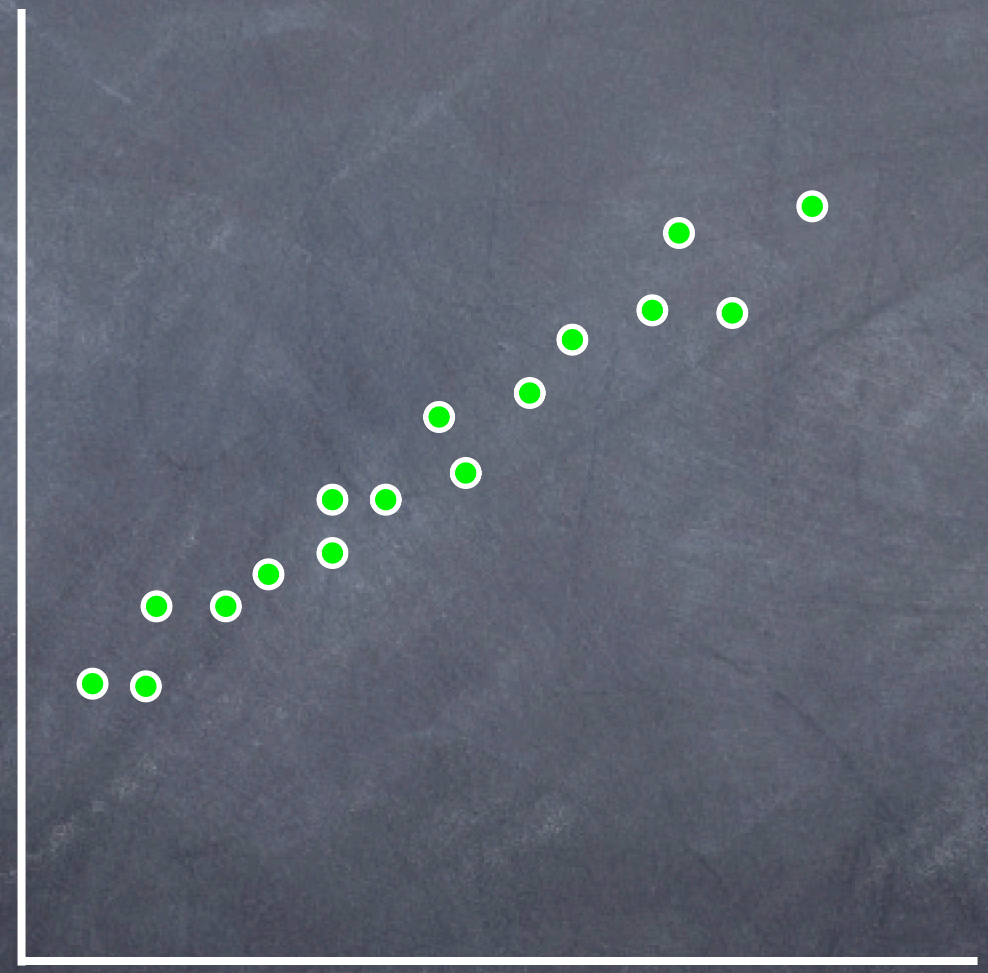
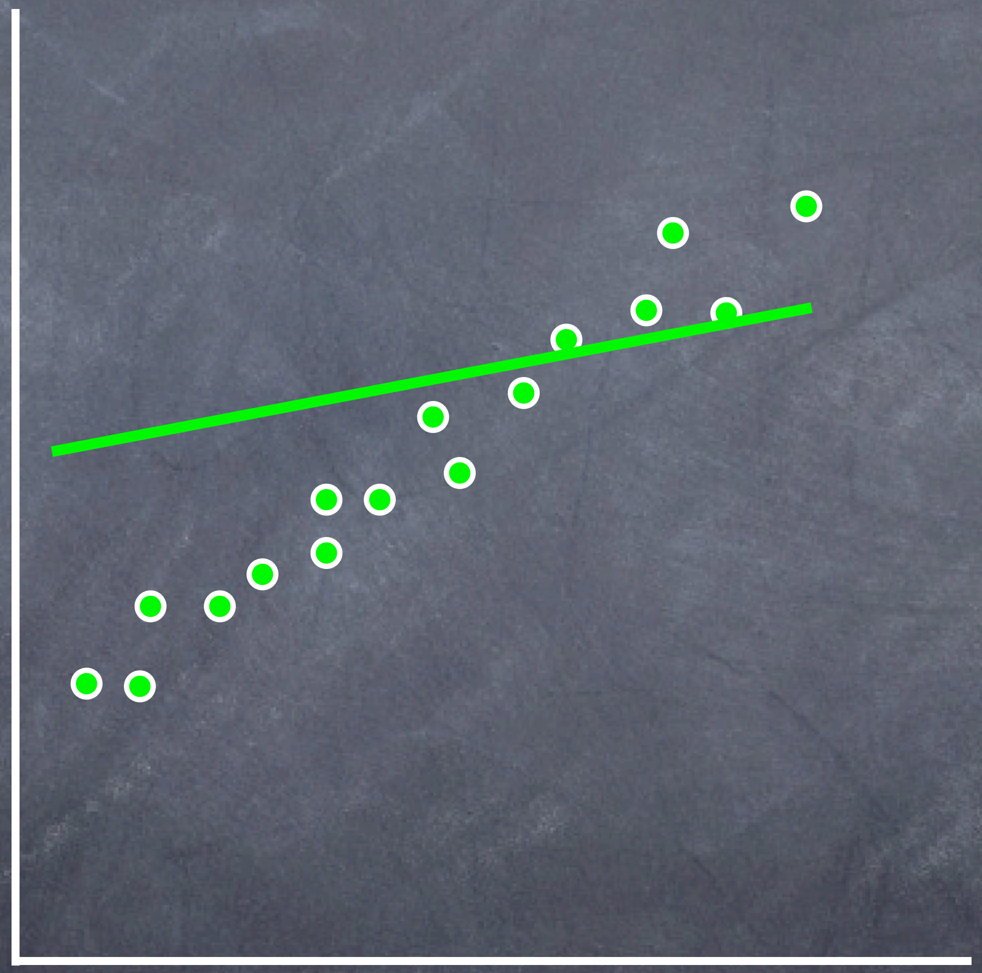
Today

- Linear regression
  - aka linear least squares
  - aka fitting data with straight lines
- Another optimization example (if time allows)

Linear regression

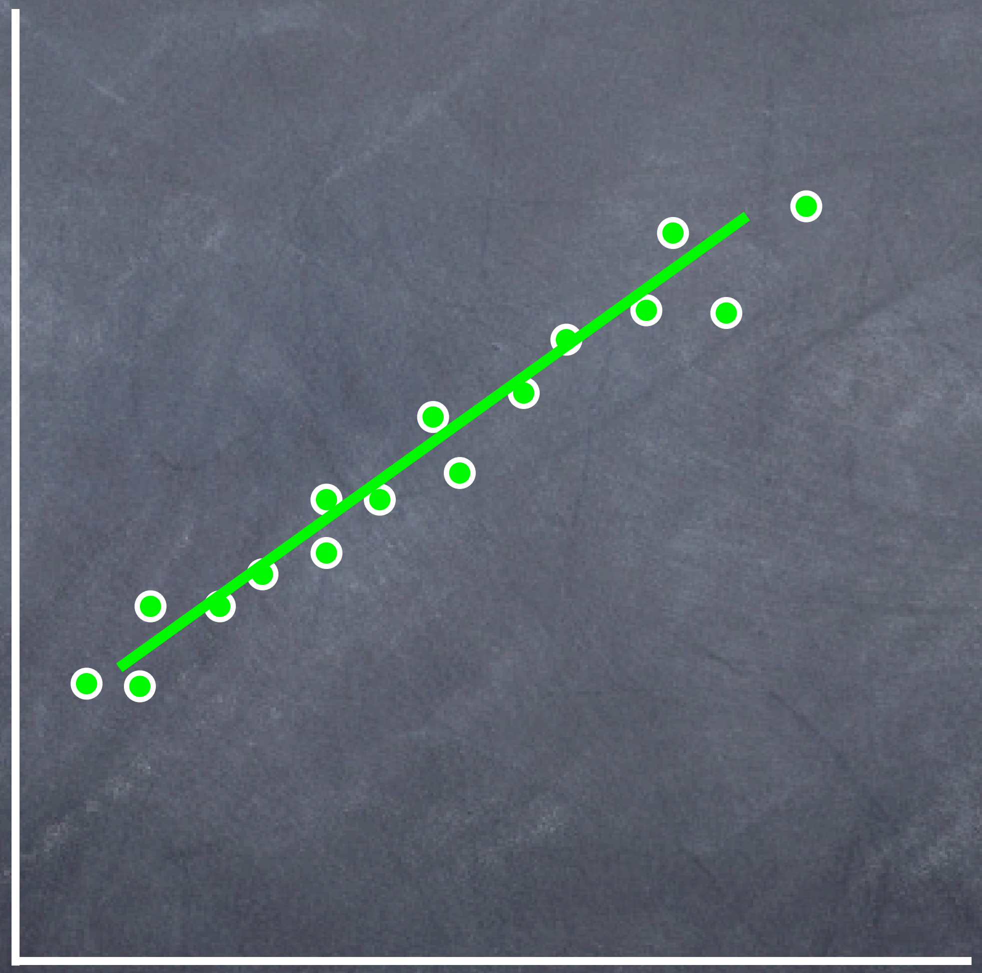


Linear regression



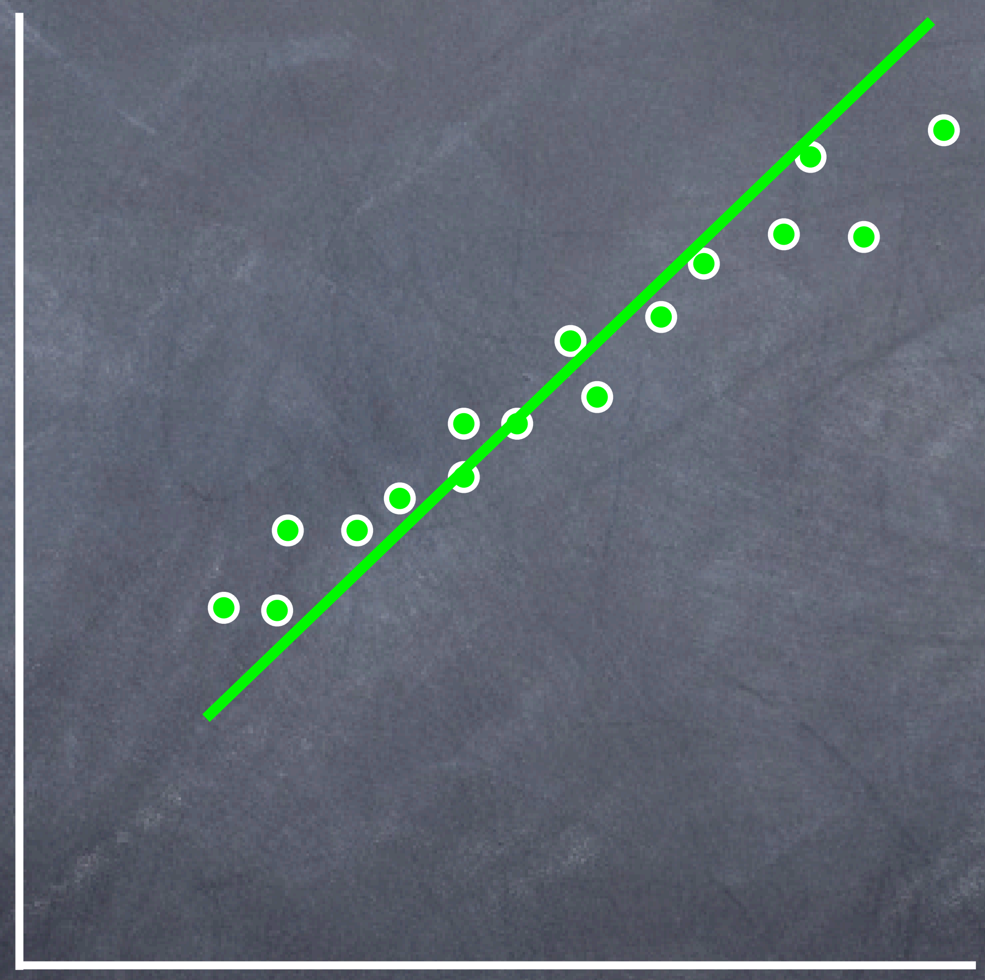
How do we find the best line
to fit through the data?

Linear regression



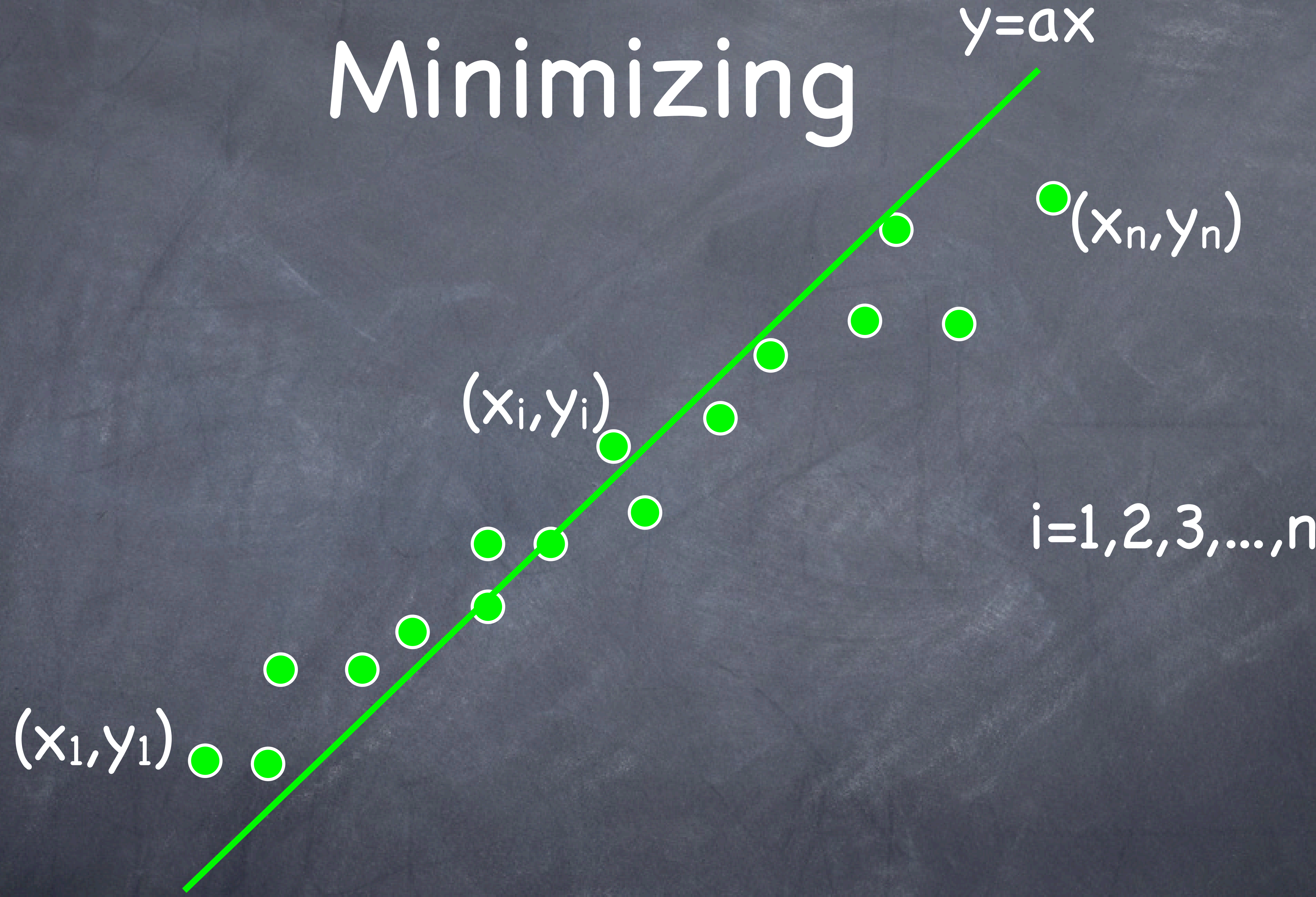
How do we find the best line to fit through the data?

Minimizing

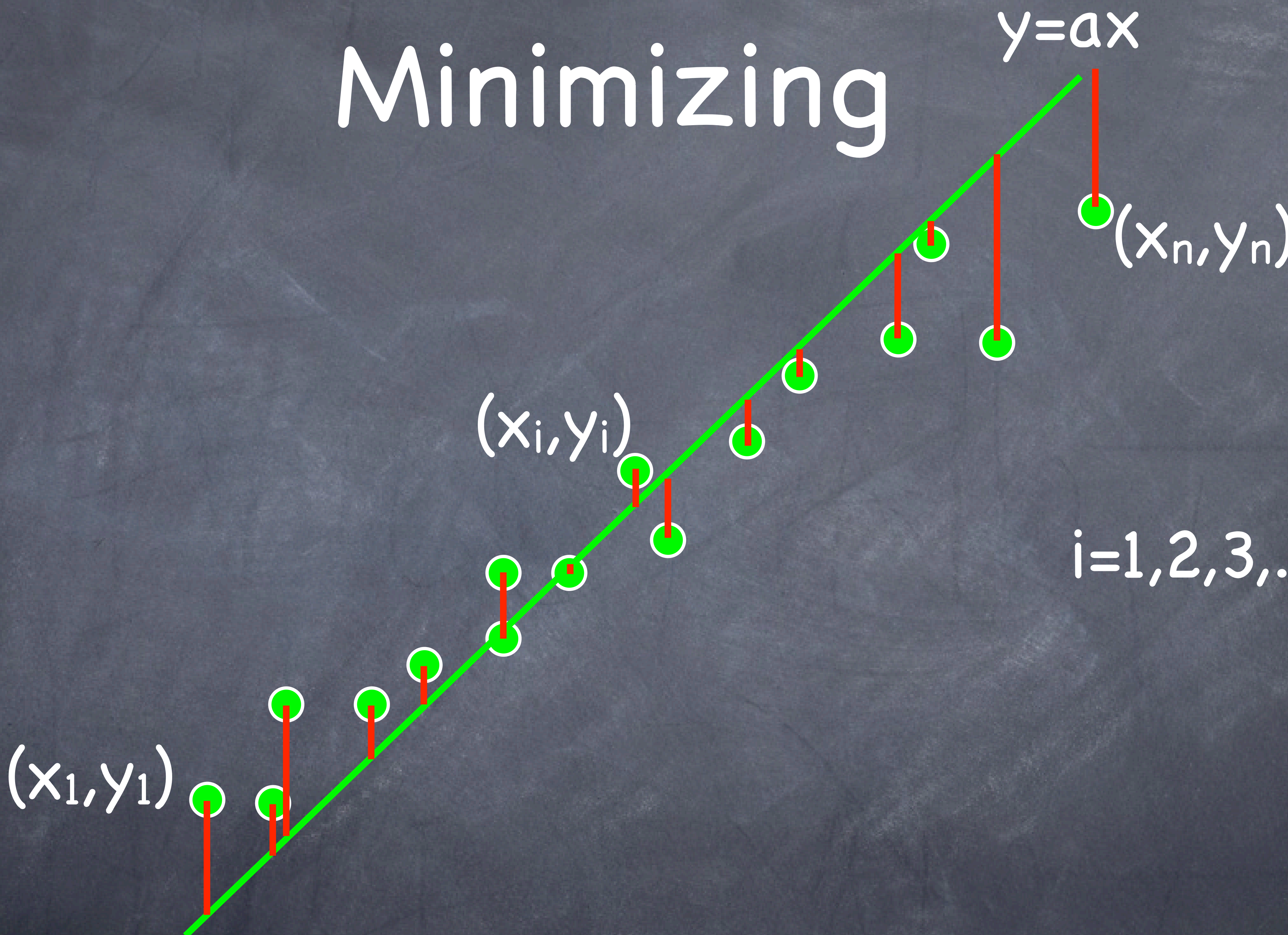


Minimizing

Minimizing

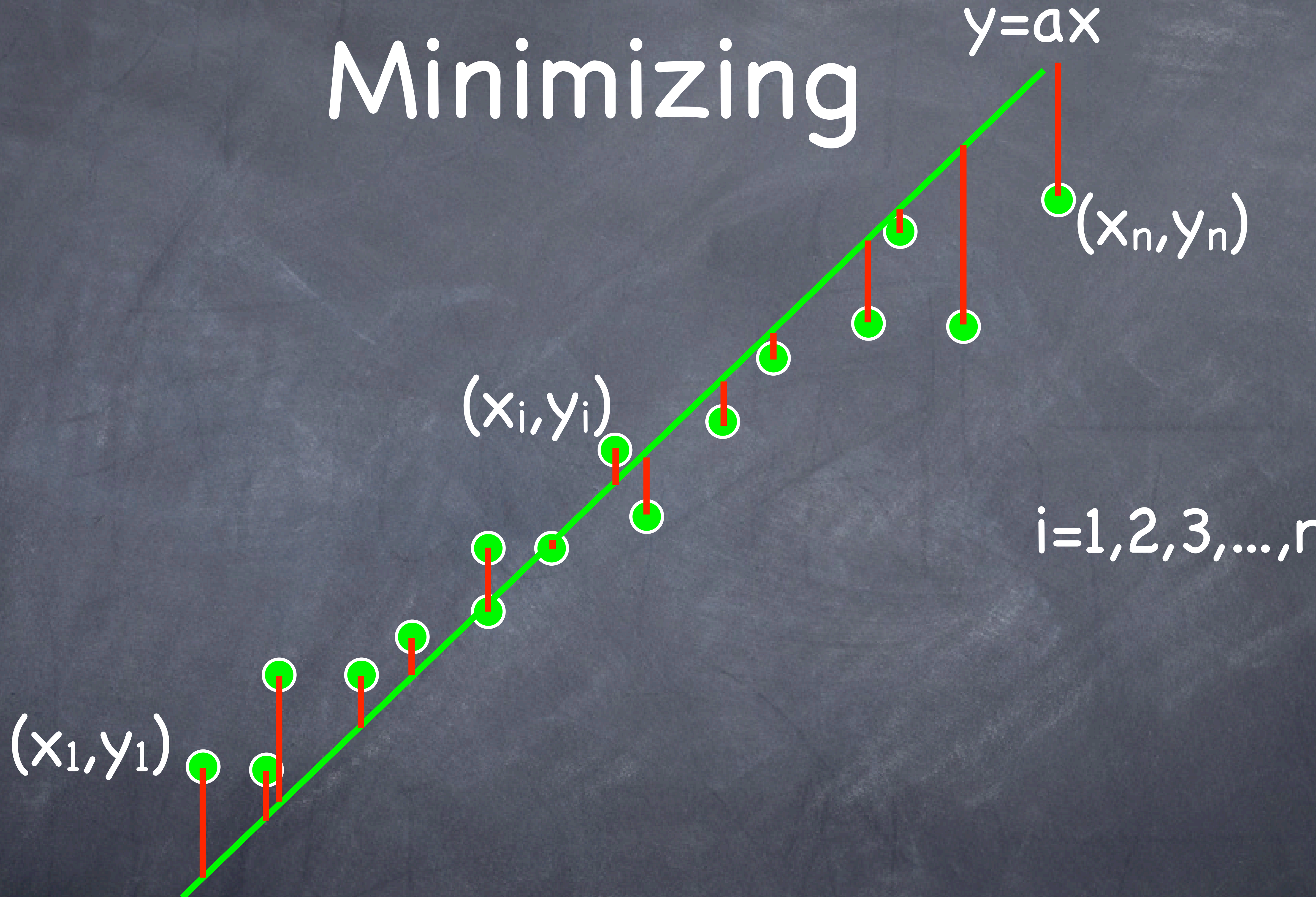


Minimizing



$i = 1, 2, 3, \ldots, n$

Minimizing



Each red bar is called a residual. We want the residuals to be as small as possible.

The residuals are...

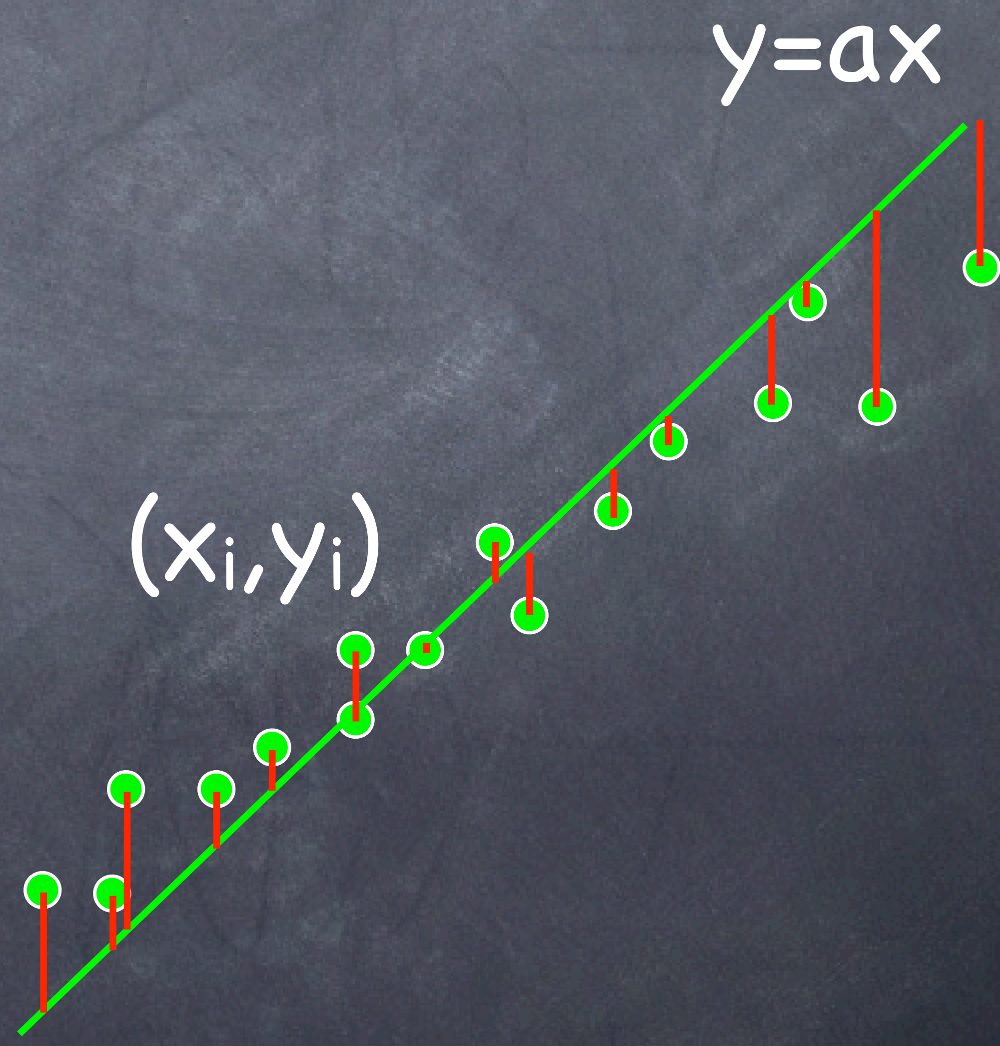
(A) $r_i = y_i^2 + x_i^2$

(B) $r_i = a^2(y_i^2 + x_i^2)$

(C) $r_i = y_i - a\, x_i$

(D) $r_i = y_i - x_i$

(E) $r_i = x_i - y_i$



The residuals are...

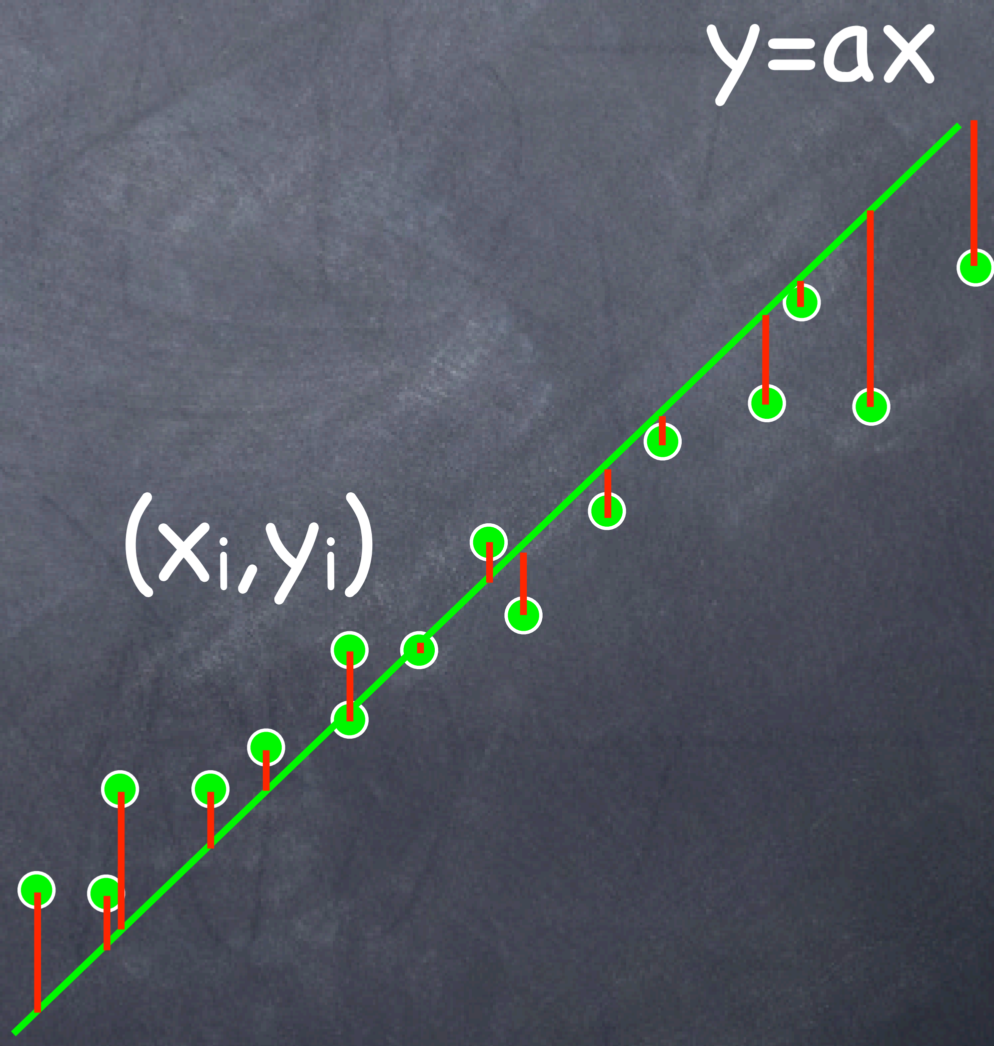
(A) $r_i = y_i^2 + x_i^2$

(B) $r_i = a^2(y_i^2 + x_i^2)$

(C) $r_i = y_i - a\,x_i$

(D) $r_i = y_i - x_i$

(E) $r_i = x_i - y_i$



To minimize the residuals, we define the objective function...

(A) $f(a) = |y_1 - ax_1| + |y_2 - ax_2| + ... + |y_n - ax_n|$

(B) $f(a) = (y_1 - ax_1)^2 + (y_2 - ax_2)^2 + ... + (y_n - ax_n)^2$

(C) $f(a) = (y_1 - ax_1)(y_2 - ax_2)...(y_n - ax_n)$

(D) $f(a) = \max((y_1 - ax_1), (y_2 - ax_2), ..., (y_n - ax_n))$

To minimize the residuals, we define the objective function...

(A) $f(a) = |y_1-ax_1| + |y_2-ax_2| +...+ |y_n-ax_n|$

(B) $f(a) = (y_1-ax_1)^2 + (y_2-ax_2)^2 +...+ (y_n-ax_n)^2$

(C) $f(a) = (y_1-ax_1)(y_2-ax_2)...(y_n-ax_n)$

(D) $f(a) = \max((y_1-ax_1),(y_2-ax_2),...,(y_n-ax_n))$

(B) is called the "sum of squared residuals".
(A) and (D) are reasonable but not as good
(take a stats class to find out more).

Find a so that y=ax fits (4,5), (6,7) in the "least squares" sense.

Define f(a):

(A) $f(a) = |5-4a| + |7-6a|$

(B) $f(a) = (4-5a)^2 + (6-7a)^2$

(C) $f(a) = (5-4a)^2 + (7-6a)^2$

(D) $f(a) = (5-4-a)^2 + (7-6-a)^2$

Find a so that y=ax fits (4,5), (6,7) in the "least squares" sense.

Define f(a):

(A) $f(a) = |5-4a| + |7-6a|$

(B) $f(a) = (4-5a)^2 + (6-7a)^2$

(C) $f(a) = (5-4a)^2 + (7-6a)^2$

(D) $f(a) = (5-4-a)^2 + (7-6-a)^2$

Find a so that y=ax fits (4,5), (6,7) in the "least squares" sense.

Find the a that minimizes f(a):

(A) $a = 7/6$

(B) $a = 5/4$

(C) $a = (7/6 + 5/4) / 2$

(D) $a = 31/26$

Find a so that y=ax fits (4,5), (6,7) in the "least squares" sense.

Find the a that minimizes f(a):

(A) $a = 7/6$

(B) $a = 5/4$

(C) $a = (7/6 + 5/4) / 2$

(D) $a = 31/26 = (4 \cdot 5 + 6 \cdot 7) / (4^2 + 6^2)$

Find a so that y=ax fits (4,5), (6,7) in the "least squares" sense.

Find the a that minimizes f(a):

(A) $a = 7/6$

(B) $a = 5/4$

(C) $a = (7/6 + 5/4) / 2$

(D) $a = 31/26 = (4 \cdot 5 + 6 \cdot 7) / (4^2 + 6^2)$

$= (x_1 \cdot y_1 + x_2 \cdot y_2) / (x_1^2 + x_2^2)$

Find a so that $y=ax$ fits (x_1,y_1), (x_2,y_2),…, (x_n,y_n) in the "least squares" sense.

Define f(a):

(A) $f(a) = |y_1-ax_1| + |y_2-ax_2| + \ldots + |y_n-ax_n|$

(B) $f(a) = (y_1-ax_1)^2 + (y_2-ax_2)^2 + \ldots + (y_n-ax_n)^2$

(C) $f(a) = (ay_1-x_1)^2 + (ay_2-x_2)^2 + \ldots + (ay_n-x_n)^2$

(D) $f(a) = (y_1-a-x_1)^2 + (y_2-a-x_2)^2 + \ldots + (y_n-a-x_n)^2$

Find a so that $y=ax$ fits (x_1,y_1), (x_2,y_2),…, (x_n,y_n) in the "least squares" sense.

Define f(a):

(A) $f(a) = |y_1-ax_1| + |y_2-ax_2| + \ldots + |y_n-ax_n|$

(B) $f(a) = (y_1-ax_1)^2 + (y_2-ax_2)^2 + \ldots + (y_n-ax_n)^2$

(C) $f(a) = (ay_1-x_1)^2 + (ay_2-x_2)^2 + \ldots + (ay_n-x_n)^2$

(D) $f(a) = (y_1-a-x_1)^2 + (y_2-a-x_2)^2 + \ldots + (y_n-a-x_n)^2$

Notation

$$\sum_{i=1}^{n} q_i = q_1 + q_2 + \ldots + q_n$$

Notation

$$\sum_{i=1}^{n} q_i = q_1 + q_2 + \dots + q_n$$

$$\sum_{i=1}^{n} (y_i - ax_i)^2 = (y_1 - ax_1)^2 + (y_2 - ax_2)^2 + \dots + (y_n - ax_n)^2$$

Find a so that y=ax fits (x_1,y_1), (x_2,y_2),...,
(x_n,y_n) in the "least squares" sense.

Find the a that minimizes f(a):

(A) $a = \sum_{i=1}^{n} y_i \,/\, \sum_{i=1}^{n} x_i$

(B) $a = \sum_{i=1}^{n} x_i \,/\, \sum_{i=1}^{n} y_i$

(C) $a = \sum_{i=1}^{n} x_i y_i \,/\, \sum_{i=1}^{n} x_i$

(D) $a = \sum_{i=1}^{n} x_i y_i \,/\, \sum_{i=1}^{n} x_i^2$

Find a so that y=ax fits (x_1,y_1), (x_2,y_2),..., (x_n,y_n) in the "least squares" sense.

Find the a that minimizes f(a):

(A) $a = \sum_{i=1}^{n} y_i \,/\, \sum_{i=1}^{n} x_i$

(C) $a = \sum_{i=1}^{n} x_i y_i \,/\, \sum_{i=1}^{n} x_i$

(B) $a = \sum_{i=1}^{n} x_i \,/\, \sum_{i=1}^{n} y_i$

(D) $a = \sum_{i=1}^{n} x_i y_i \,/\, \sum_{i=1}^{n} x_i^2$

For best fits using y=ax+b, see course notes supplement.

$$a = \frac{P_{avg} - \bar{x}\bar{y}}{X^2_{avg} - \bar{x}^2}$$

$$b = \bar{y} - a\bar{x}$$

$$P_{avg} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i$$

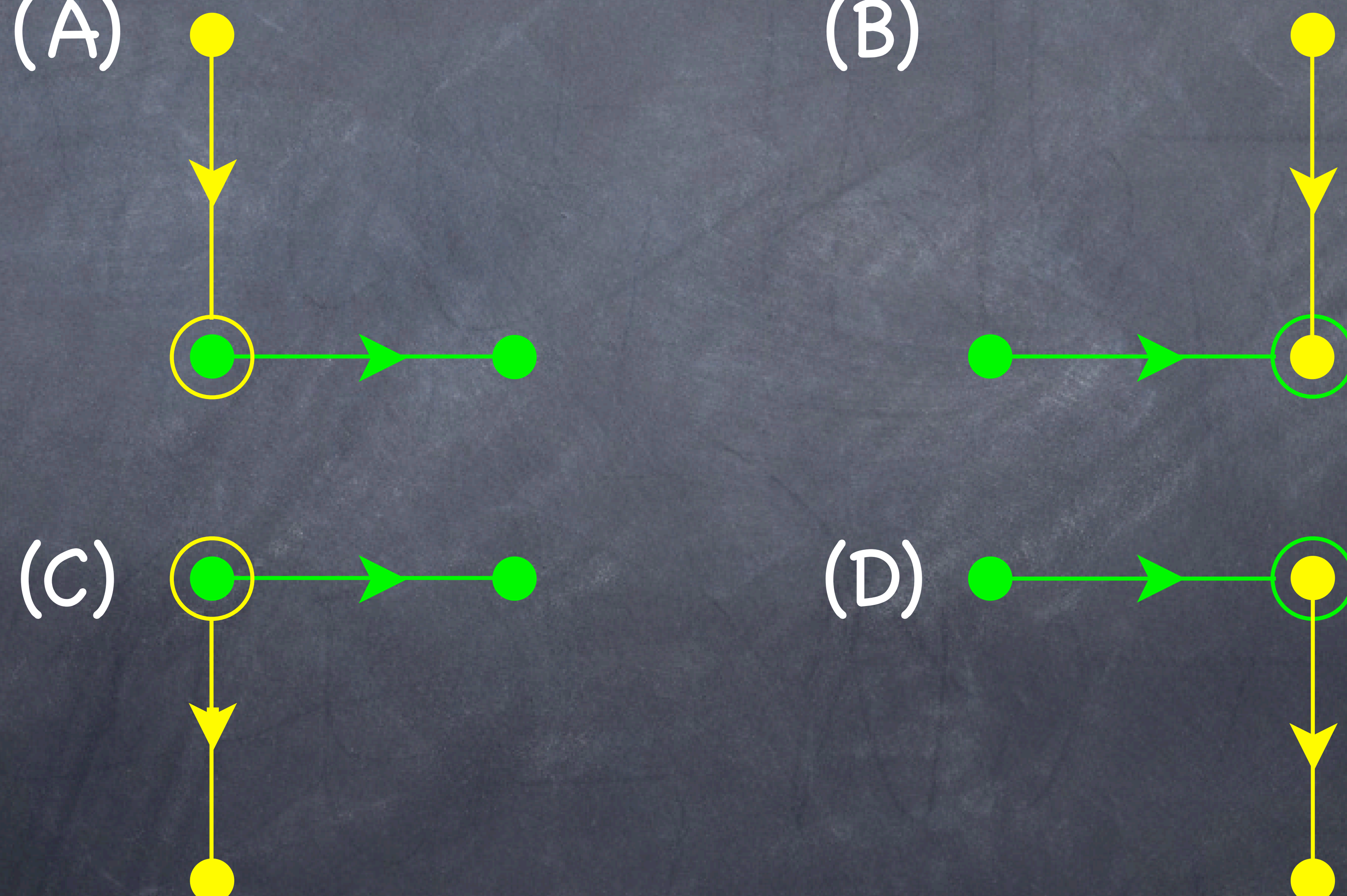
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$X^2_{avg} = \frac{1}{n}\sum_{i=1}^{n} (x_i^2)$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

A boat leaves a dock at 2:00 P.M. and travels due south at a speed of 5 km/h. Another boat has been heading due east at 5 km/h and reaches the same dock at 3:00 P.M. How many minutes past 2:00 P.M. were the boats closest together?

Sketch:



A boat leaves a dock at 2:00 P.M. and travels due south at a speed of 5 km/h. Another boat has been heading due east at 5 km/h and reaches the same dock at 3:00 P.M. How many minutes past 2:00 P.M. were the boats closest together?

Two quantities relevant to solving this problem are:

(A) $x = 5/60\ t$, $y = 5/60\ (60-t)$.

(B) $x = 5(t-2)$, $y=5(3-t)$.

(C) $x = 5-2$, $y=5+3$.

(D) $x = 5t-2$, $y=5t-3$.

A boat leaves a dock at 2:00 P.M. and travels due south at a speed of 5 km/h. Another boat has been heading due east at 5 km/h and reaches the same dock at 3:00 P.M. How many minutes past 2:00 P.M. were the boats closest together?

Objective function to be minimized:

(A) $f(t) = 25|t| + 25|60-t|$

(B) $f(t) = 5/60 \text{ sqrt}(2t^2)$

(C) $f(t) = t^2 + (60-t)^2$

(D) $f(t) = \text{sqrt}(25(t-2)^2 + 25(3-t)^2)$

A boat leaves a dock at 2:00 P.M. and travels due south at a speed of 5 km/h. Another boat has been heading due east at 5 km/h and reaches the same dock at 3:00 P.M. How many minutes past 2:00 P.M. were the boats closest together?

Expectation: The boats will be closest together...

(A) at 2 pm.

(B) at 3 pm.

(C) sometime between 2 pm and 3 pm.

(D) before 2 pm.

(E) after 2 pm.

A boat leaves a dock at 2:00 P.M. and travels due south at a speed of 5 km/h. Another boat has been heading due east at 5 km/h and reaches the same dock at 3:00 P.M. How many minutes past 2:00 P.M. were the boats closest together?

Constraint:

(A) The minimum distance must occur between 2 pm and 3 pm.

(B) $x(t)^2+y(t)^2 = t^2/6$.

(C) $x(t) = 60-y(t)$.

(D) There isn't really a constraint for this problem.

A boat leaves a dock at 2:00 P.M. and travels due south at a speed of 5 km/h. Another boat has been heading due east at 5 km/h and reaches the same dock at 3:00 P.M. How many minutes past 2:00 P.M. were the boats closest together?

Answer (in minutes past 2 pm):

(A) $t = 0$.

(B) $t = 15$.

(C) $t = 30$.

(D) $t = 45$

(E) $t = 60$.