(f(x_1), \ldots, f(x_m))$ is defined to be $\left(\max_{i=1,\ldots,m} |y_i - f(x_i)|\right)$

For $1 \leq p \leq \infty$, l_p-distance satisfy the metric axioms:

1. $d(\vec{u}, \vec{v}) = 0 \Leftrightarrow \vec{u} = \vec{v}$; 2. $d(\vec{u}, \vec{v}) = d(\vec{v}, \vec{u})$; 3. $d(\vec{u}, \vec{v}) \leq d(\vec{u}, \vec{w}) + d(\vec{w}, \vec{v})$

Triangle inequality

We want to minimize the "distance" between two vectors of points:

$(y_1, y_2, \ldots, y_m)$ and $(f(x_1), f(x_2), \ldots, f(x_m))$

"observations at $(x_1, \ldots, x_m)$" "predicted values at $(x_1, \ldots, x_m)$"

Three criterion for this distance

① Chebyshev Approximation Criterion

Choose f to Minimize $\max\limits_{i=1,\ldots,m} \overbrace{|y_i - f(x_i)|}^{\text{absolute deviation / error}}$

② Average Deviation Criterion

Choose f to Minimize $\sum\limits_{i=1}^{m} |y_i - f(x_i)|$

③ Least-squares Criterion

Choose f to Minimize $\sum\limits_{i=1}^{m} |y_i - f(x_i)|^2$

We want to minimize the "distance" between two vectors of points:

$(y_1, y_2, \ldots, y_m)$ and $(f(x_1), f(x_2), \ldots, f(x_m))$

"observations at $(x_1, \ldots, x_m)$" "predicted values at $(x_1, \ldots, x_m)$"

Three criterion for this distance

① Chebyshev Approximation Criterion

Choose f to Minimize $\max_{i=1,\ldots,m} \overbrace{|y_i - f(x_i)|}^{\text{absolute deviation / error}}$ ℓ_∞-distance

② Average Deviation Criterion

Choose f to Minimize $\sum_{i=1}^{m} |y_i - f(x_i)|$ ℓ_1-distance

③ Least-squares Criterion

Choose f to Minimize $\sum_{i=1}^{m} |y_i - f(x_i)|^2$ ℓ_2-distance

Given data (x_i, y_i), $i=1,\dots,m$, find "best" model of the form $y=f(x)$.

Chebyshev Approximation Criterion

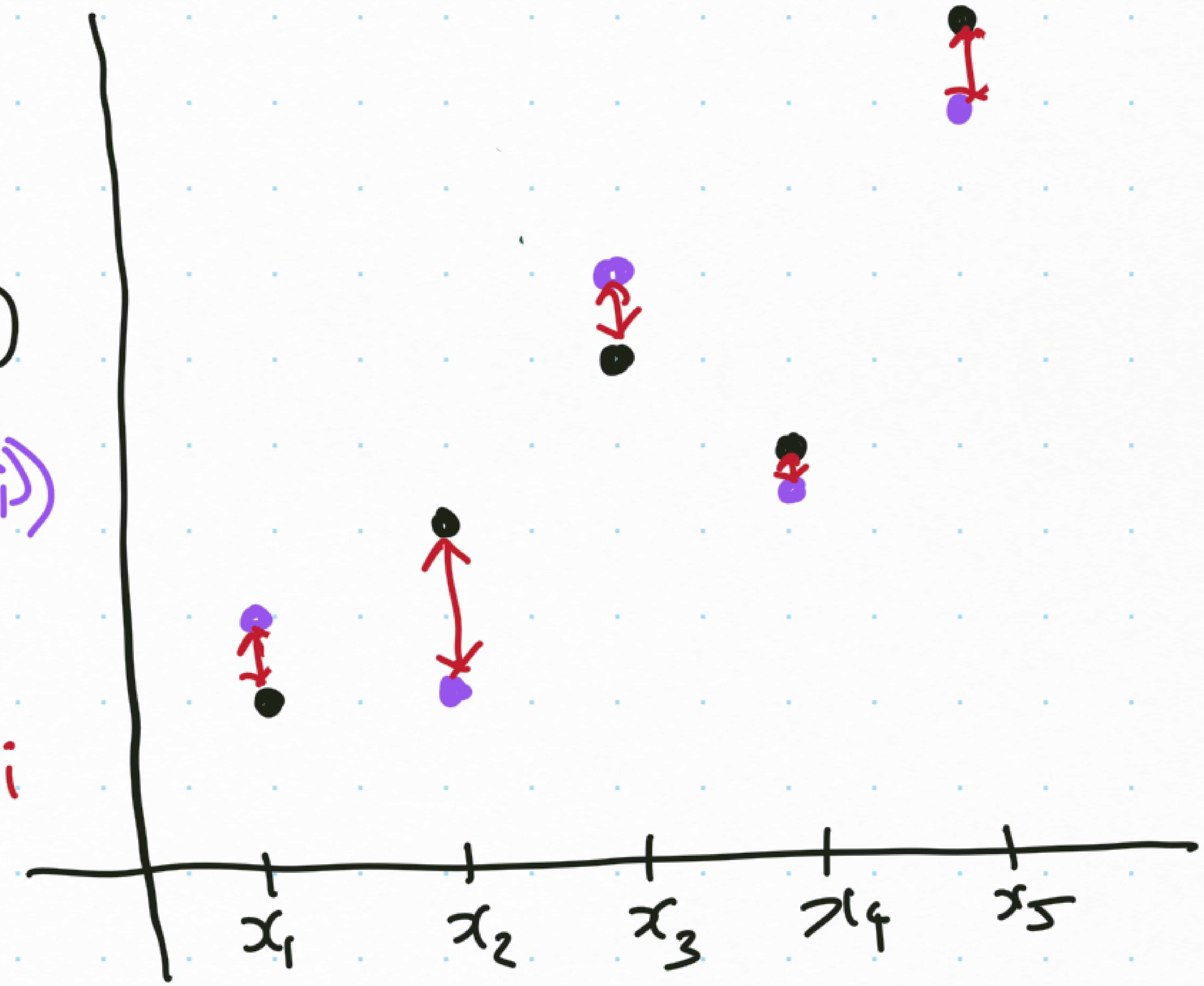
minimize the largest deviation ("pointwise error").

Minimize $\displaystyle \max_{i=1,\dots,m} |y_i - f(x_i)|$

• = (x_i, y_i)

• = $(x_i, f(x_i))$

↕ deviation at each x_i



Model #1

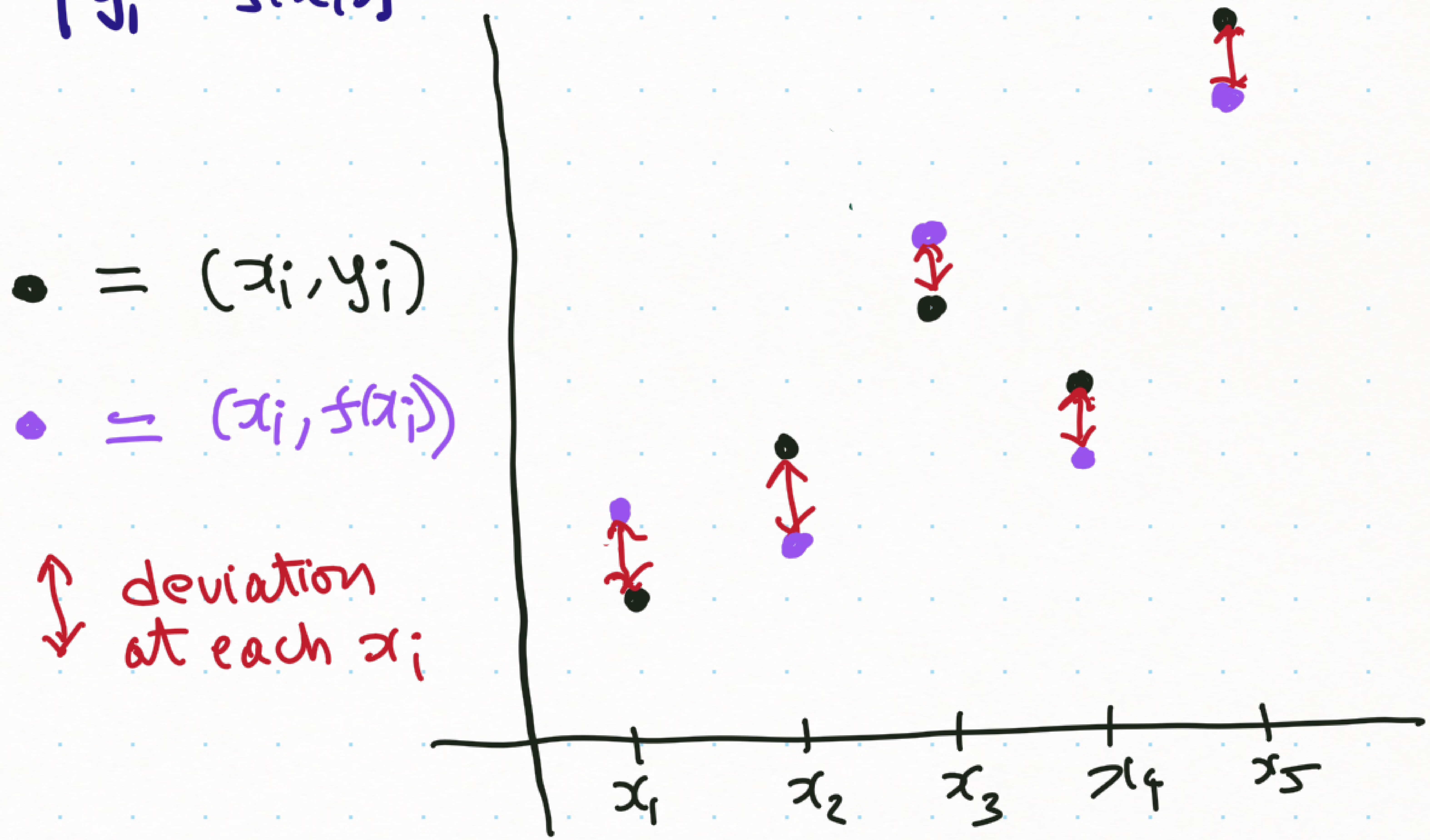
Given data $(x_i, y_i), i=1,\dots,m$, find "best" model of the form $y = f(x)$.

Chebyshev Approximation Criterion

minimize the largest deviation (pointwise error).

Minimize $\max_{i=1,\dots,m} |y_i - f(x_i)|$

Model #2



Given data (x_i, y_i), $i = 1, \ldots, m$, find "best" model of the form $y = f(x)$.

Chebyshev Approximation Criterion

minimize the largest deviation ("pointwise error").

Minimize $\max_{i=1,\ldots,m} |y_i - f(x_i)|$

Model #2 (vs. Model #1)

at x_1 worse

at x_2 better

at x_3 same

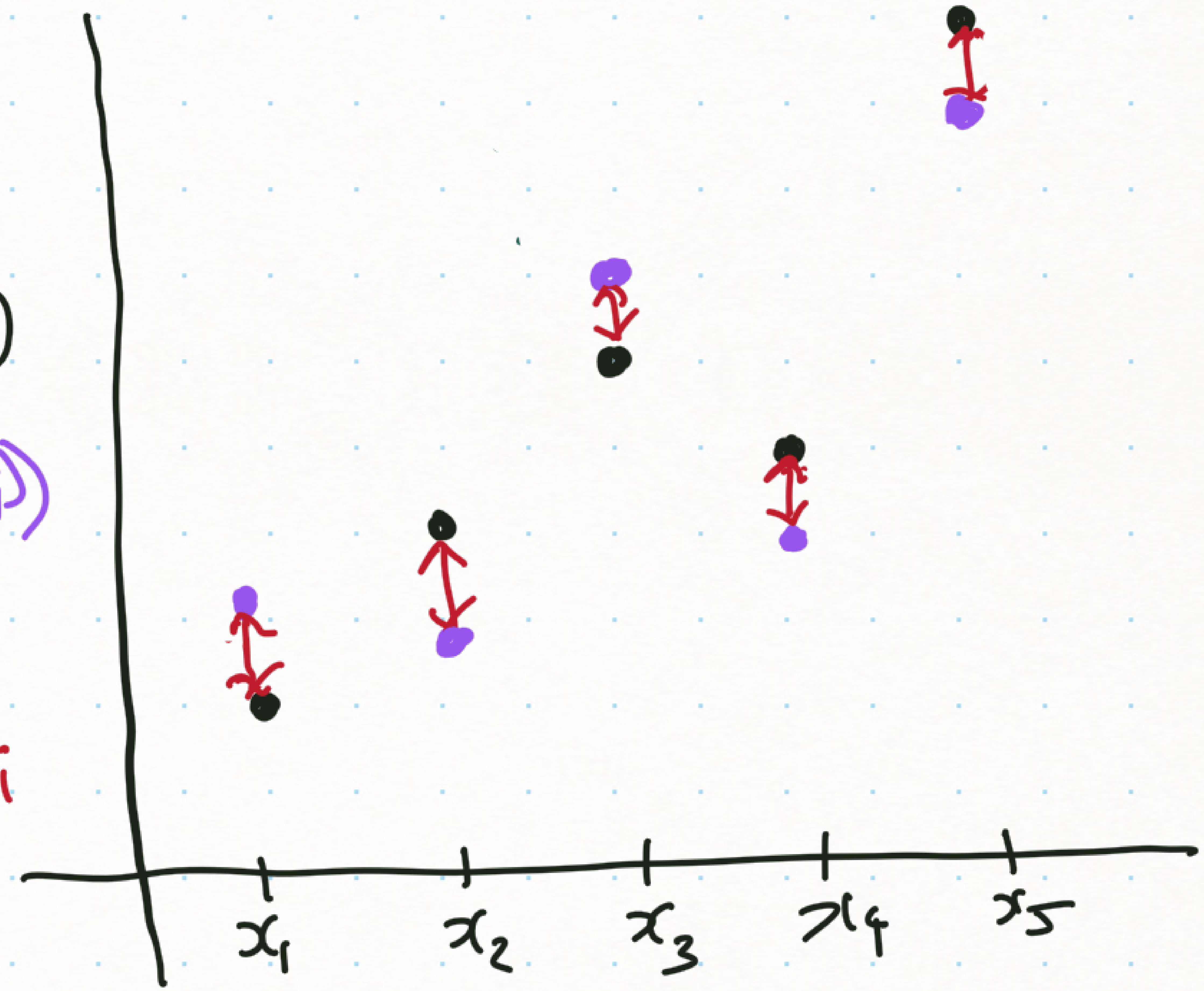
at x_4 worse

at x_5 same

● = (x_i, y_i)

● = $(x_i, f(x_i))$

$\updownarrow$ deviation at each x_i



But overall according to CAC?

Given data (x_i, y_i), $i=1,\ldots,m$, find "best" model of the form $y=f(x)$.

Chebyshev Approximation Criterion

minimize the largest deviation (pointwise error).

Minimize $\max_{i=1,\ldots,m} |y_i - f(x_i)|$

If $r = \max_{i=1,\ldots,m} |y_i - f(x_i)|$ then we want

$\min r$ such that $r \geq |y_i - f(x_i)| \ \forall i$

i.e., $r \geq y_i - f(x_i) \geq -r \ \forall i$

i.e., $r - (y_i - f(x_i)) \geq 0$ and $r + (y_i - f(x_i)) \geq 0 \ \forall i$

Given data $(x_i, y_i), i=1,\dots,m$, find "best" model of the form $y=f(x)$.

Chebyshev Approximation Criterion

minimize the largest deviation (pointwise error).

$$\text{Minimize} \quad \max_{i=1,\dots,m} |y_i - f(x_i)|$$

If $r = \max_{i=1,\dots,m} |y_i - f(x_i)|$ then we want

min r such that $r \geq |y_i - f(x_i)| \;\forall i$

i.e., $r \geq y_i - f(x_i) \geq -r \;\forall i$

i.e., $r - (y_i - f(x_i)) \geq 0$ and $r + (y_i - f(x_i)) \geq 0 \;\forall i$

We get an optimization problem:

min r ← objective function

subject to

$$\left.\begin{aligned} & r - (y_1 - f(x_1)) \geq 0 \\ & r + (y_1 - f(x_1)) \geq 0 \\ & \vdots \\ & r + (y_m - f(x_m)) \geq 0 \end{aligned}\right\} \text{2m constraints}$$

Given data $(x_i, y_i), i = 1, \ldots, m$, find "best" model of the form $y = f(x)$.

Chebyshev Approximation Criterion

minimize the largest deviation (pointwise error).

Minimize $\max_{i=1,\ldots,m} |y_i - f(x_i)|$

Our assumption is that the model is $\boxed{f(x) = ax + b.}$

So our optimization problem becomes:

min r

subject to

$$r - (y_1 - f(x_1)) \geq 0$$
$$r + (y_1 - f(x_1)) \geq 0$$
$$\vdots$$
$$r - (y_m - f(x_m)) \geq 0$$
$$r + (y_m - f(x_m)) \geq 0$$

Given data $(x_i, y_i), i=1,\dots,m$, find "best" model of the form $y = f(x)$.

Chebyshev Approximation Criterion

minimize the largest deviation (pointwise error).

Minimize $\max\limits_{i=1,\dots,m} |y_i - f(x_i)|$

Our assumption is that the model is $\boxed{f(x) = ax+b}$.

So our optimization problem becomes:

min r

subject to

$$
\begin{aligned}
r - (y_1 - (ax_1+b)) &\geq 0 \\
r + (y_1 - (ax_1+b)) &\geq 0 \\
&\vdots \\
r - (y_m - (ax_m+b)) &\geq 0 \\
r + (y_m - (ax_m+b)) &\geq 0
\end{aligned}
$$

What is known?

what is unknown?

Input data?

Variables?

Given data (x_i, y_i), $i = 1, \dots, m$, find "best" model of the form $y = f(x)$.

Chebyshev Approximation Criterion

minimize the largest deviation (pointwise error).

$$\text{Minimize} \quad \max_{i=1,\dots,m} |y_i - f(x_i)|$$

Our assumption is that the model is $f(x) = ax + b$.

So our optimization problem becomes:

$$
\begin{aligned}
&\min \ r \\
&\text{subject to} \quad r - (y_1 - (ax_1 + b)) \geq 0 \\
&\qquad\qquad\quad\ \ r + (y_1 - (ax_1 + b)) \geq 0 \\
&\qquad\qquad\qquad\quad \vdots \qquad\qquad \vdots \\
&\qquad\qquad\quad\ \ r - (y_m - (ax_m + b)) \geq 0 \\
&\qquad\qquad\quad\ \ r + (y_m - (ax_m + b)) \geq 0
\end{aligned}
$$

Only 3 variables: r, a, b. x_i, y_i are given data.

Given data $(x_i, y_i), i=1,\dots,m$, find "best" model of the form $y=f(x)$.

<u>Chebyshev Approximation Criterion</u>

minimize the largest deviation (pointwise error).

Minimize $\max\limits_{i=1,\dots,m} |y_i - f(x_i)|$

Our assumption is that the model is $f(x)=ax+b$.

So our optimization problem becomes:

$\min\ r$ ⟵ linear objective function

subject to

$$\left.\begin{array}{l} r-(y_1-(ax_1+b))\geq 0 \\ r+(y_1-(ax_1+b))\geq 0 \\ \vdots \qquad\qquad \vdots \\ r-(y_m-(ax_m+b))\geq 0 \\ r+(y_m-(ax_m+b))\geq 0 \end{array}\right\} \text{2m linear constraints}$$

Linear Program or, Linear Optimization Problem

Only 3 variables: r, a, b. x_i, y_i are given data.

Linear Programs are a building block / foundation of Discrete Optimization problems in Computer Sc., Networks, Combinatorics (see Math 435/535).

For this course, all you need is an inbuilt solver in Matlab or Mathematica or R or Python or .... .

<u>example</u> Given data

x	1.0	2.3	3.7	4.2	6.1	7.0
y	3.6	3.0	3.2	5.1	5.3	6.8

Find a, b s.t. $y = f(x) = ax + b$ minimizes the largest deviation between the data and the model.

Let r be the largest deviation.

$$\begin{aligned} \min\ & r \\ \text{s.t. } & r \geq |3.6 - f(1.0)| \\ & r \geq |3.0 - f(2.3)| \\ & \vdots \\ & r \geq |6.8 - f(7.0)| \end{aligned} \quad \Longleftrightarrow \quad \begin{aligned} \min\ & r \\ \text{s.t. } & r \geq |3.6 - (a(1.0) + b)| \\ & r \geq |3.0 - (a(2.3) + b)| \\ & \vdots \\ & r \geq |6.8 - (a(7.0) + b)| \end{aligned}$$

<u>example</u> Given data

x	1.0	2.3	3.7	4.2	6.1	7.0
y	3.6	3.0	3.2	5.1	5.3	6.8

Find a, b s.t. $y = f(x) = ax + b$ minimizes the largest deviation between the data and the model.

Let r be the largest deviation.

$$\min r$$

s.t. $r - (1.0)a - b + 3.6 \geq 0$

$r + (1.0)a + b - 3.6 \geq 0$

$\vdots$

$r - (7.0)a - b + 6.8 \geq 0$

$r + (7.0)a + b - 6.8 \geq 0$

example Given data

x	1.0	2.3	3.7	4.2	6.1	7.0
y	3.6	3.0	3.2	5.1	5.3	6.8

Find a, b s.t. $y = f(x) = ax + b$ minimizes the largest deviation between the data and the model.

Let r be the largest deviation.

$$\min r$$

s.t.

$$r - (1.0)a - b + 3.6 \geq 0$$
$$r + (1.0)a + b - 3.6 \geq 0$$
$$\vdots$$
$$r - (7.0)a - b + 6.8 \geq 0$$
$$r + (7.0)a + b - 6.8 \geq 0$$

$\leftrightarrow$ matrix form

$$\min \begin{bmatrix} \quad \end{bmatrix} \begin{bmatrix} \quad \end{bmatrix}$$

s.t. $$\begin{bmatrix} \quad \end{bmatrix} \begin{bmatrix} \quad \end{bmatrix} \geq \begin{bmatrix} \quad \end{bmatrix}$$

__example__ Given data

x	1.0	2.3	3.7	4.2	6.1	7.0
y	3.6	3.0	3.2	5.1	5.3	6.8

Find a, b s.t. $y = f(x) = ax + b$ minimizes the largest deviation between the data and the model.

Let r be the largest deviation.

$$\min r$$
s.t.
$$r - (1.0)a - b + 3.6 \geq 0$$
$$r + (1.0)a + b - 3.6 \geq 0$$
$$\vdots$$
$$r - (7.0)a - b + 6.8 \geq 0$$
$$r + (7.0)a + b - 6.8 \geq 0$$

$\leftrightarrow$ matrix form

$$\min \begin{bmatrix} & & \end{bmatrix} \begin{bmatrix} r \\ a \\ b \end{bmatrix}$$

$$\text{s.t.} \quad \underbrace{\begin{bmatrix} \\ \\ \\ \end{bmatrix}}_{2m \times 3} \begin{bmatrix} r \\ a \\ b \end{bmatrix} \geq \underbrace{\begin{bmatrix} \\ \\ \\ \end{bmatrix}}_{2m \times 1}$$

example Given data

x	1.0	2.3	3.7	4.2	6.1	7.0
y	3.6	3.0	3.2	5.1	5.3	6.8

Find a, b s.t. $y = f(x) = ax + b$ minimizes the largest deviation between the data and the model.

Let r be the largest deviation.

$$\min r$$

s.t.
$$r - (1.0)a - b + 3.6 \geq 0$$
$$r + (1.0)a + b - 3.6 \geq 0$$
$$\vdots$$
$$r - (7.0)a - b + 6.8 \geq 0$$
$$r + (7.0)a + b - 6.8 \geq 0$$

$\longleftrightarrow$ matrix form

$$\min \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} r \\ a \\ b \end{bmatrix}$$

s.t.
$$\underbrace{\begin{bmatrix} 1 & -1.0 & -1 \\ 1 & +1.0 & +1 \\ \vdots & & \\ 1 & -7.0 & -1 \\ 1 & +7.0 & +1 \end{bmatrix}}_{2m \times 3} \begin{bmatrix} r \\ a \\ b \end{bmatrix} \geq \underbrace{\begin{bmatrix} -3.6 \\ 3.6 \\ \vdots \\ -6.8 \\ 6.8 \end{bmatrix}}_{2m \times 1}$$

Given data (x_i, y_i), $i=1,\dots,m$, find "best" model of the form $y=f(x)$.

Least-Squares Criterion

minimize the sum of the squares of the deviations.

minimize $\sum_{i=1}^{m} |y_i - f(x_i)|^2$, that is, $\min \underbrace{\sum_{i=1}^{m} (y_i - f(x_i))^2}_{\text{differentiable if } f \text{ is "nice"}}$.

Our assumption is $f(x) = ax+b$, so

$\min \quad z = \sum_{i=1}^{m} (y_i - (ax_i+b))^2$ where a, b are the two unknowns

Apply Calculus (2-variable)!

Given data (x_i, y_i), $i=1,\dots,m$, find "best" model of the form $y=f(x)$.

Least-Squares Criterion

minimize the sum of the squares of the deviations.

minimize $\displaystyle\sum_{i=1}^{m} |y_i - f(x_i)|^2$, that is, $\underbrace{\min \displaystyle\sum_{i=1}^{m} (y_i - f(x_i))^2}_{\text{differentiable if } f \text{ is "nice".}}$

Our assumption is $f(x) = ax+b$, so

$$\min \quad z = \sum_{i=1}^{m} (y_i - (ax_i + b))^2$$

where a, b are the two unknowns

Apply Calculus (2-variable)!

Set $\dfrac{\partial z}{\partial a} = 0$ and $\dfrac{\partial z}{\partial b} = 0$

$$\left.\begin{aligned} \frac{\partial z}{\partial a} &= \sum_i 2(y_i - ax_i - b)(-x_i) = 0 \\ \frac{\partial z}{\partial b} &= \sum_i 2(y_i - ax_i - b)(-1) = 0 \end{aligned}\right\} \Rightarrow$$

Solve for a & b

$$\boxed{\begin{aligned} a\sum_i x_i^2 + b\sum_i x_i &= \sum_i x_i y_i \\ a\sum_i x_i + bm &= \sum_i y_i \end{aligned}}$$

Normal Equations

Given data (x_i, y_i), $i = 1, \dots, m$, find "best" model of the form $y = f(x)$.

Least-Squares Criterion

minimize the sum of the squares of the deviations.

minimize $\sum_{i=1}^{m} |y_i - f(x_i)|^2$, that is, $\underbrace{\min \sum_{i=1}^{m} (y_i - f(x_i))^2}$

differentiable if f is "nice".

Our assumption is $f(x) = ax + b$, so

$$\min \quad z = \sum_{i=1}^{m} (y_i - (ax_i + b))^2$$

where a, b are the two unknowns

Solving the normal equations for a & b gives

$$a = \frac{m \sum x_i y_i - \sum x_i \sum y_i}{m \sum x_i^2 - (\sum x_i)^2}$$

$$b = \frac{\sum x_i^2 \sum y_i - \sum x_i y_i \sum x_i}{m \sum x_i^2 - (\sum x_i)^2}$$

Using 2nd derivative test from Calc III, we can verify that this critical point is a local minimum.

e.g.

x	1	5	8
y	1	10	6

Estimate parameters for fitting the model $y = ax + b$ using the least squares criterion.

We want to find a, b such that $L(a,b) = \sum_{i=1}^{m} (y_i - f(x_i))^2$ is minimized.

$$L(a,b) = \sum_{i=1}^{3} (y_i - (ax_i + b))^2 = (1 - (a(1) + b))^2 + (10 - (a(5) + b))^2 + (6 - (a(8) + b))^2$$

By setting $\frac{\partial L}{\partial a} = 0$ & $\frac{\partial L}{\partial b} = 0$, we get the Normal Equations.

Solving the Normal Equations using Matlab | Mathematica | ..

we get $a = \frac{59}{74} \approx 0.797$ $\quad b = \frac{72}{37} \approx 1.946$

and $L\left(\frac{59}{74}, \frac{72}{37}\right) = \frac{1844}{74} \approx 24.986$ is the minimum possible value of the sum of the squares of deviations over all choices of a & b

Models for the Data

x	0.5	1.0	1.5	2.0	2.5
y	0.7	3.4	7.2	12.4	20.1

$m = 5$ data points

① Model $y \propto x^2$

Estimate parameters of $y = ax^2$ using

Models for the Data

x	0.5	1.0	1.5	2.0	2.5
y	0.7	3.4	7.2	12.4	20.1

$m = 5$ data points

① Model $y \propto x^2$

Estimate parameters of $y = ax^2$ using Chebyshev Approx. Criterion

Models for the Data

x	0.5	1.0	1.5	2.0	2.5
y	0.7	3.4	7.2	12.4	20.1

$m=5$ data points

① Model $y \propto x^2$

Estimate parameters of $y = ax^2$ using Chebyshev Approx. Criterion

Apply CAC to the transformed data

x^2	$(0.5)^2$	$(1.0)^2$	$(1.5)^2$	$(2.0)^2$	(2.5)
y	0.7	3.4	7.2	12.4	20.1

As discussed before, ?

set up the Linear Program

min r

s.t

$$\left.\begin{aligned} r - (y_i - (ax_i^2 + \underbrace{b}_{=0})) &\geq 0 \\ r + (y_i - (ax_i^2 + b)) &\geq 0 \end{aligned}\right\} \text{ for } i = 1,2,3,4,5$$

↑ note the data we are using y_i vs. x_i^2

Models for the Data

x	0.5	1.0	1.5	2.0	2.5
y	0.7	3.4	7.2	12.4	20.1

$m=5$ data points

① Model $y \propto x^2$
Estimate parameters of $y = ax^2$ using Chebyshev Approx. Criterion

Apply CAC to the transformed data

x^2	$(0.5)^2$	$(1.0)^2$	$(1.5)^2$	$(2.0)^2$	(2.5)
y	0.7	3.4	7.2	12.4	20.1

As discussed before, set up the Linear Program

min r
s.t

$$\left.\begin{aligned} r - (y_i - (ax_i^2 + \cancel{b})) &\geq 0 \\ r + (y_i - (ax_i^2 + \cancel{b})) &\geq 0 \end{aligned}\right\} \text{ for } i = 1,2,3,4,5$$

Convert into matrix form.
Input into a solver, and get the optimal values for $a, b = 0, r$.

$$\boxed{y = 3.171x^2}$$

$r = 0.2829$

Models for the Data

x	0.5	1.0	1.5	2.0	2.5
y	0.7	3.4	7.2	12.4	20.1

$m=5$ data points

② Model $y \propto x^2$

Estimate parameters of $y = ax^2$ using Least Squares Criterion.

Apply LSC to the transformed data

x^2	$(0.5)^2$	$(1.0)^2$	$(1.5)^2$	$(2.0)^2$	(2.5)
y	0.7	3.4	7.2	12.4	20.1

Models for the Data

x	0.5	1.0	1.5	2.0	2.5
y	0.7	3.4	7.2	12.4	20.1

$m = 5$ data points

② Model $y \propto x^2$

Estimate parameters of $y = ax^2$ using Least Squares Criterion.

Apply LSC to the transformed data

x^2	$(0.5)^2$	$(1.0)^2$	$(1.5)^2$	$(2.0)^2$	(2.5)
y	0.7	3.4	7.2	12.4	20.1

As discussed before, set up the Normal Equations & solve them for optimal values of a & $b=0$

or more simply

$\frac{dL}{da} = 0$, since no b.

single variable calculus

Models for the Data

x	0.5	1.0	1.5	2.0	2.5
y	0.7	3.4	7.2	12.4	20.1

$m=5$ data points

② Model $y \propto x^2$

Estimate parameters of $y = ax^2$ using Least Squares Criterion.

Apply LSC to the transformed data

x^2	$(0.5)^2$	$(1.0)^2$	$(1.5)^2$	$(2.0)^2$	(2.5)
y	0.7	3.4	7.2	12.4	20.1

As discussed before, set up the Normal Equations & solve them for optimal values of a & $b=0$

or more simply $\frac{dL}{da} = 0$, since no b.

$$\boxed{y = (3.1869)x^2}$$

& min sum of squares of deviations is 0.20954

Models for the Data

x	0.5	1.0	1.5	2.0	2.5
y	0.7	3.4	7.2	12.4	20.1

$m=5$ data points

(3) Are we sure $y \propto x^2$? Maybe $y \propto x^b$ for some unknown b?

Model $y \propto x^b$

Estimate parameters of $y = ax^b$ using Least Squares Criterion.

Models for the Data

x	0.5	1.0	1.5	2.0	2.5
y	0.7	3.4	7.2	12.4	20.1

$m=5$ data points

③ Are we sure $y \propto x^2$? Maybe $y \propto x^b$ for some unknown b?

Model $y \propto x^b$

Estimate parameters of $y = ax^b$ using Least Squares Criterion.

b is unknown so we don't know how to transform data directly.

$$y = ax^b \Leftrightarrow \underbrace{\ln y}_{\text{t. data}} = \ln a + b \underbrace{\ln x}_{\text{t. data}}$$

unknown parameters

$\longleftarrow$ linear fit to the transformed data $(\ln x_i, \ln y_i)$] Transformed LSC

Apply LSC to find A & B in $Y = A + BX$ using data $(X_i, Y_i) = ?$

Models for the Data

x	0.5	1.0	1.5	2.0	2.5
y	0.7	3.4	7.2	12.4	20.1

$m=5$ data points

(3) Are we sure $y \propto x^2$? Maybe $y \propto x^b$ for some unknown b?

Model $y \propto x^b$

Estimate parameters of $y = ax^b$ using Least Squares Criterion.

b is unknown so we don't know how to transform data directly.

$$y = ax^b \Leftrightarrow \underbrace{\ln y}_{\text{t. data}} = \ln a + b \underbrace{\ln x}_{\text{t. data}}$$

unknown parameters

$\leftarrow$ linear fit to the transformed data $(\ln x_i, \ln y_i)$ — Transformed LSC

Apply LSC to find A & B in $Y = A + BX$ using data $(X_i, Y_i) = (\ln x_i, \ln y_i)$

X_i	$\ln x_1$	$\ln x_2$	$\cdots$	$\ln x_5$
Y_i	$\ln y_1$	$\ln y_2$		$\ln y_5$

we will get $A = 1.1266$, $B = 2.063$

i.e. $\ln y = 1.1266 + 2.063 \ln x$

Models for the Data

x	0.5	1.0	1.5	2.0	2.5
y	0.7	3.4	7.2	12.4	20.1

$m=5$ data points

③ Are we sure $y \propto x^2$? Maybe $y \propto x^b$ for some unknown b?

Model $y \propto x^b$

Estimate parameters of $y = a x^b$ using Least Squares Criterion.

b is unknown so we don't know how to transform data directly.

$$y = a x^b \iff \underbrace{\ln y}_{\text{t. data}} = \ln a + b \underbrace{\ln x}_{\text{t. data}}$$

unknown parameters

$\longleftarrow$ linear fit to the transformed data $(\ln x_i, \ln y_i)$] Transformed LSC

Apply LSC to find A & B in $Y = A + BX$ using data $(X_i, Y_i) = (\ln x_i, \ln y_i)$

we will get $A = 1.1266$, $B = 2.063$

i.e. $\ln y = 1.1266 + 2.063 \ln x$

X_i	$\ln x_1$	$\ln x_2$	$\cdots$	$\ln x_5$
Y_i	$\ln y_1$	$\ln y_2$		$\ln y_5$

In original variables] $y = e^{1.1266} x^{2.063}$ since $a = e^A$, $b = B$

i.e., $\boxed{y = (3.0852)\, x^{2.063}}$

Models for the Data

x	0.5	1.0	1.5	2.0	2.5
y	0.7	3.4	7.2	12.4	20.1

$m = 5$ data points

(4) The exponent 2.063 is so close to 2, might as well try $y \propto x^2$

Model $y \propto x^2$

Models for the Data

x	0.5	1.0	1.5	2.0	2.5
y	0.7	3.4	7.2	12.4	20.1

$m=5$ data points

④ The exponent 2.063 is so close to 2, might as well try $y \propto x^2$

Model $y \propto x^2$

Estimate parameters of $y = ax^2$ using the transformed LSC.

$$y = ax^2 \iff \ln y = \ln a + 2\ln x \iff Y = A + 2X$$

$\leftarrow$ Apply Transforme LSC to the data $(X, Y) = (\ln x, \ln y)$

Models for the Data

x	0.5	1.0	1.5	2.0	2.5
y	0.7	3.4	7.2	12.4	20.1

$m = 5$ data points

④ The exponent 2.063 is so close to 2, might as well try $y \propto x^2$

Model $y \propto x^2$

Estimate parameters of $y = ax^2$ using the transformed LSC.

$y = ax^2 \Leftrightarrow \ln y = \ln a + 2\ln x \Leftrightarrow Y = A + 2X \leftarrow$ Apply Transformed LSC to the data $(X, Y) = (\ln x, \ln y)$

LSC applied to $Y = A + 2X$ with data

X	$\ln x_i$	...
Y	$\ln y_i$	...

gives $Y = 1.1432 + 2X$

Models for the Data

x	0.5	1.0	1.5	2.0	2.5
y	0.7	3.4	7.2	12.4	20.1

$m=5$ data points

④ The exponent 2.063 is so close to 2, might as well try $y \propto x^2$

Model $y \propto x^2$

Estimate parameters of $y = ax^2$ using the transformed LSC.

$y = ax^2 \Leftrightarrow \ln y = \ln a + 2\ln x \Leftrightarrow Y = A + 2X \leftarrow$ Apply Transformed LSC to the data $(X,Y) = (\ln x, \ln y)$

LSC applied to $Y = A + 2X$ with data

X	$\ln x_i$	$\cdots$
Y	$\ln y_i$	$\cdots$

gives $Y = 1.1432 + 2X$

In original variables, $\boxed{y = 3.1368x^2}$

since $A = \ln a \Leftrightarrow a = e^A = e^{1.1432} = 3.1368$

Models for the Data

x	0.5	1.0	1.5	2.0	2.5
y	0.7	3.4	7.2	12.4	20.1

$m = 5$ data points

(5) Someone else comes along & says its not really a quadratic relation but an exponential relation.

$y \propto e^{bx}$, i.e., $y = a e^{bx}$

Models for the Data

x	0.5	1.0	1.5	2.0	2.5
y	0.7	3.4	7.2	12.4	20.1

$m=5$ data points

5. Someone else comes along & says its not really a quadratic relation but an exponential relation.

$y \propto e^{bx}$, i.e., $y = a e^{bx}$

We can estimate parameters using transformed LSC again.

$$y = a e^{bx} \iff \ln y = \ln a + bx \iff Y = A + BX$$

where $A = \ln a$ or $a = e^{A}$ with data $(X, Y) = (x, \ln y)$

and $B = b$

⋮

$y = (\cdots) e^{(\cdots)x}$ & so on.

So Far

→ Based on our understanding of the phenomenon under study, either based on some qualitative assumptions or based on a numerical study (based on observations/data) we propose a model (or many models :-)

So Far

→ Based on our understanding of the phenomenon under study, either based on some qualitative assumptions or based on a numerical study (based on observations/data) we propose a model (or many models :-))

→ These models will have some unknown parameters. We estimate these unknown constants based on either CAC or LSC (with or without transformations) using the given data.

So Far

→ Based on our understanding of the phenomenon under study, either based on some qualitative assumptions or based on a numerical study (based on observations/data) we propose a model (or many models :-)

→ These models will have some unknown parameters we estimate these unknown constants based on either CAC or LSC (with or without transformations) using the given data.

no matter how they were created

→ When we have several models for the same observations/data we can compare them quantitatively using CAC and LSC

Based on the data (& in order to understand it) we created several models for

x	0.5	1.0	1.5	2.0	2.5
y	0.7	3.4	7.2	12.4	20.1

① $y = (3.171)x^2$ (based on CAC for $y \propto x^2$)

② $y = (3.1869)x^2$ (based on LSC for $y \propto x^2$)

③ $y = (3.0852)x^{2.063}$ (based on transformed LSC for $y \propto x^b$)

④ $y = (3.1368)x^2$ (based on transformed LSC for $y \propto x^2$)

Based on the data (& in order to understand it) we created several models for

x	0.5	1.0	1.5	2.0	2.5
y	0.7	3.4	7.2	12.4	20.1

① $y = (3.171)x^2$ (based on CAC for $y \propto x^2$)

② $y = (3.1869)x^2$ (based on LSC for $y \propto x^2$)

③ $y = (3.0852)x^{2.063}$ (based on transformed LSC for $y \propto x^b$)

④ $y = (3.1368)x^2$ (based on transformed LSC for $y \propto x^2$)

By calculating the deviations at each of the 15 original data points for each of these 4 models,

we can calculate the
→ sum of squares of deviations (LSC)
→ max of absolute deviations (CAC)

for each model.

Data		Model #1	Model #2	Model #3	Model #4
x_i	y_i	$y_i - 3.171x_i^2$	$y_i - 3.1869x^2$	$y_i - 3.0852x^{2.063}$	$y_i - 3.137x^2$
0.5	0.7	−0.0927	−0.0967	−0.0384	−0.0842
1.0	3.4	0.2293	0.2131	0.3148	0.2632
1.5	7.2	0.0659	0.0295	0.0792	0.1422
2.0	12.4	−0.2829	−0.3476	−0.4899	−0.1472
2.5	20.1	0.28293	0.18187	−0.3247	0.4950

Data		Model #1	Model #2	Model #3	Model #4
x_i	y_i	$y_i - 3.171 x_i^2$	$y_i - 3.1869 x^2$	$y_i - 3.0852 x^{2.063}$	$y_i - 3.137 x^2$
0.5	0.7	−0.0927	−0.0967	−0.0384	−0.0842
1.0	3.4	0.2293	0.2131	0.3148	0.2632
1.5	7.2	0.0659	0.0295	0.0792	0.1422
2.0	12.4	−0.2829	−0.3476	−0.4899	−0.1472
2.5	20.1	0.28293	0.18187	−0.3247	0.4950

	Model #1	Model #2	Model #3	Model #4
LSC = $\sum (y_i - f(x_i))^2$	0.2256	0.2095	0.4523	0.3633
CAC = $\max \lvert y_i - f(x_i) \rvert$	0.2823	0.3476	0.4899	0.4950