Multiple Regression

General Matrix form for Multiple Regression

Rahul Singh
rsingh@arrsingh.com

What is Simple Linear Regression

re·gres·sion

noun

A statistical method used to **predict** the relationship between a dependent variable and one or more independent variables

y is the dependent variable

What is Simple Linear Regression

re·gres·sion

noun

A statistical method used to **predict** the relationship between a dependent variable and one or more independent variables

in other words…

if we see some data (x, y) we can use linear regression to **predict** the y values for other values of x

y is the dependent variable

What is Simple Linear Regression

re·gres·sion

noun

A statistical method used to **predict** the relationship between a dependent variable and one or more independent variables

$$y = f(x)$$

in other words...

if we see some data (x, y) we can use linear regression to **predict** the y values for other values of x

y is the dependent variable

What is Simple Linear Regression

re·gres·sion

noun

A statistical method used to **predict** the relationship between a dependent variable and one or more independent variables

$$y = f(x)$$

in other words…

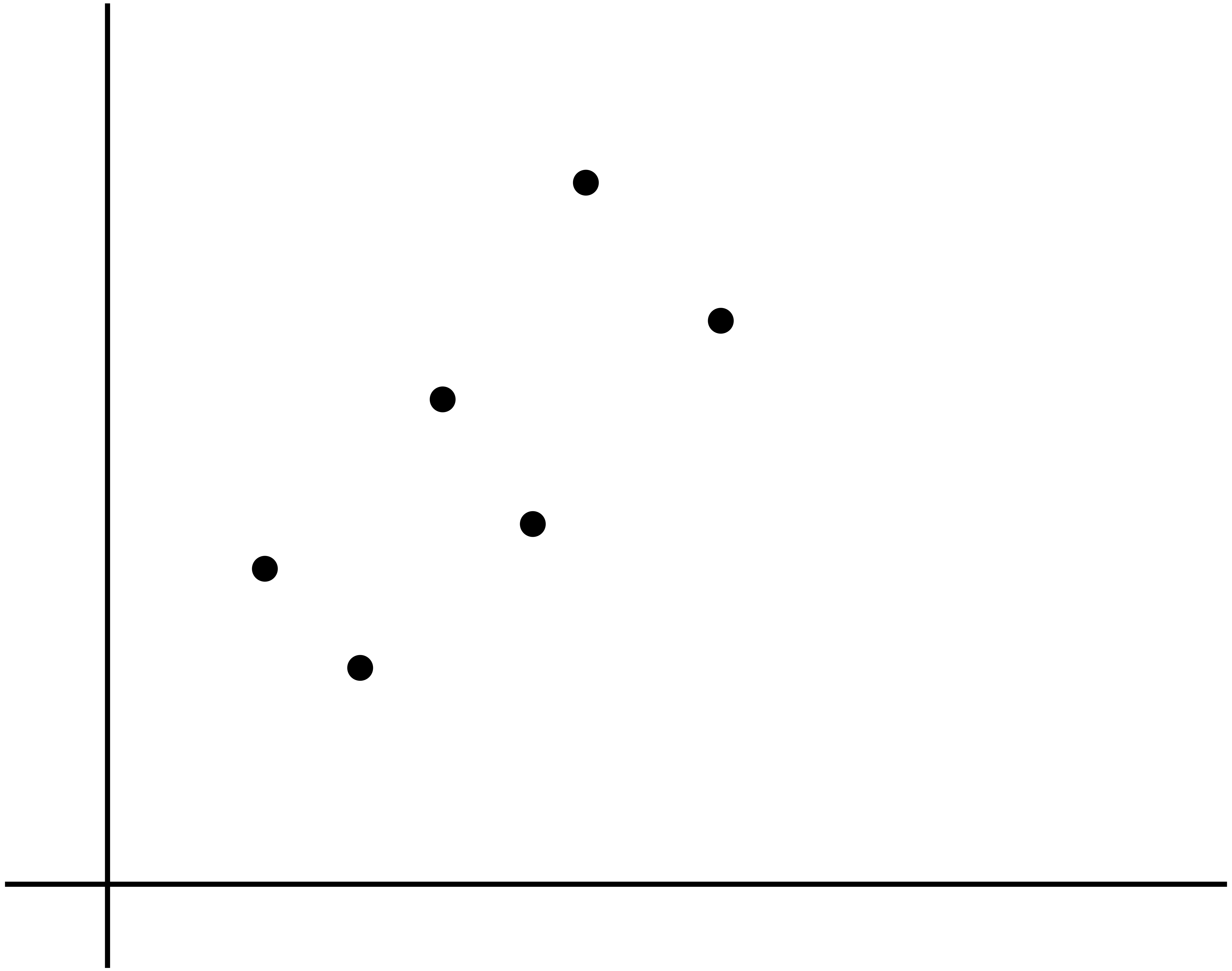
if we see some data (x, y) we can use linear regression to **predict** the y values for other values of x

x is the independent variable

y is the dependent variable

Simple Linear Regression

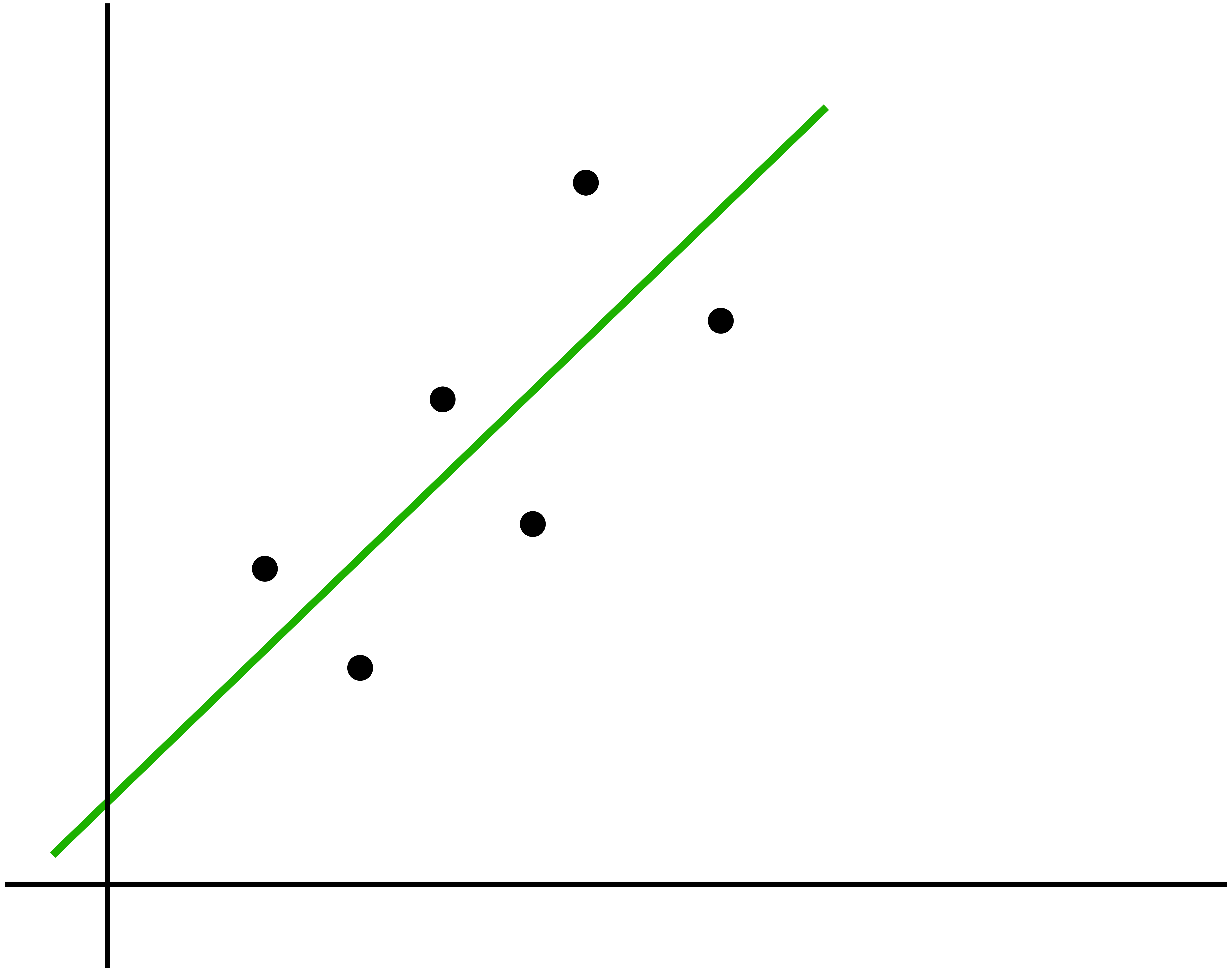
Fundamental Concept: Given a set of data (observations), find the values of β_0 and β_1 for the line that best fits the given data.



Data is synthetic and not plotted to scale

Simple Linear Regression

Fundamental Concept: Given a set of data (observations), find the values of β_0 and β_1 for the line that best fits the given data.

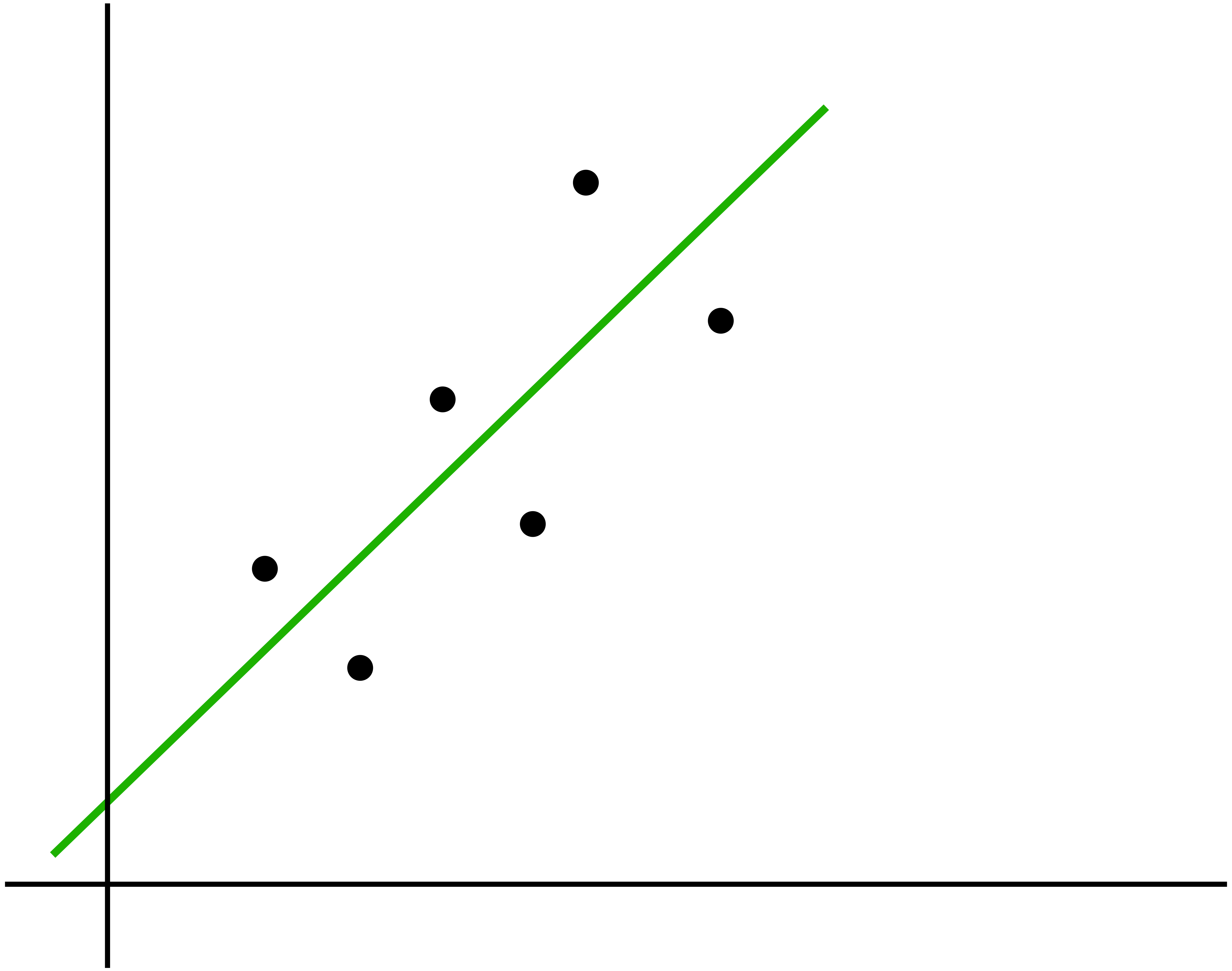


Data is synthetic and not plotted to scale

Simple Linear Regression

Fundamental Concept: Given a set of data (observations), find the values of β_0 and β_1 for the line that best fits the given data.

The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$



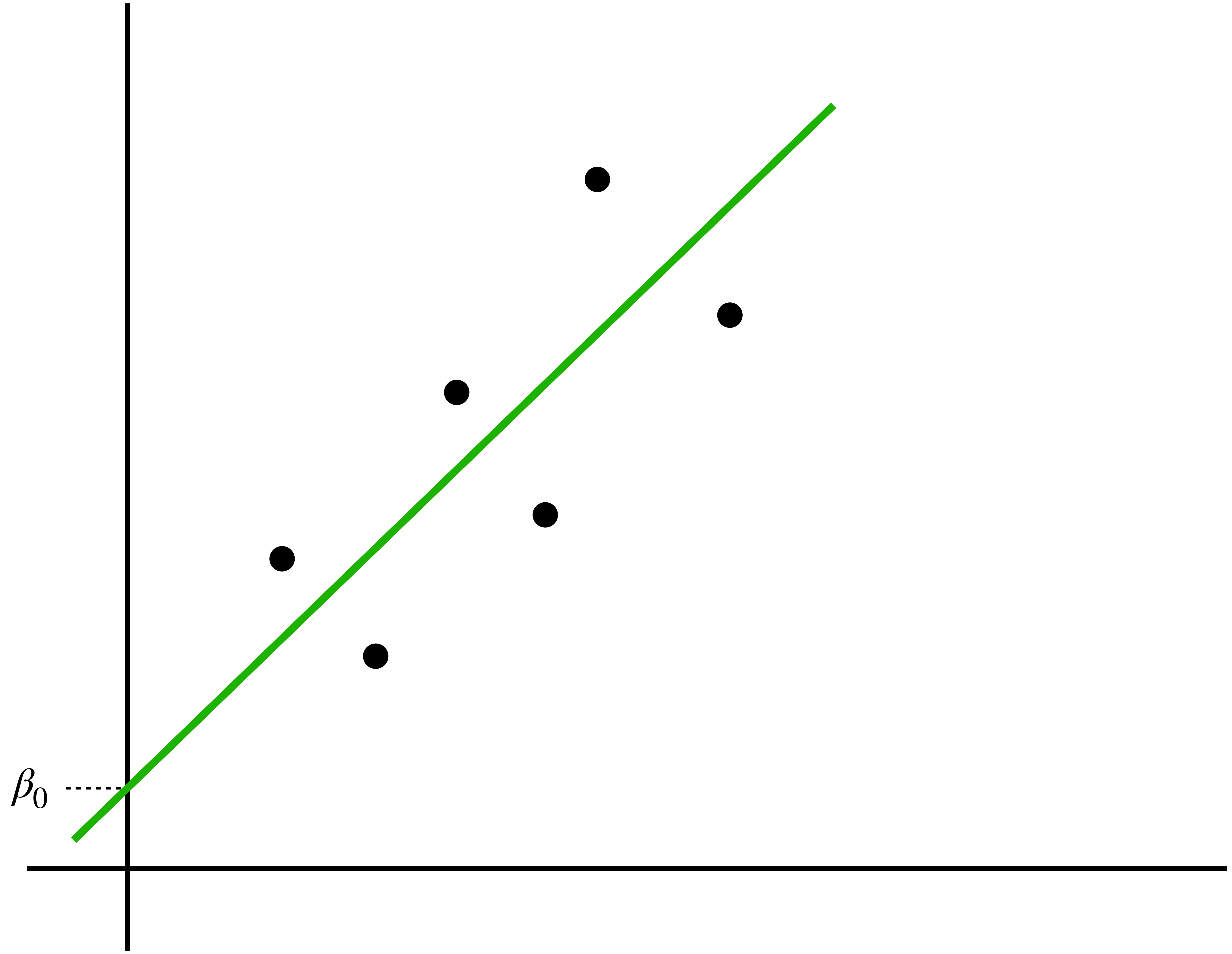
Data is synthetic and not plotted to scale

Simple Linear Regression

Fundamental Concept: Given a set of data (observations), find the values of β_0 and β_1 for the line that best fits the given data.

The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$

β_0 Is the Y intercept



Data is synthetic and not plotted to scale

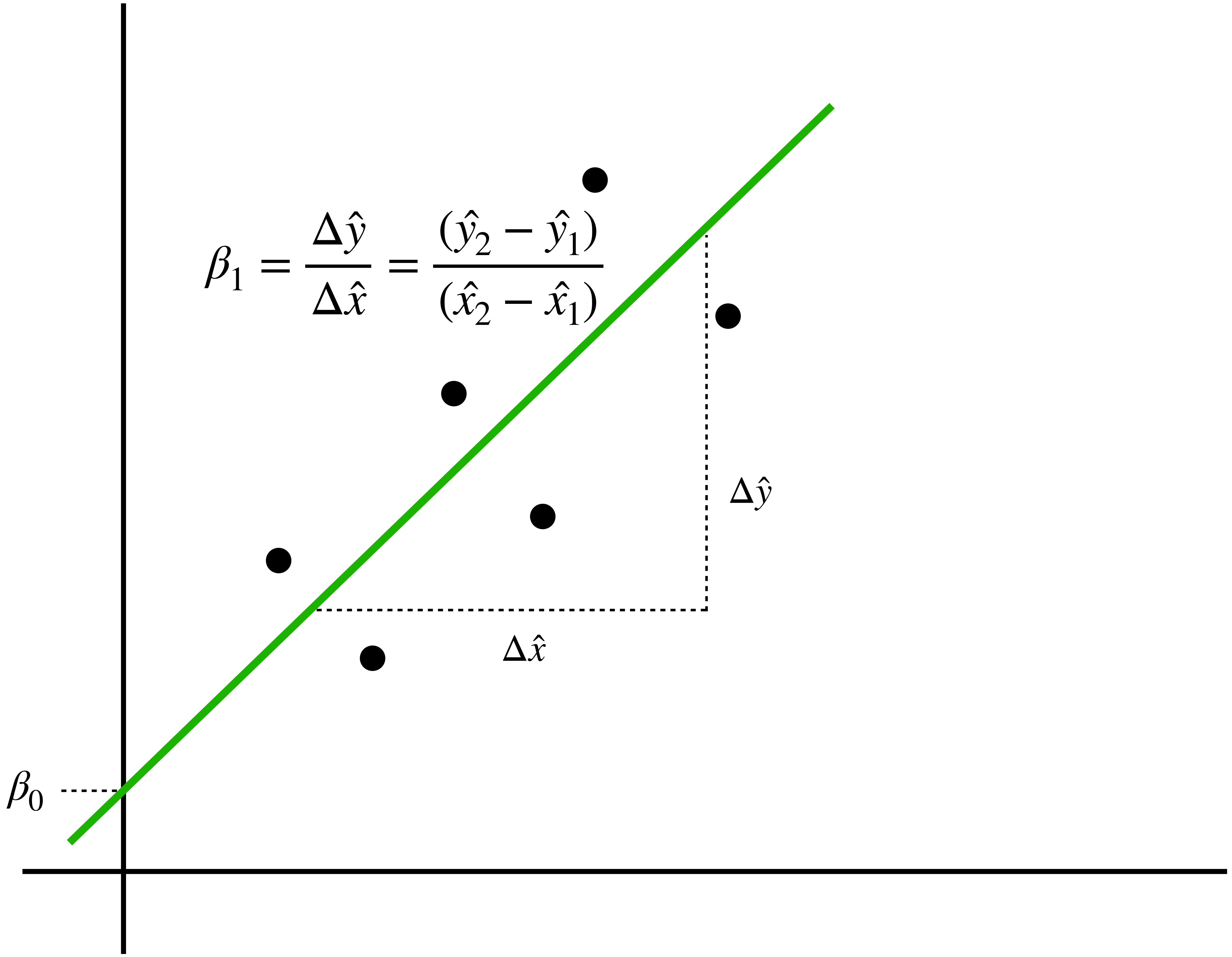
Simple Linear Regression

Fundamental Concept: Given a set of data (observations), find the values of β_0 and β_1 for the line that best fits the given data.

The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$

β_0 Is the Y intercept

β_1 Is the slope of the line



Data is synthetic and not plotted to scale

Simple Linear Regression

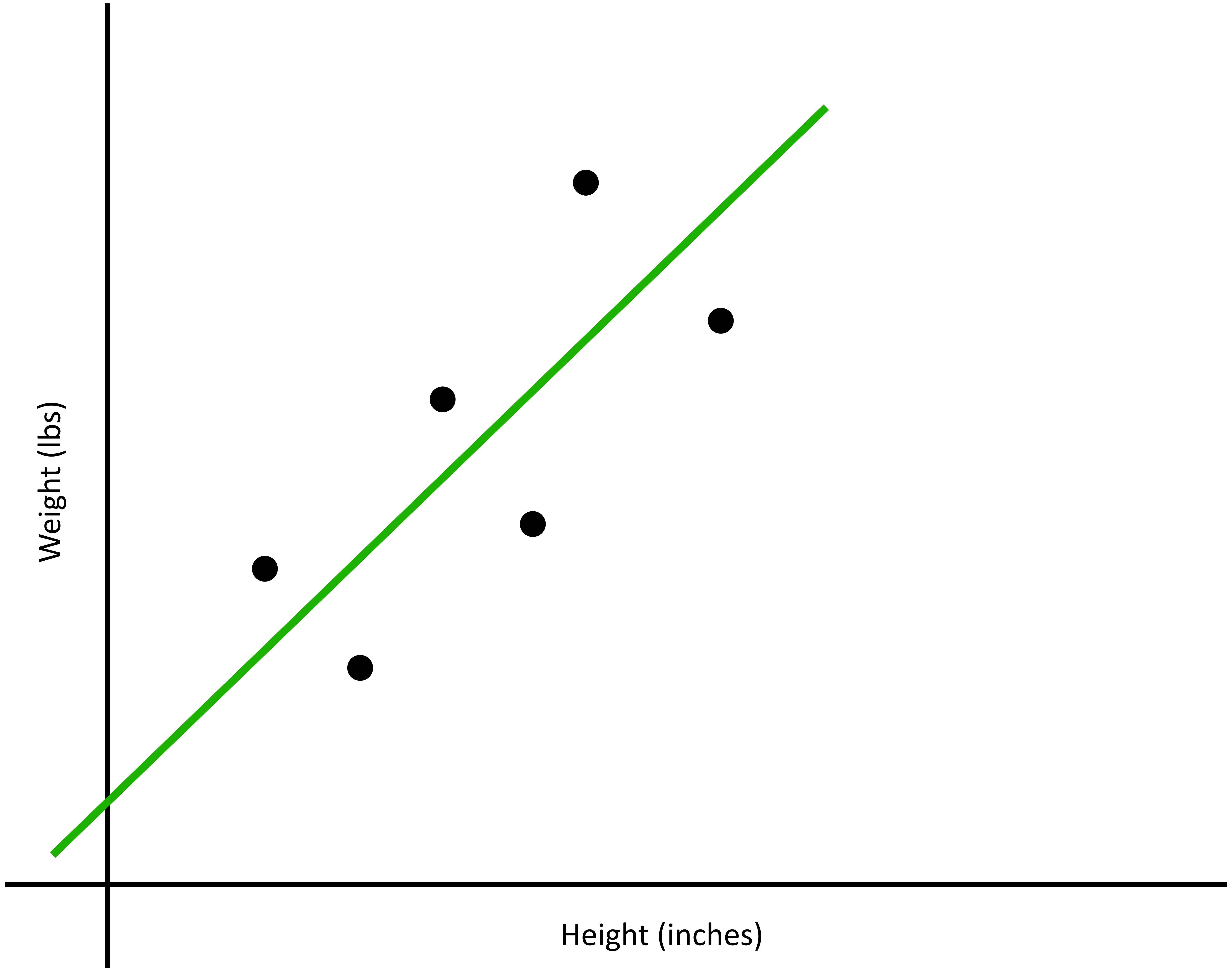
Fundamental Concept: Given a set of data (observations), find the values of β_0 and β_1 for the line that best fits the given data.

$$\hat{y} = \beta_0 + \beta_1 \hat{x}$$

x is the independent variable

y is the dependent variable

Eg: Predict Weight (y) given Height (x)



Data is synthetic and not plotted to scale

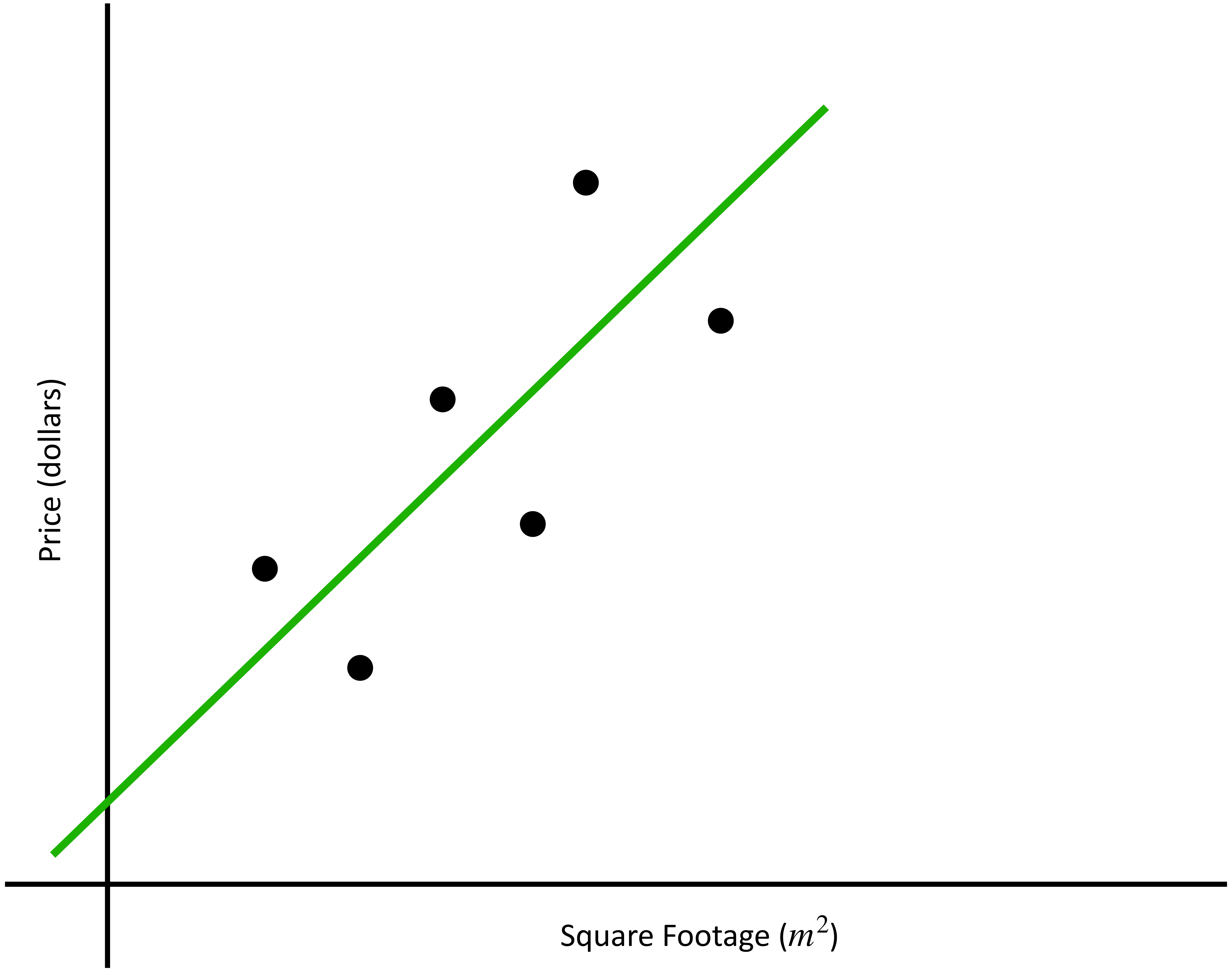
Simple Linear Regression

Fundamental Concept: Given a set of data (observations), find the values of β_0 and β_1 for the line that best fits the given data.

$$\hat{y} = \beta_0 + \beta_1 \hat{x}$$

x represents square footage (m^2)

y represents price (dollars)



Eg: Predict Price of a house (y) given Square Footage (x)

Data is synthetic and not plotted to scale

Simple Linear Regression

Fundamental Concept: Given a set of data (observations), find the values of β_0 and β_1 for the line that best fits the given data.

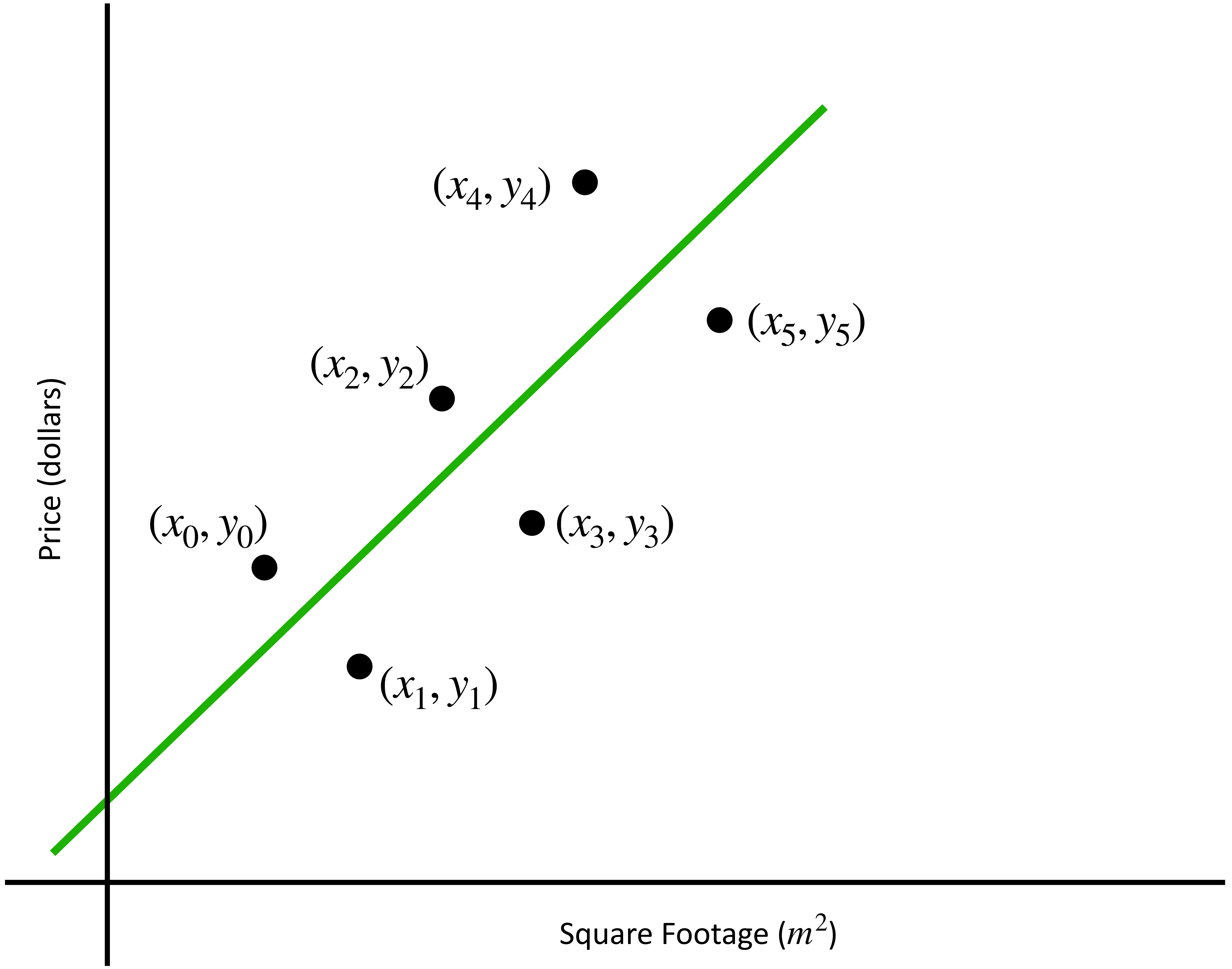
The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$

For each point calculate the squared distance to the line. Divide that by the number of data points.

$$\frac{(y_0 - \hat{y}_0)^2 + (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2}{n}$$

This is the **Mean Squared Error (MSE)**

$$\frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1 \hat{x}_i)^2$$



Data is synthetic and not plotted to scale

Simple Linear Regression

Fundamental Concept: Given a set of data (observations), find the values of β_0 and β_1 for the line that best fits the given data.

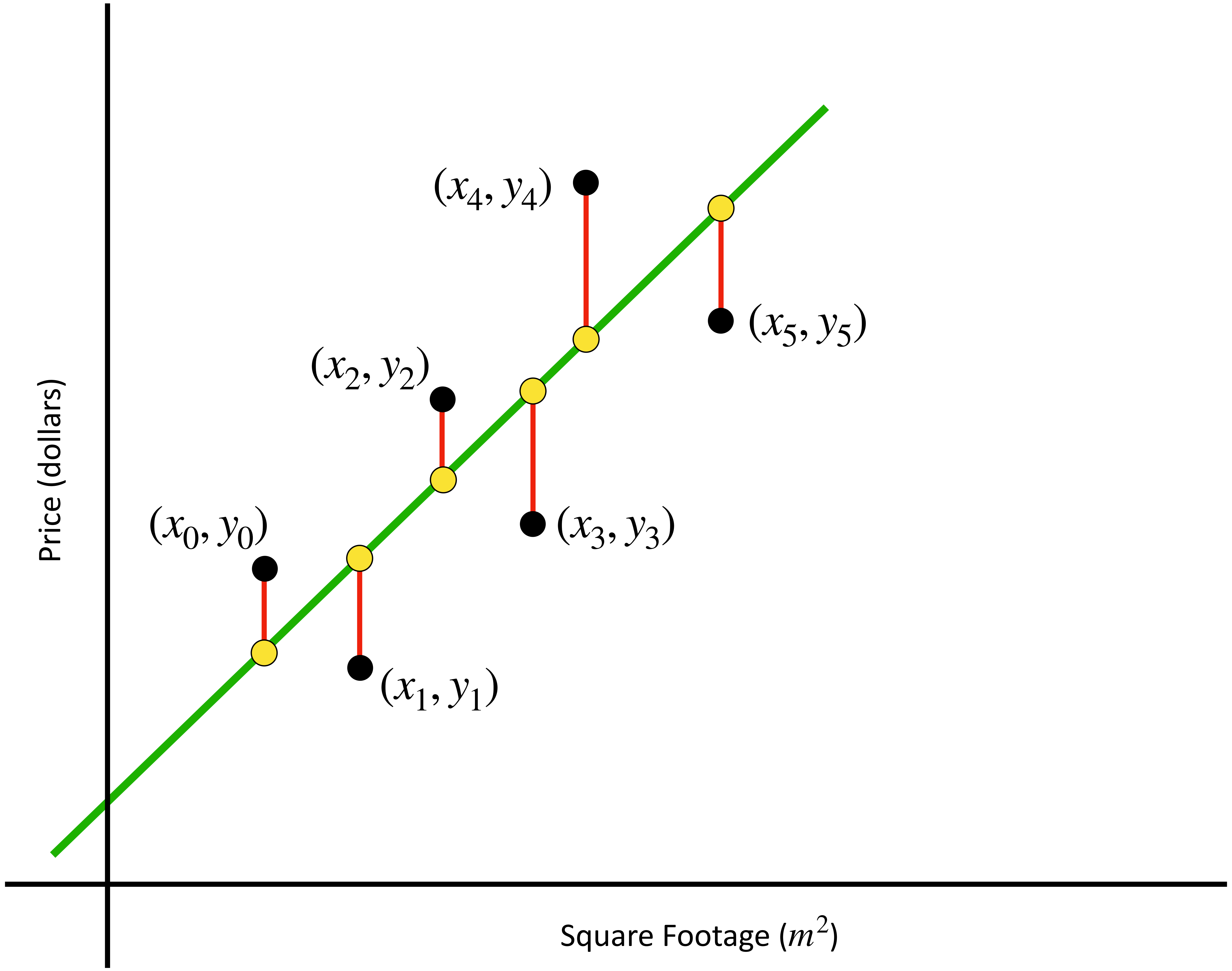
The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$

For each point calculate the squared distance to the line. Divide that by the number of data points.

$$\frac{(y_0 - \hat{y}_0)^2 + (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2}{n}$$

This is the **Mean Squared Error (MSE)**

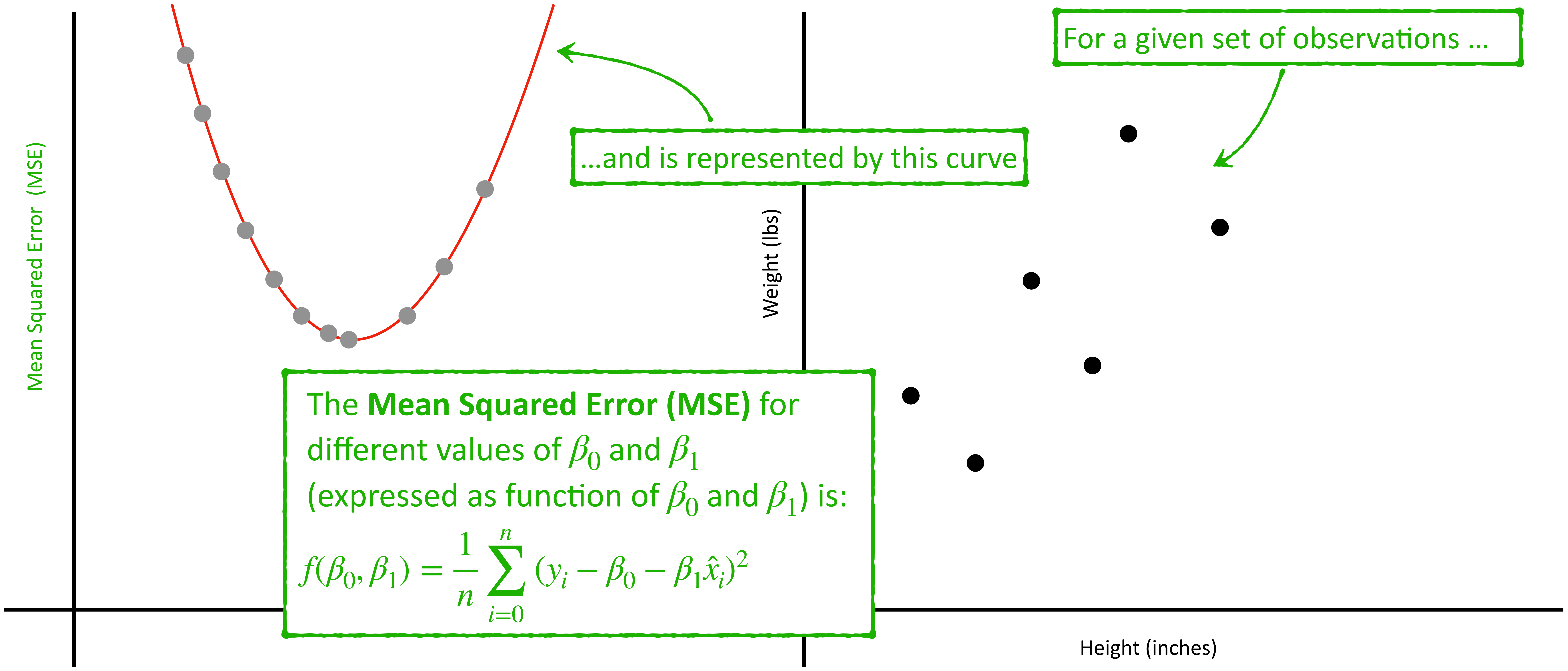
$$\frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1 \hat{x}_i)^2$$



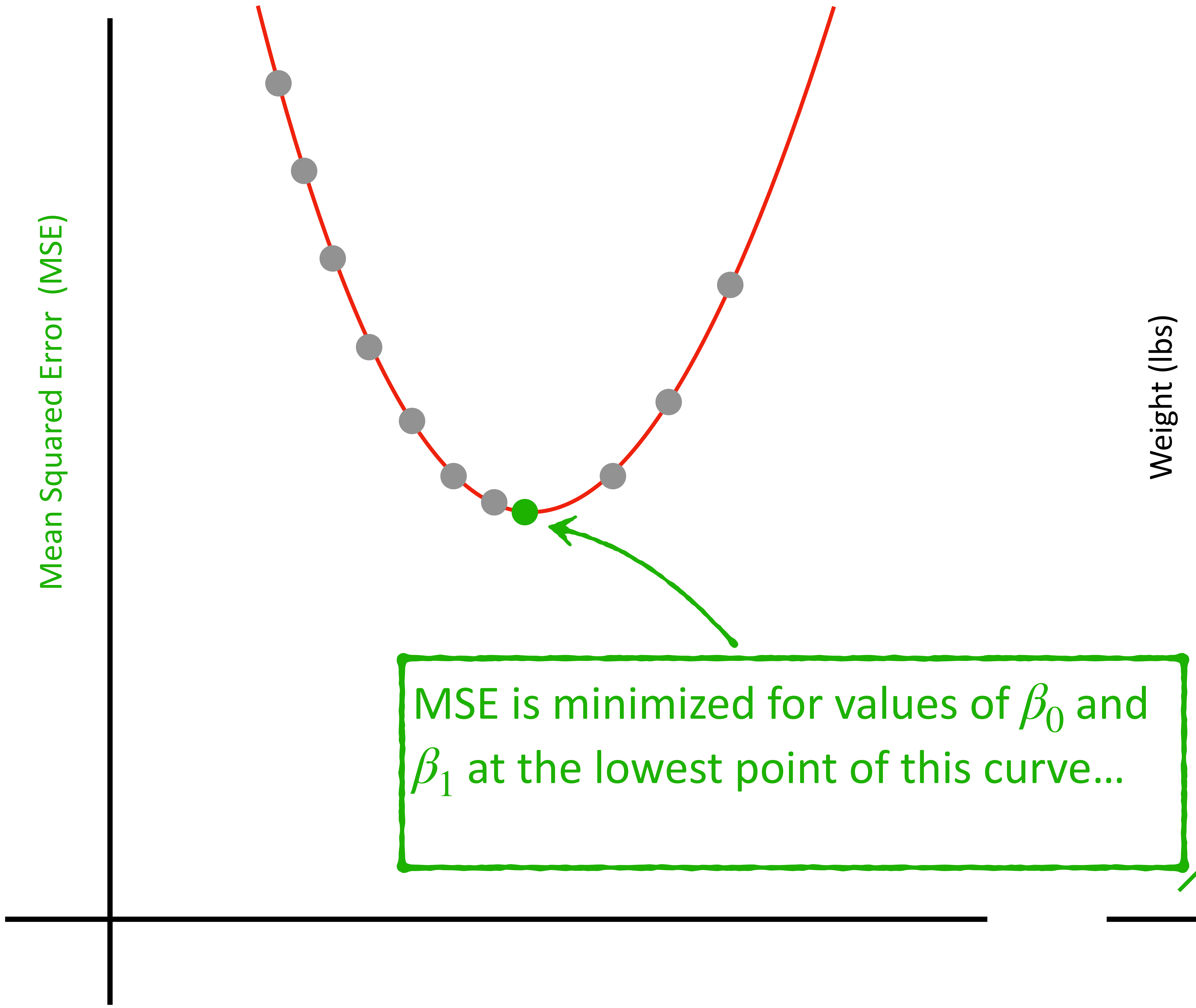
Data is synthetic and not plotted to scale

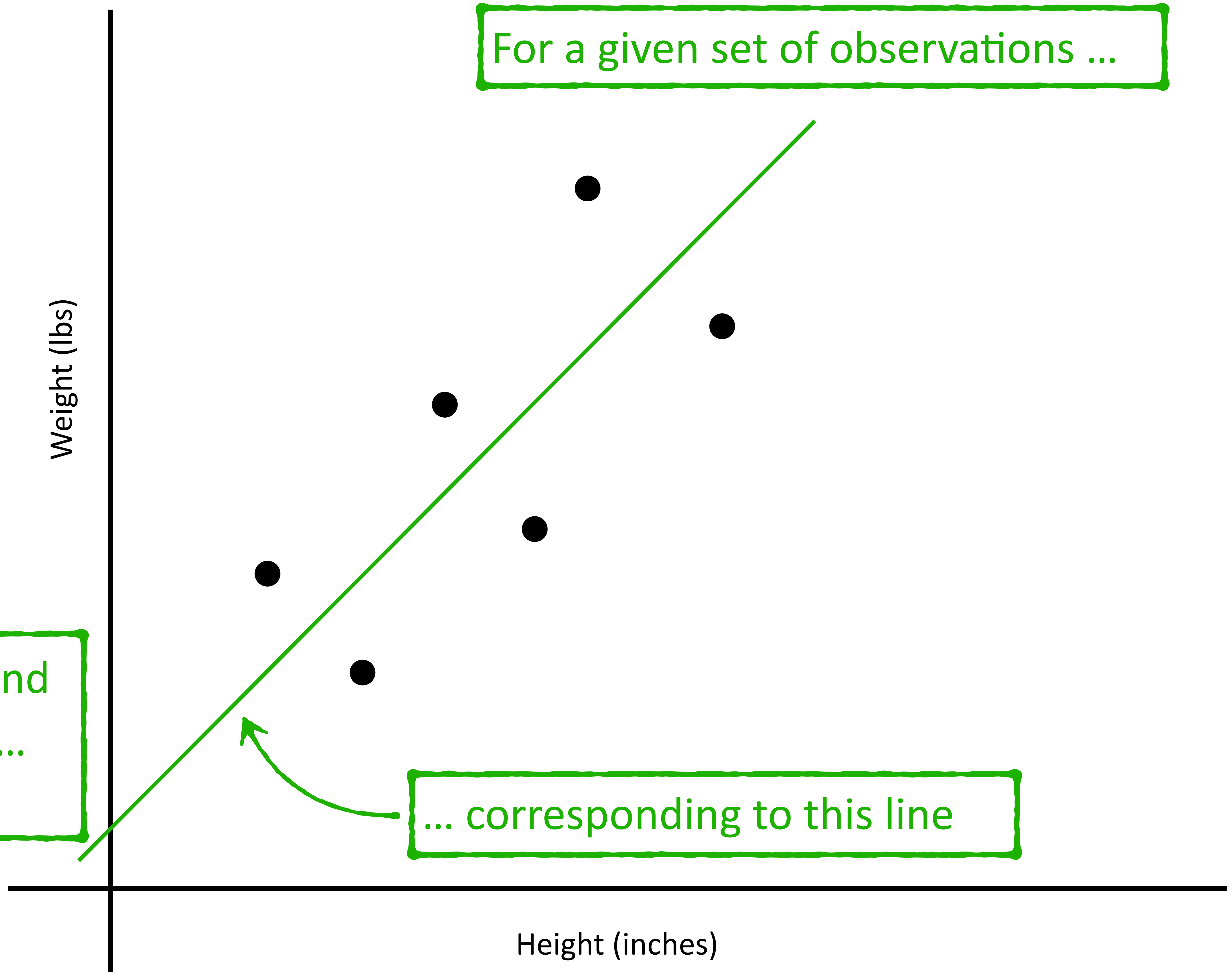
Mean Squared Error (MSE) for various values of β_0 and β_1

Simple Linear Regression



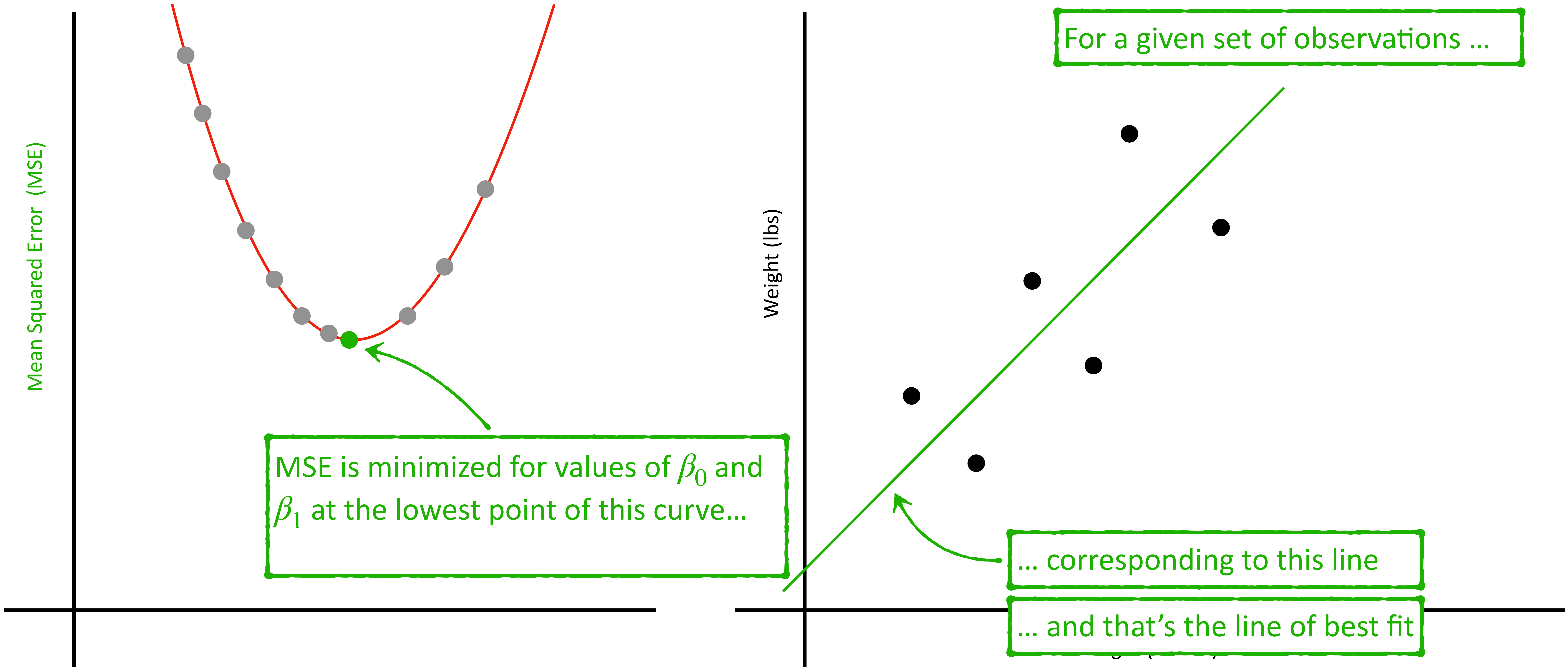
Mean Squared Error (MSE) for various values of β_0 and β_1



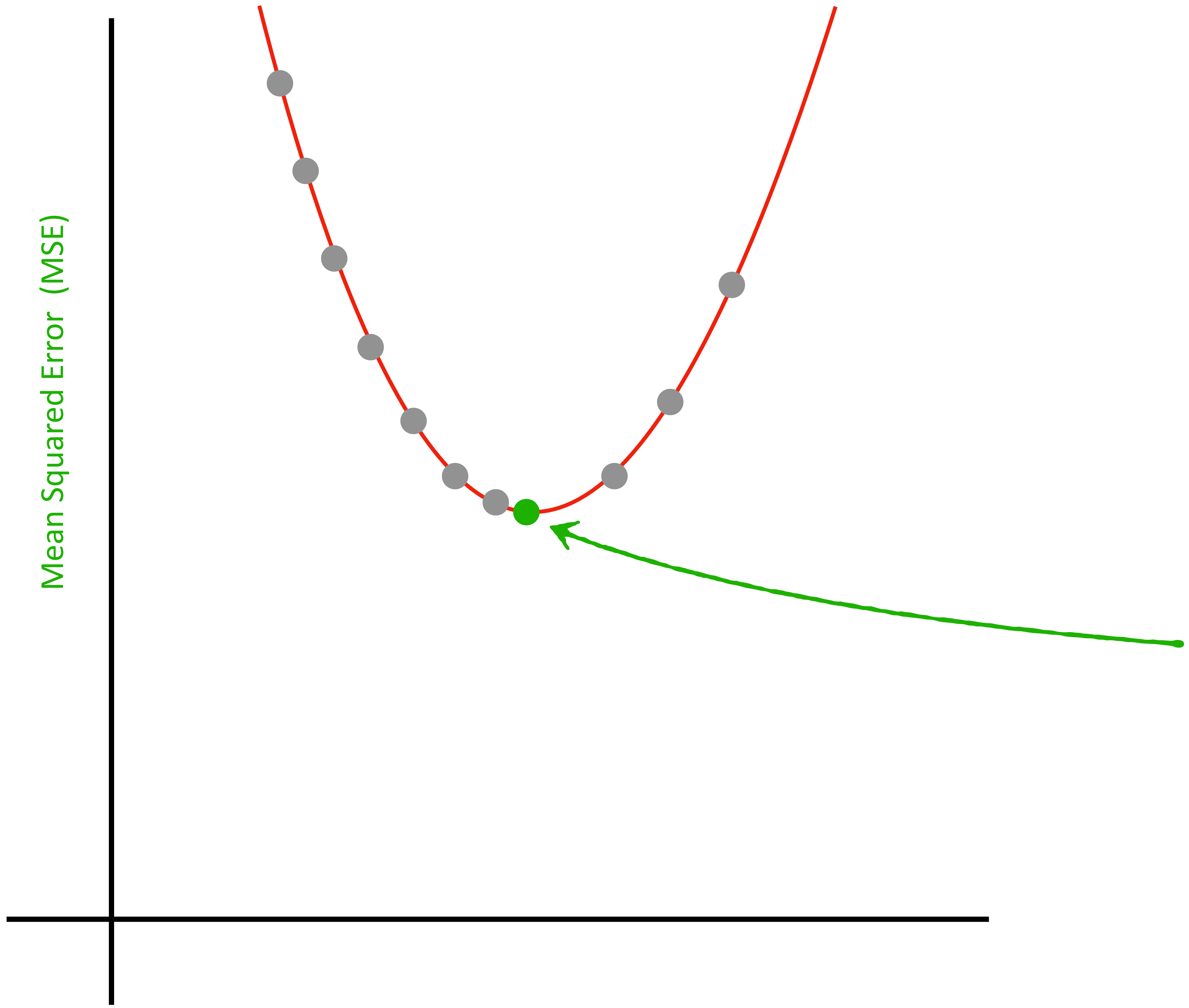


Mean Squared Error (MSE) for various values of β_0 and β_1

Simple Linear Regression



Mean Squared Error (MSE) for various values of β_0 and β_1



MSE is minimized when the first derivative w.r.t β_0 and β_1 equals 0

See Tutorial on Differential Calculus

Mean Squared Error (MSE) for various values of β_0 and β_1

The **Mean Squared Error (MSE)** is

$$\frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1\hat{x}_i)^2$$

Partial Derivatives w.r.t β_0 andβ_1

$$\frac{\partial}{\partial\beta_0}\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = 0 \quad \cdots \quad eq(1)$$

$$\frac{\partial}{\partial\beta_1}\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = 0 \quad \cdots \quad eq(2)$$

MSE is minimized when the first derivative w.r.t β_0 and β_1 equals 0

See Tutorial on Differential Calculus

Mean Squared Error (MSE) for various values of β_0 and β_1

The **Mean Squared Error (MSE)** is

$$\frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1\hat{x}_i)^2$$

Partial Derivatives w.r.t β_0 andβ_1

$$\frac{\partial}{\partial\beta_0}\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = 0 \quad \cdots \quad eq(1)$$

$$\frac{\partial}{\partial\beta_1}\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = 0 \quad \cdots \quad eq(2)$$

MSE is minimized when the first derivative w.r.t β_0 and β_1 equals 0
See Tutorial on Differential Calculus

Solving both equations for β_0 and β_1 we get…

Mean Squared Error (MSE) for various values of β_0 and β_1

Simple Linear Regression

The **Mean Squared Error (MSE)** is

$$\frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1\hat{x}_i)^2$$

Partial Derivatives w.r.t β_0 andβ_1

$$\frac{\partial}{\partial\beta_0}\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = 0 \quad \cdots \; eq(1)$$

$$\frac{\partial}{\partial\beta_1}\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = 0 \quad \cdots \; eq(2)$$

$$\beta_0 = \frac{\sum_{i=0}^{n} y_i - \beta_1 \sum_{i=0}^{n} x_i}{n}$$

$$\beta_1 = \frac{n\sum_{i=0}^{n} x_i y_i - \sum_{i=0}^{n} x_i \sum_{i=0}^{n} y_i}{n\sum_{i=0}^{n} x_i^2 - \left(\sum_{i=0}^{n} x_i\right)^2}$$

This is known as the **Closed Form Solution** for Simple Linear Regression

For the details on how the two equations are solved see
Proof of the Closed Form Solution

Mean Squared Error (MSE) for various values of β_0 and β_1

Simple Linear Regression

The **Mean Squared Error (MSE)** is

$$\frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1\hat{x}_i)^2$$

Partial Derivatives w.r.t β_0 andβ_1

$$\frac{\partial}{\partial\beta_0}\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = 0 \quad \cdots\; eq(1)$$

$$\frac{\partial}{\partial\beta_1}\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = 0 \quad \cdots\; eq(2)$$

Solving both equations for β_0 and β_1 we get…

$$\beta_0 = \frac{\sum_{i=0}^{n} y_i - \beta_1 \sum_{i=0}^{n} x_i}{n}$$

$$\beta_1 = \frac{n\sum_{i=0}^{n} x_i y_i - \sum_{i=0}^{n} x_i \sum_{i=0}^{n} y_i}{n\sum_{i=0}^{n} x_i^2 - \left(\sum_{i=0}^{n} x_i\right)^2}$$

This is known as the **Closed Form Solution** for Simple Linear Regression

For the details on how the two equations are solved see
Proof of the Closed Form Solution

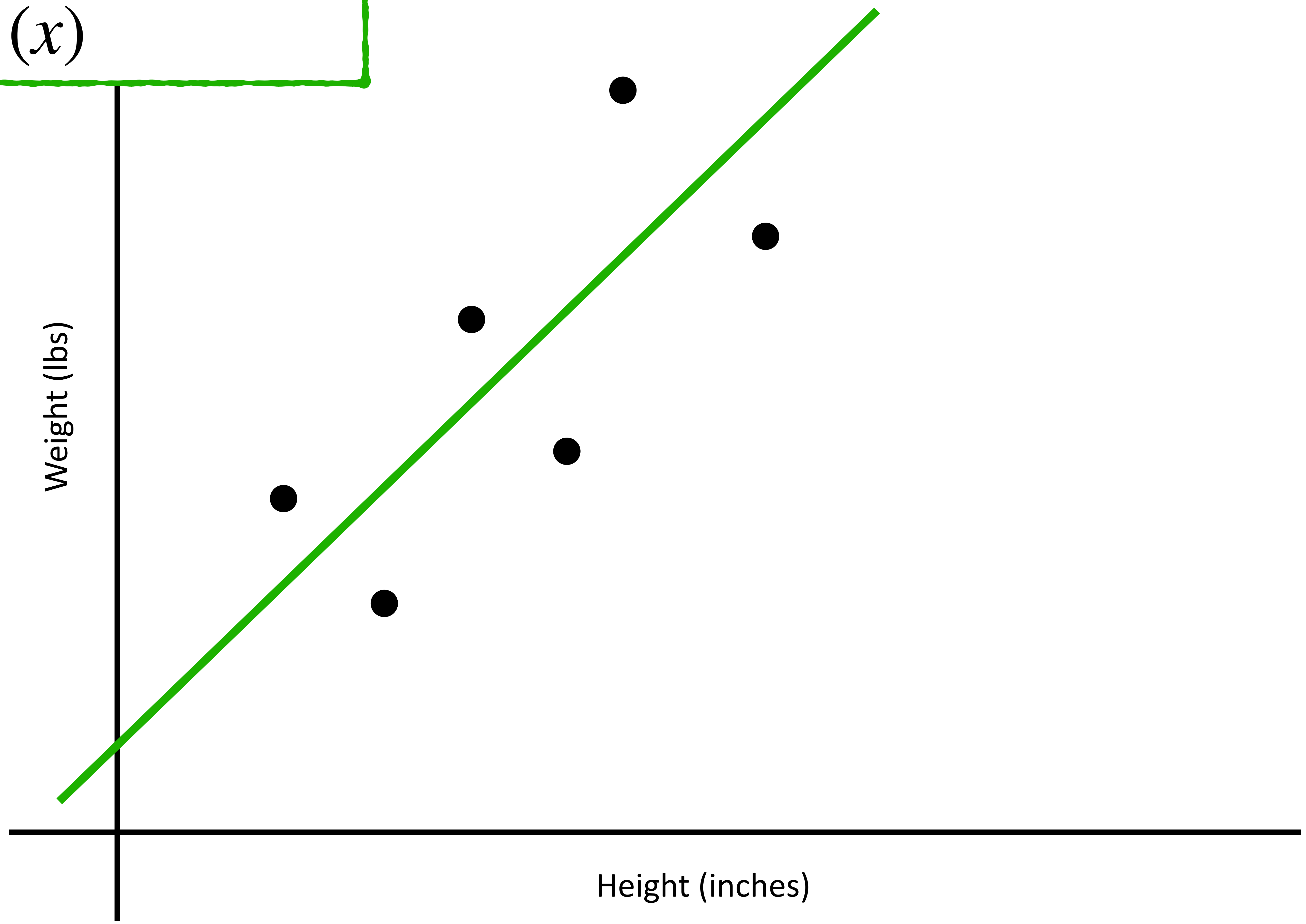
Simple Linear Regression

$$\hat{y} = \beta_0 + \beta_1 \hat{x}$$

1 dependent variable

1 independent variable

Eg: Predict Weight (y) of a person given Height (x)



Data is synthetic and not plotted to scale

Simple Linear Regression

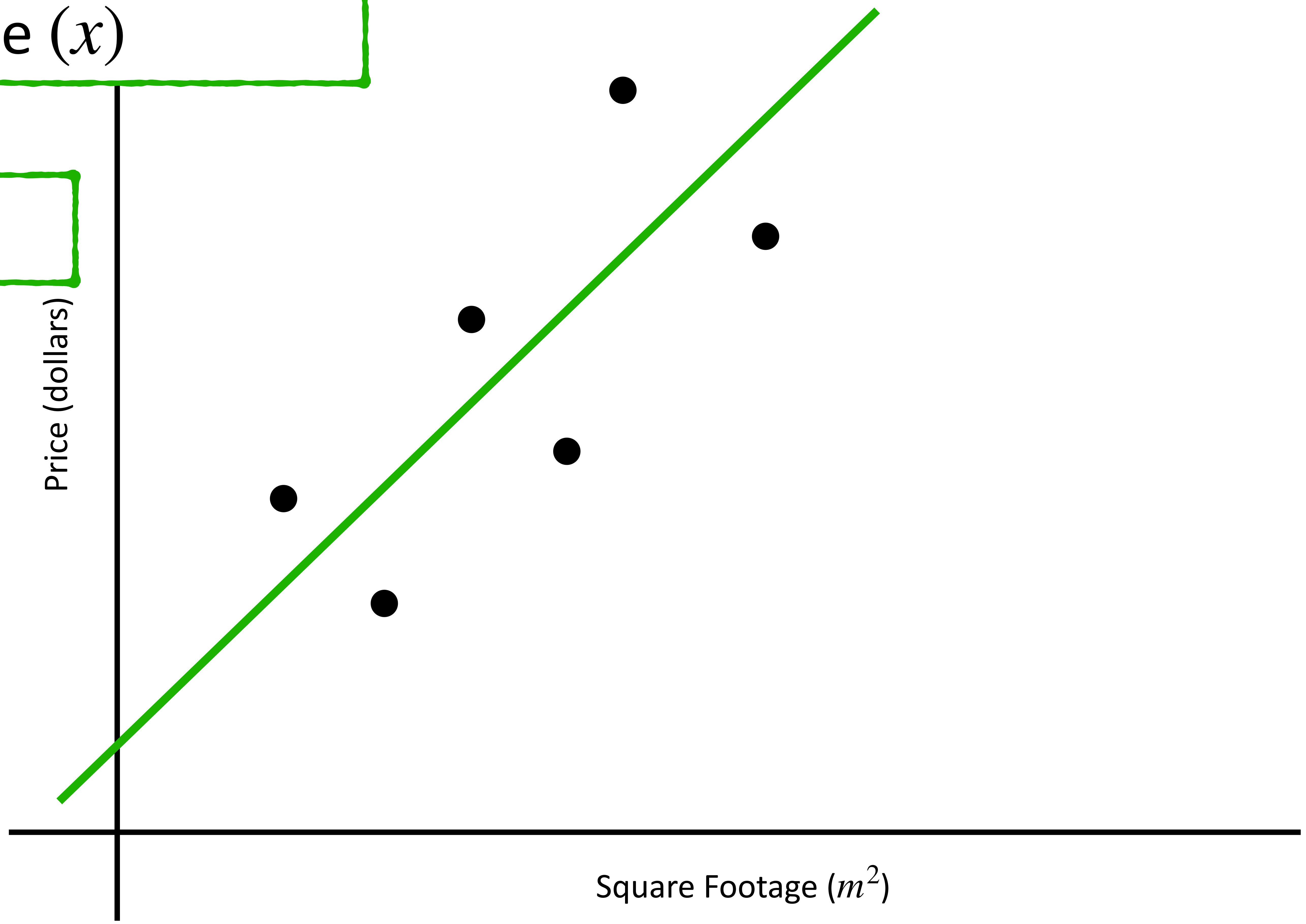
$$\hat{y} = \beta_0 + \beta_1 \hat{x}$$

1 dependent variable

1 independent variable

Eg: Predict Price of a house (y) given Square Footage (x)

2 Parameters - β_0 and β_1



Data is synthetic and not plotted to scale

Simple Linear Regression

$$\hat{y} = \beta_0 + \beta_1 \hat{x}$$

1 dependent variable

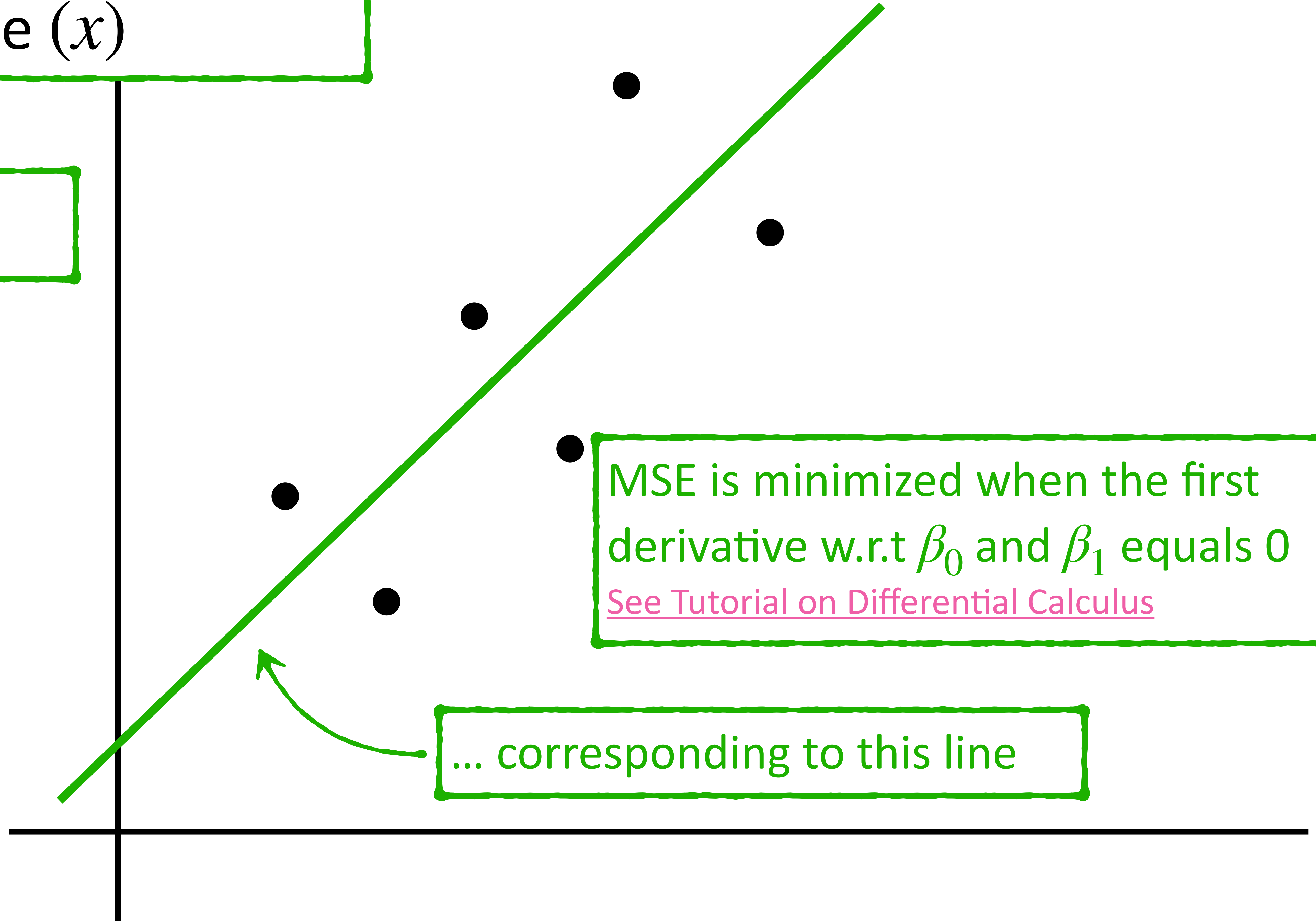
1 independent variable

Eg: Predict Price of a house (y) given Square Footage (x)

2 Parameters - β_0 and β_1

Goal: Find the values of β_0 and β_1 that minimizes the Mean Squared Error (MSE)

$$\frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2$$



MSE is minimized when the first derivative w.r.t β_0 and β_1 equals 0

See Tutorial on Differential Calculus

… corresponding to this line

Data is synthetic and not plotted to scale

Simple Linear Regression

$$\hat{y} = \beta_0 + \beta_1 \hat{x}$$

1 dependent variable

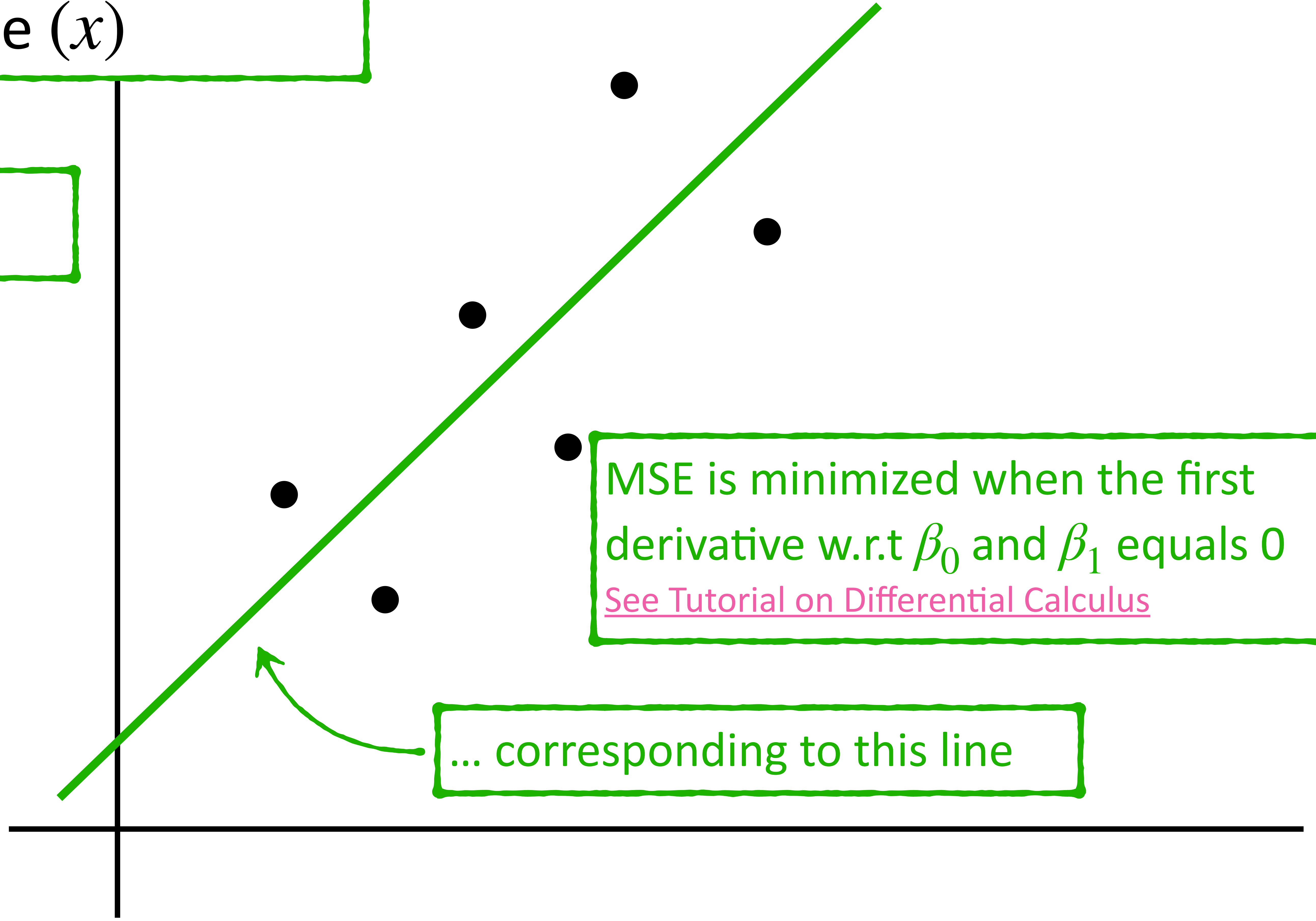
1 independent variable

Eg: Predict Price of a house (y) given Square Footage (x)

2 Parameters - β_0 and β_1

Goal: Find the values of β_0 and β_1 that minimizes the Mean Squared Error (MSE)

$$\frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1 \hat{x}_i)^2$$



MSE is minimized when the first derivative w.r.t β_0 and β_1 equals 0

See Tutorial on Differential Calculus

... corresponding to this line

Data is synthetic and not plotted to scale

Simple Linear Regression

$$\hat{y} = \beta_0 + \beta_1 \hat{x}$$

1 dependent variable

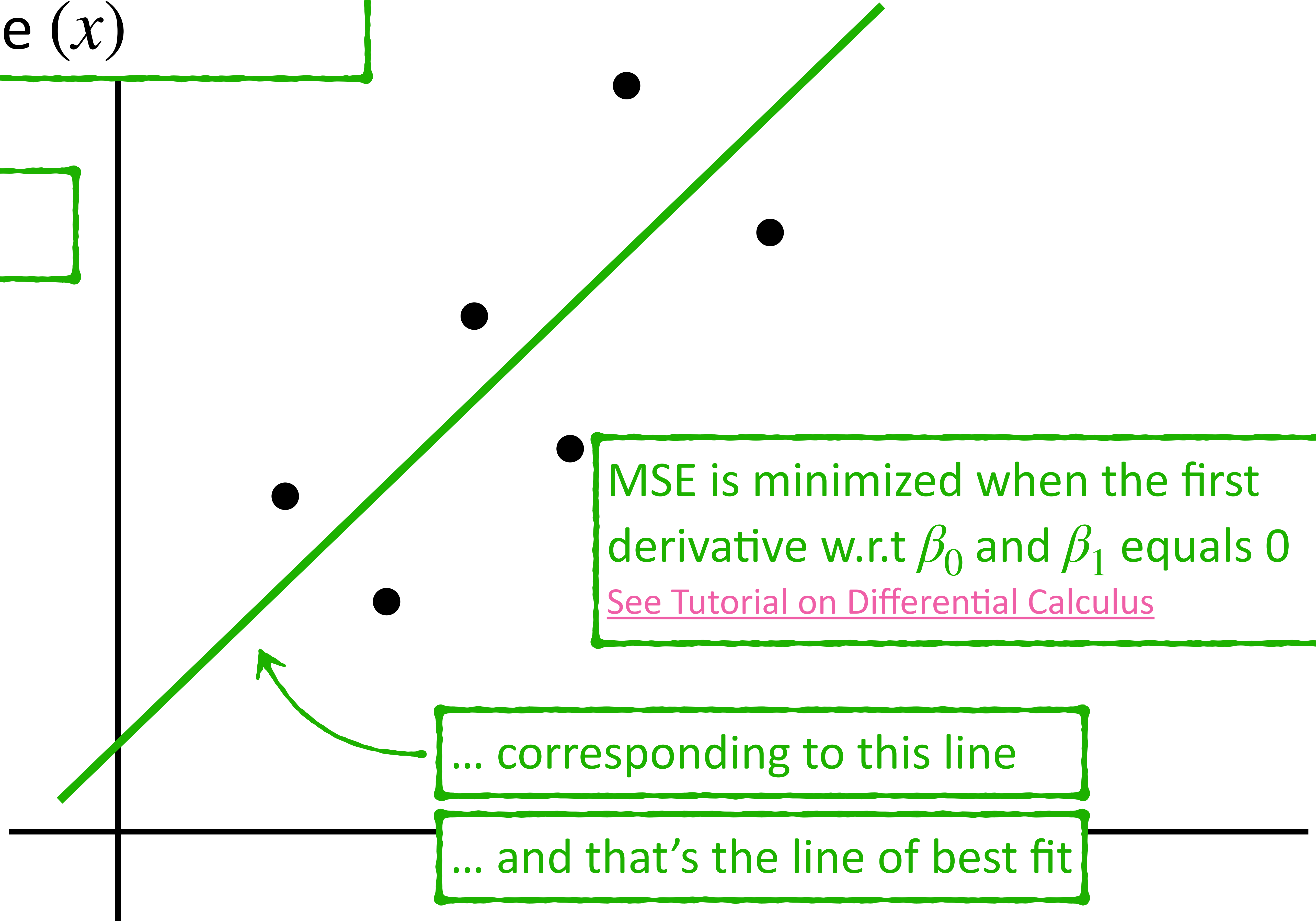
1 independent variable

Eg: Predict Price of a house (y) given Square Footage (x)

2 Parameters - β_0 and β_1

Goal: Find the values of β_0 and β_1 that minimizes the Mean Squared Error (MSE)

$$\frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1 \hat{x}_i)^2$$



MSE is minimized when the first derivative w.r.t β_0 and β_1 equals 0

See Tutorial on Differential Calculus

... corresponding to this line

... and that's the line of best fit

Data is synthetic and not plotted to scale

Simple Linear Regression

1 dependent variable

1 independent variable

$$\hat{y} = \beta_0 + \beta_1 \hat{x}$$

$$\beta_0 = \frac{\sum_{i=0}^{n} y_i - \beta_1 \sum_{i=0}^{n} x_i}{n}$$

$$\beta_1 = \frac{n \sum_{i=0}^{n} x_i y_i - \sum_{i=0}^{n} x_i \sum_{i=0}^{n} y_i}{n \sum_{i=0}^{n} x_i^2 - \left(\sum_{i=0}^{n} x_i \right)^2}$$

We solve for two Parameters - β_0 and β_1 - by solving two equations…

$$\frac{\partial}{\partial \beta_0} \frac{1}{n} \sum_{i=0}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 = 0 \quad \cdots \; eq(1)$$

$$\frac{\partial}{\partial \beta_1} \frac{1}{n} \sum_{i=0}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 = 0 \quad \cdots \; eq(2)$$

MSE is minimized when the first derivative w.r.t β_0 and β_1 equals 0

See Tutorial on Differential Calculus

Data is synthetic and not plotted to scale

Simple Linear Regression

$$\hat{y} = \beta_0 + \beta_1 \hat{x}$$

Predict Price (y) of a house given Square Footage (x)

Simple Linear Regression

- 1 dependent variable y
- 1 independent variable x
- 2 Parameters - β_0 and β_1
- We solve 2 equations to find the values of β_0 and β_1

Data is synthetic and not plotted to scale

Multiple Regression

What if we wanted to predict price of a house, given Square Footage **and** Number of bedrooms?

Data is synthetic and not plotted to scale

Multiple Regression

1 dependent variable y

2 independent variables

$$\hat{y} = \beta_0 + \beta_1 \hat{x}_1 + \beta_2 \hat{x}_2$$

$\hat{x}_1$ represents the square footage

$\hat{x}_2$ represents the number of bedrooms

What if we wanted to predict price of a house, given Square Footage **and** Number of bedrooms?

Data is synthetic and not plotted to scale

Multiple Regression

1 dependent variable y

2 independent variables

$$\hat{y} = \beta_0 + \beta_1 \hat{x}_1 + \beta_2 \hat{x}_2$$

$\hat{x}_1$ represents the square footage

$\hat{x}_2$ represents the number of bedrooms

Goal: Find the values of β_0, β_1 and β_2 that minimizes the Sum of Squared Residuals (SSR)

$$\frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2$$

MSE is minimized when the first derivative of the SSR w.r.t β_0, β_1 and β_2 equals 0

See Tutorial on Differential Calculus

We solve for three Parameters - β_0, β_1 and β_2 - by solving three equations…

$$\frac{\partial}{\partial \beta_0}\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1\hat{x}_{1i} - \beta_2\hat{x}_{2i})^2 = 0$$

$$\frac{\partial}{\partial \beta_1}\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1\hat{x}_{1i} - \beta_2\hat{x}_{2i})^2 = 0$$

$$\frac{\partial}{\partial \beta_2}\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1\hat{x}_{1i} - \beta_2\hat{x}_{2i})^2 = 0$$

Data is synthetic and not plotted to scale

Multiple Regression

1 dependent variable y

2 independent variables

$$\hat{y} = \beta_0 + \beta_1 \hat{x}_1 + \beta_2 \hat{x}_2$$

$\hat{x}_1$ represents the square footage

$\hat{x}_2$ represents the number of bedrooms

Goal: Find the values of β_0, β_1 and β_2 that minimizes the Sum of Squared Residuals (SSR)

$$\frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1\hat{x}_{1i} - \beta_2\hat{x}_{2i})^2$$

MSE is minimized when the first derivative of the SSR w.r.t β_0, β_1 and β_2 equals 0

See Tutorial on Differential Calculus

We solve for three Parameters - β_0, β_1 and β_2 - by solving three equations…

$$\frac{\partial}{\partial \beta_0}\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1\hat{x}_{1i} - \beta_2\hat{x}_{2i})^2 = 0$$

$$\frac{\partial}{\partial \beta_1}\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1\hat{x}_{1i} - \beta_2\hat{x}_{2i})^2 = 0$$

$$\frac{\partial}{\partial \beta_2}\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1\hat{x}_{1i} - \beta_2\hat{x}_{2i})^2 = 0$$

Data is synthetic and not plotted to scale

Lets generalize this…

Multiple Regression

A linear model with...

1 dependent variable y

2 independent variables

$$\hat{y} = \beta_0 + \beta_1 \hat{x}_1 + \beta_2 \hat{x}_2$$

Has 3 parameters

And a cost function...

$$\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1\hat{x}_{1i} - \beta_2\hat{x}_{2i})^2$$

That can be minimized by solving 3 equations

$$\frac{\partial}{\partial \beta_0}\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1\hat{x}_{1i} - \beta_2\hat{x}_{2i})^2 = 0$$

$$\frac{\partial}{\partial \beta_1}\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1\hat{x}_{1i} - \beta_2\hat{x}_{2i})^2 = 0$$

$$\frac{\partial}{\partial \beta_2}\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1\hat{x}_{1i} - \beta_2\hat{x}_{2i})^2 = 0$$

Data is synthetic and not plotted to scale

Multiple Regression

A linear model with...

1 dependent variable y

3 independent variables

$$\hat{y} = \beta_0 + \beta_1 \hat{x}_1 + \beta_2 \hat{x}_2 + \beta_3 \hat{x}_3$$

Has 4 parameters

And a cost function...

$$\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1\hat{x}_{1i} - \beta_2\hat{x}_{2i} - \beta_3\hat{x}_{3i})^2$$

That can be minimized by solving 4 equations

$$\frac{\partial}{\partial\beta_0}\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1\hat{x}_{1i} - \beta_2\hat{x}_{2i} - \beta_3\hat{x}_{3i})^2 = 0$$

$$\frac{\partial}{\partial\beta_1}\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1\hat{x}_{1i} - \beta_2\hat{x}_{2i} - \beta_3\hat{x}_{3i})^2 = 0$$

$$\frac{\partial}{\partial\beta_2}\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1\hat{x}_{1i} - \beta_2\hat{x}_{2i} - \beta_3\hat{x}_{3i})^2 = 0$$

$$\frac{\partial}{\partial\beta_3}\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1\hat{x}_{1i} - \beta_2\hat{x}_{2i} - \beta_3\hat{x}_{3i})^2 = 0$$

Data is synthetic and not plotted to scale

Multiple Regression

A linear model with…

1 dependent variable y

k independent variables $x_1 \,..\, x_k$

$$\hat{y} = \beta_0 + \beta_1\hat{x}_1 + \beta_2\hat{x}_2 + \beta_3\hat{x}_3 + \ldots + \beta_k\hat{x}_k$$

Has $k+1$ parameters

And a cost function…

$$\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1\hat{x}_{1i} - \beta_2\hat{x}_{2i} - \beta_3\hat{x}_{3i} - \ldots - \beta_k\hat{x}_{ki})^2$$

That can be minimized by solving $k+1$ equations

$$\frac{\partial}{\partial\beta_0}\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1\hat{x}_{1i} - \beta_2\hat{x}_{2i} - \beta_3\hat{x}_{3i} - \ldots - \beta_k\hat{x}_{ki})^2 = 0$$

$$\frac{\partial}{\partial\beta_1}\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1\hat{x}_{1i} - \beta_2\hat{x}_{2i} - \beta_3\hat{x}_{3i} - \ldots - \beta_k\hat{x}_{ki})^2 = 0$$

$$\frac{\partial}{\partial\beta_2}\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1\hat{x}_{1i} - \beta_2\hat{x}_{2i} - \beta_3\hat{x}_{3i} - \ldots - \beta_k\hat{x}_{ki})^2 = 0$$

$$\vdots$$

$$\frac{\partial}{\partial\beta_n}\frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1\hat{x}_{1i} - \beta_2\hat{x}_{2i} - \beta_3\hat{x}_{3i} - \ldots - \beta_k\hat{x}_{ki})^2 = 0$$

k independent variables $x_1 \ldots x_k$

$$\hat{y} = \beta_0 + \beta_1\hat{x}_1 + \beta_2\hat{x}_2 + \beta_3\hat{x}_3 + \ldots + \beta_k\hat{x}_k$$

Has $k+1$ p

And a cost f

$$\sum_{i=0}^{n} (y_i - \beta_0 - \beta_1\hat{x}_{1i} - \beta_2\hat{x}_{2i} - \beta_3\hat{x}_{3i} - \ldots - \beta_k\hat{x}_{ki})^2$$

That can be minimized by solving $k+1$ equations

$$- \beta_k\hat{x}_{ki})^2 = 0$$

$$- \beta_k\hat{x}_{ki})^2 = 0$$

$$- \beta_k\hat{x}_{ki})^2 = 0$$

$$\vdots$$

$$\frac{\partial}{\partial \beta_n} \sum_{i=0}^{n} (y_i - \beta_0 - \beta_1\hat{x}_{1i} - \beta_2\hat{x}_{2i} - \beta_3\hat{x}_{3i} - \ldots - \beta_k\hat{x}_{ki})^2 = 0$$

Solving $k+1$ linear equations isn't practical

Lets use a Matrix

Simple Linear Regression

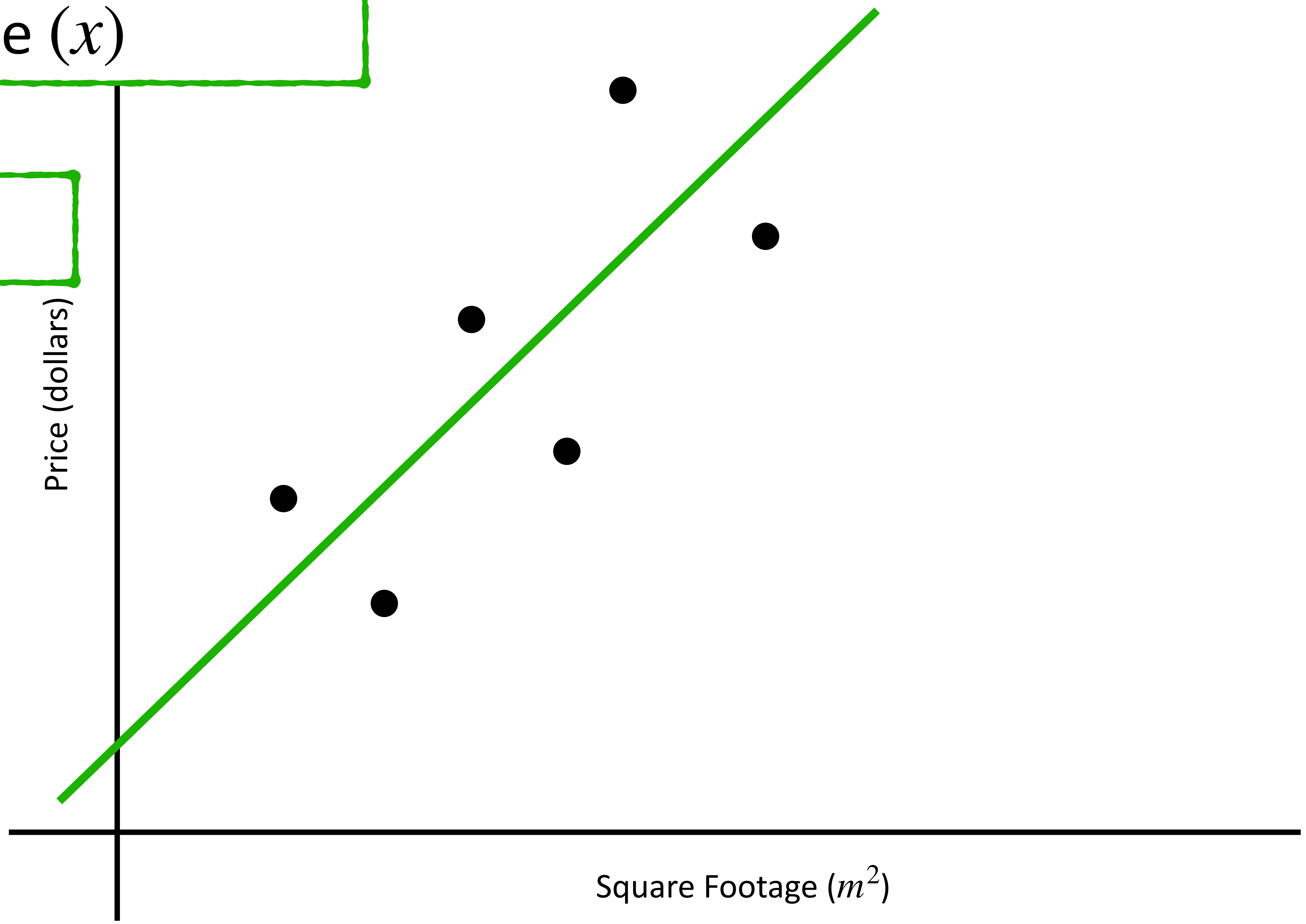
1 dependent variable

1 independent variable

$$\hat{y} = \beta_0 + \beta_1 \hat{x}$$

Eg: Predict Price of a house (y) given Square Footage (x)

2 Parameters - β_0 and β_1

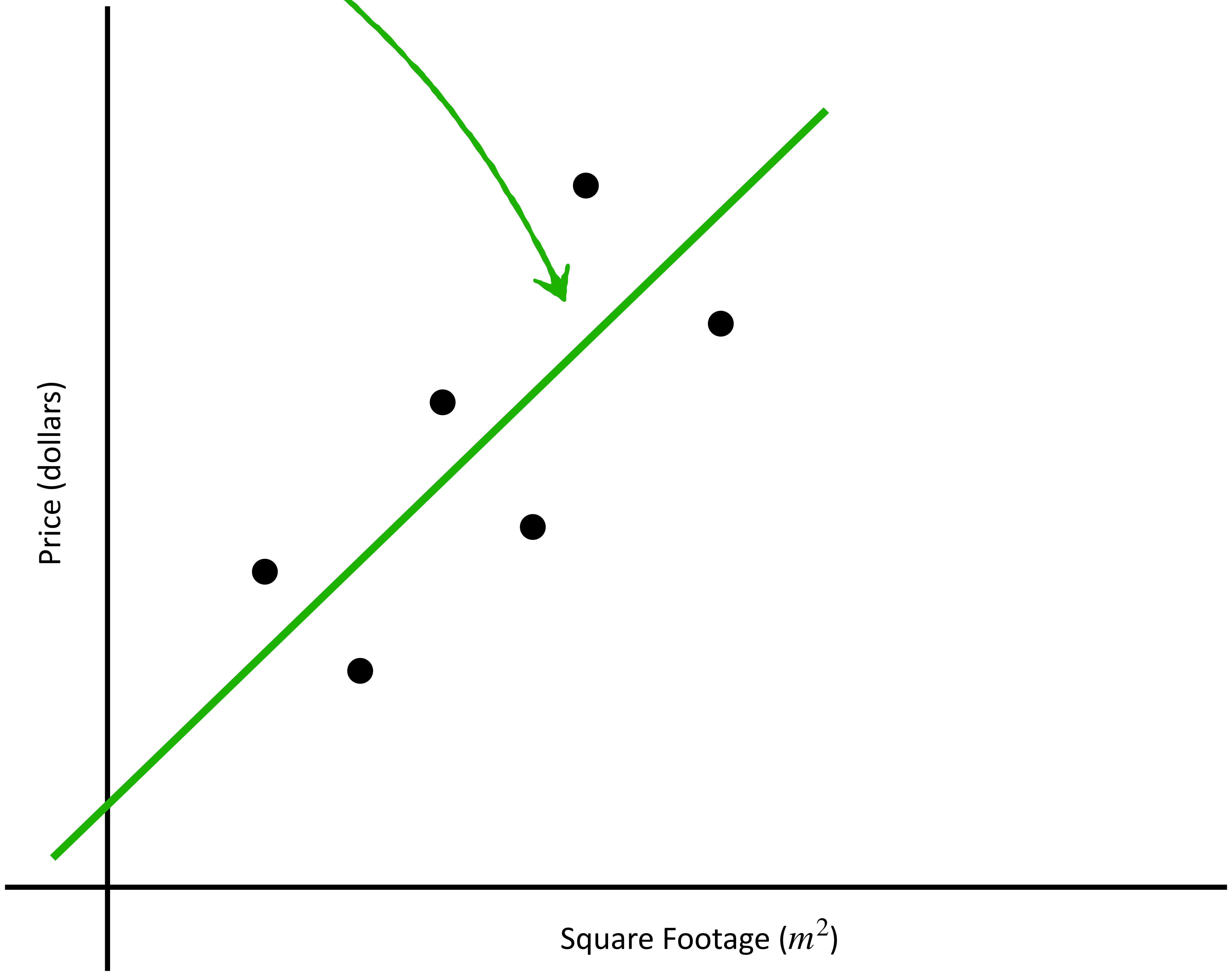


Data is synthetic and not plotted to scale

Simple Linear Regression

Line of Best Fit...

$$\hat{y} = \beta_0 + \beta_1 \hat{x}$$



Data is synthetic and not plotted to scale

Simple Linear Regression

Line of Best Fit…

…can be represented as a matrix

$$\hat{y} = \beta_0 + \beta_1 \hat{x}$$

$$\hat{y}_0 = 1 \times \beta_0 + \hat{x}_0 \times \beta_1$$

$$\hat{y}_1 = 1 \times \beta_0 + \hat{x}_1 \times \beta_1$$

$$\hat{y}_2 = 1 \times \beta_0 + \hat{x}_2 \times \beta_1$$

$$\hat{y}_3 = 1 \times \beta_0 + \hat{x}_3 \times \beta_1$$

$$\vdots$$

$$\hat{y}_n = 1 \times \beta_0 + \hat{x}_n \times \beta_1$$

$$\begin{bmatrix} \hat{y}_0 \\ \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ . \\ . \\ . \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} 1 & \hat{x}_0 \\ 1 & \hat{x}_1 \\ 1 & \hat{x}_2 \\ 1 & \hat{x}_3 \\ . & . \\ . & . \\ . & . \\ 1 & \hat{x}_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

- 1 dependent variable $\hat{y}$
- 1 independent variables $\hat{x}$
- 2 parameters - β_0 and β_1

Data is synthetic and not plotted to scale

Simple Linear Regression

Linear Model in 2 Dimensions

$$\hat{y} = \beta_0 + \beta_1 \hat{x}$$

$$\hat{y}_0 = 1 \times \beta_0 + \hat{x}_0 \times \beta_1$$

$$\hat{y}_1 = 1 \times \beta_0 + \hat{x}_1 \times \beta_1$$

$$\hat{y}_2 = 1 \times \beta_0 + \hat{x}_2 \times \beta_1$$

$$\hat{y}_3 = 1 \times \beta_0 + \hat{x}_3 \times \beta_1$$

$$\vdots$$

$$\hat{y}_n = 1 \times \beta_0 + \hat{x}_n \times \beta_1$$

$$\begin{bmatrix} \hat{y}_0 \\ \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ . \\ . \\ . \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} 1 & \hat{x}_0 \\ 1 & \hat{x}_1 \\ 1 & \hat{x}_2 \\ 1 & \hat{x}_3 \\ . & . \\ . & . \\ . & . \\ 1 & \hat{x}_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

- 1 dependent variable $\hat{y}$
- 1 independent variables $\hat{x}$
- 2 parameters - β_0 and β_1

Data is synthetic and not plotted to scale

Multiple Regression

Linear Model in 3 Dimensions

$$\hat{y} = \beta_0 + \beta_1 \hat{x}_1 + \beta_2 \hat{x}_2$$

$$\hat{y}_0 = 1 \times \beta_0 + \hat{x}_{10} \times \beta_1 + \hat{x}_{20} \times \beta_2$$

$$\hat{y}_1 = 1 \times \beta_0 + \hat{x}_{11} \times \beta_1 + \hat{x}_{21} \times \beta_1$$

$$\hat{y}_2 = 1 \times \beta_0 + \hat{x}_{12} \times \beta_1 + \hat{x}_{22} \times \beta_2$$

$$\hat{y}_3 = 1 \times \beta_0 + \hat{x}_{13} \times \beta_1 + \hat{x}_{23} \times \beta_2$$

$$\vdots$$

$$\hat{y}_n = 1 \times \beta_0 + \hat{x}_{1n} \times \beta_1 + \hat{x}_{2n} \times \beta_2$$

$$\begin{bmatrix} \hat{y}_0 \\ \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ . \\ . \\ . \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} 1 & \hat{x}_{10} & \hat{x}_{20} \\ 1 & \hat{x}_{11} & \hat{x}_{21} \\ 1 & \hat{x}_{12} & \hat{x}_{22} \\ 1 & \hat{x}_{13} & \hat{x}_{23} \\ . & . & . \\ . & . & . \\ . & . & . \\ 1 & \hat{x}_{1n} & \hat{x}_{2n} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

- 1 dependent variable $\hat{y}$
- 2 independent variables $\hat{x}_1$ and $\hat{x}_2$
- 3 parameters - β_0, β_1 and β_2

Data is synthetic and not plotted to scale

Multiple Regression

Linear Model in 4 Dimensions

$$\hat{y} = \beta_0 + \beta_1\hat{x}_1 + \beta_2\hat{x}_2 + \beta_3\hat{x}_3$$

$$\hat{y}_n = 1 \times \beta_0 + \hat{x}_{1n} \times \beta_1 + \hat{x}_{2n} \times \beta_2 + \hat{x}_{3n} \times \beta_3$$

$$\begin{bmatrix} \hat{y}_0 \\ \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ . \\ . \\ . \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} 1 & \hat{x}_{10} & \hat{x}_{20} & \hat{x}_{30} \\ 1 & \hat{x}_{11} & \hat{x}_{21} & \hat{x}_{21} \\ 1 & \hat{x}_{12} & \hat{x}_{22} & \hat{x}_{22} \\ 1 & \hat{x}_{13} & \hat{x}_{23} & \hat{x}_{23} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ 1 & \hat{x}_{1n} & \hat{x}_{2n} & \hat{x}_{3n} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$$

- 1 dependent variable $\hat{y}$
- 3 independent variables $\hat{x}_1$, $\hat{x}_2$ and $\hat{x}_3$
- 4 parameters - β_0, β_1, β_2 and β_3

Data is synthetic and not plotted to scale

Multiple Regression

Linear Model in
$k+1$ Dimensions

$$\hat{y} = \beta_0 + \beta_1\hat{x}_1 + \beta_2\hat{x}_2 + \beta_3\hat{x}_3 + \ldots + \beta_k\hat{x}_k$$

$$\hat{y}_n = 1 \times \beta_0 + \hat{x}_{1n} \times \beta_1 + \hat{x}_{2n} \times \beta_2 + \hat{x}_{3n} \times \beta_3 + \ldots + \hat{x}_{kn} \times \beta_k$$

$$\begin{bmatrix} \hat{y}_0 \\ \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ . \\ . \\ . \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} 1 & \hat{x}_{10} & \hat{x}_{20} & \hat{x}_{30} & . & . & . & \hat{x}_{k0} \\ 1 & \hat{x}_{11} & \hat{x}_{21} & \hat{x}_{21} & . & . & . & \hat{x}_{k1} \\ 1 & \hat{x}_{12} & \hat{x}_{22} & \hat{x}_{22} & . & . & . & \hat{x}_{k2} \\ 1 & \hat{x}_{13} & \hat{x}_{23} & \hat{x}_{23} & . & . & . & \hat{x}_{k3} \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ 1 & \hat{x}_{1n} & \hat{x}_{2n} & \hat{x}_{3n} & . & . & . & \hat{x}_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ . \\ . \\ . \\ \beta_k \end{bmatrix}$$

- 1 dependent variable $\hat{y}$
- k independent variables $\hat{x}_1, \hat{x}_2, \hat{x}_3 \ldots \hat{x}_k$
- $k+1$ parameters - $\beta_0, \beta_1, \beta_2, \beta_3 \ldots \beta_k$

Data is synthetic and not plotted to scale

Multiple Regression

Linear Model in
$k+1$ Dimensions

$$\hat{y} = \beta_0 + \beta_1\hat{x}_1 + \beta_2\hat{x}_2 + \beta_3\hat{x}_3 + \ldots + \beta_k\hat{x}_k$$

$$\hat{Y} = \hat{X}\beta$$

$\hat{Y}$ and $\hat{X}$ are matrices

$$\begin{bmatrix} \hat{y}_0 \\ \hat{y}_1 \\ \\ \\ \\ \\ . \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} 1 & \hat{x}_{10} & \hat{x}_{20} & \hat{x}_{30} & . & . & . & \hat{x}_{k0} \\ 1 & \hat{x}_{11} & \hat{x}_{21} & \hat{x}_{21} & . & . & . & \hat{x}_{k1} \\ 1 & \hat{x}_{12} & & & & . & . & \hat{x}_{k2} \\ 1 & \hat{x}_{13} & & & & . & . & \hat{x}_{k3} \\ . & . & . & & & . & . & . \\ . & . & . & & & . & . & . \\ . & . & . & . & . & . & . & . \\ 1 & \hat{x}_{1n} & \hat{x}_{2n} & \hat{x}_{3n} & . & . & . & \hat{x}_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \\ \\ \\ \\ . \\ \beta_k \end{bmatrix}$$

$\hat{Y}$: $(n+1)\times 1$

$\hat{X}$: $(n+1)\times(k+1)$

β: $(k+1)\times 1$

- 1 dependent variable $\hat{y}$
- k independent variables $\hat{x}_1, \hat{x}_2, \hat{x}_3 \ldots \hat{x}_k$
- $k+1$ parameters - $\beta_0, \beta_1, \beta_2, \beta_3 \ldots \beta_k$

Data is synthetic and not plotted to scale

Multiple Regression

$$\hat{y} = \beta_0 + \beta_1 \hat{x}$$

Given a matrix (Y) of observations

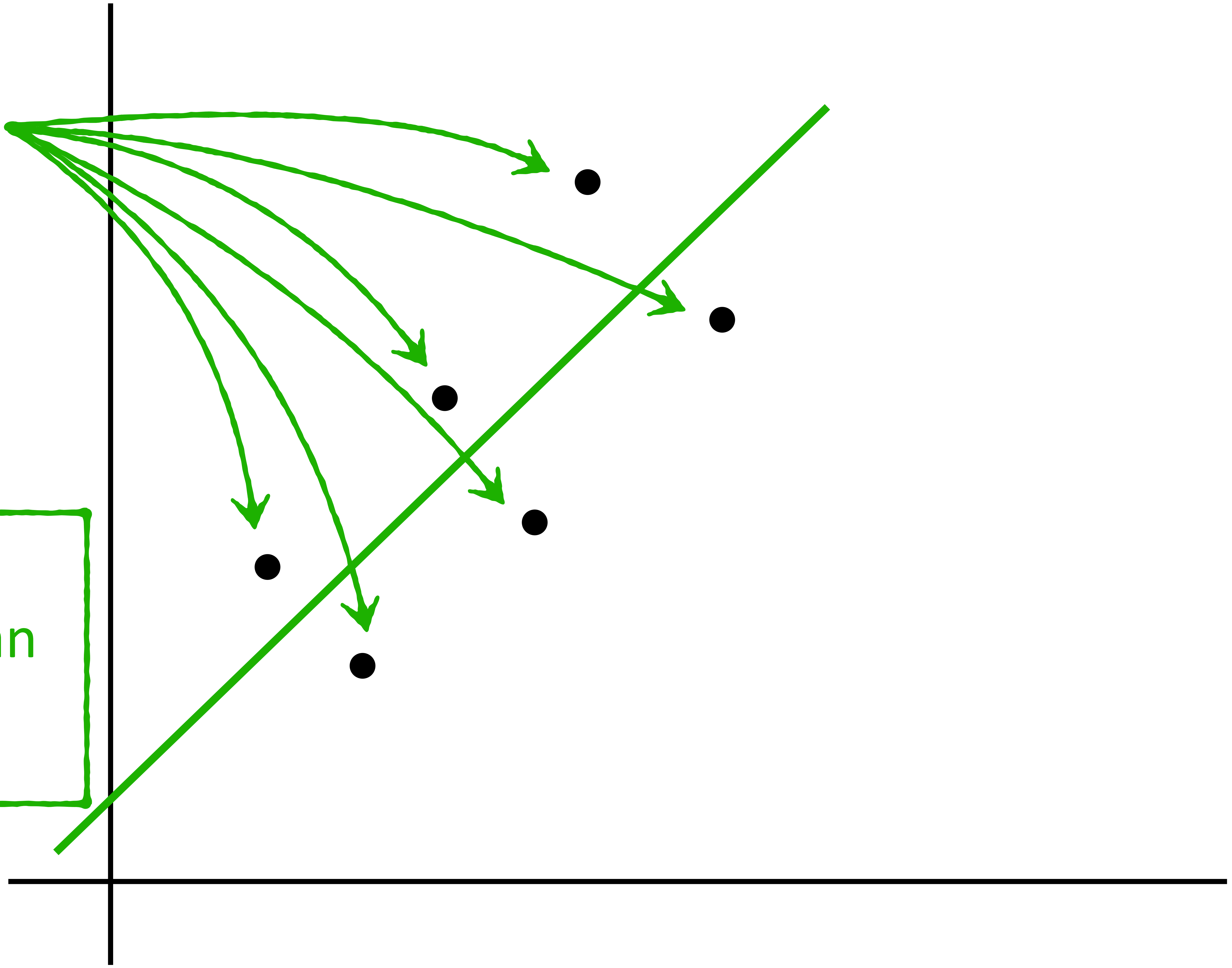
$$\hat{Y} = \hat{X}\beta$$

The **Mean Squared Error (MSE)**

$$\frac{1}{n} \| Y - \hat{Y} \|^2$$

The two parallel vertical lines mean that this is the Euclidean Norm of the matrix

See Tutorial on Vectors & Matrices



Data is synthetic and not plotted to scale

$$\hat{y} = \beta_0 + \beta_1 \hat{x}$$

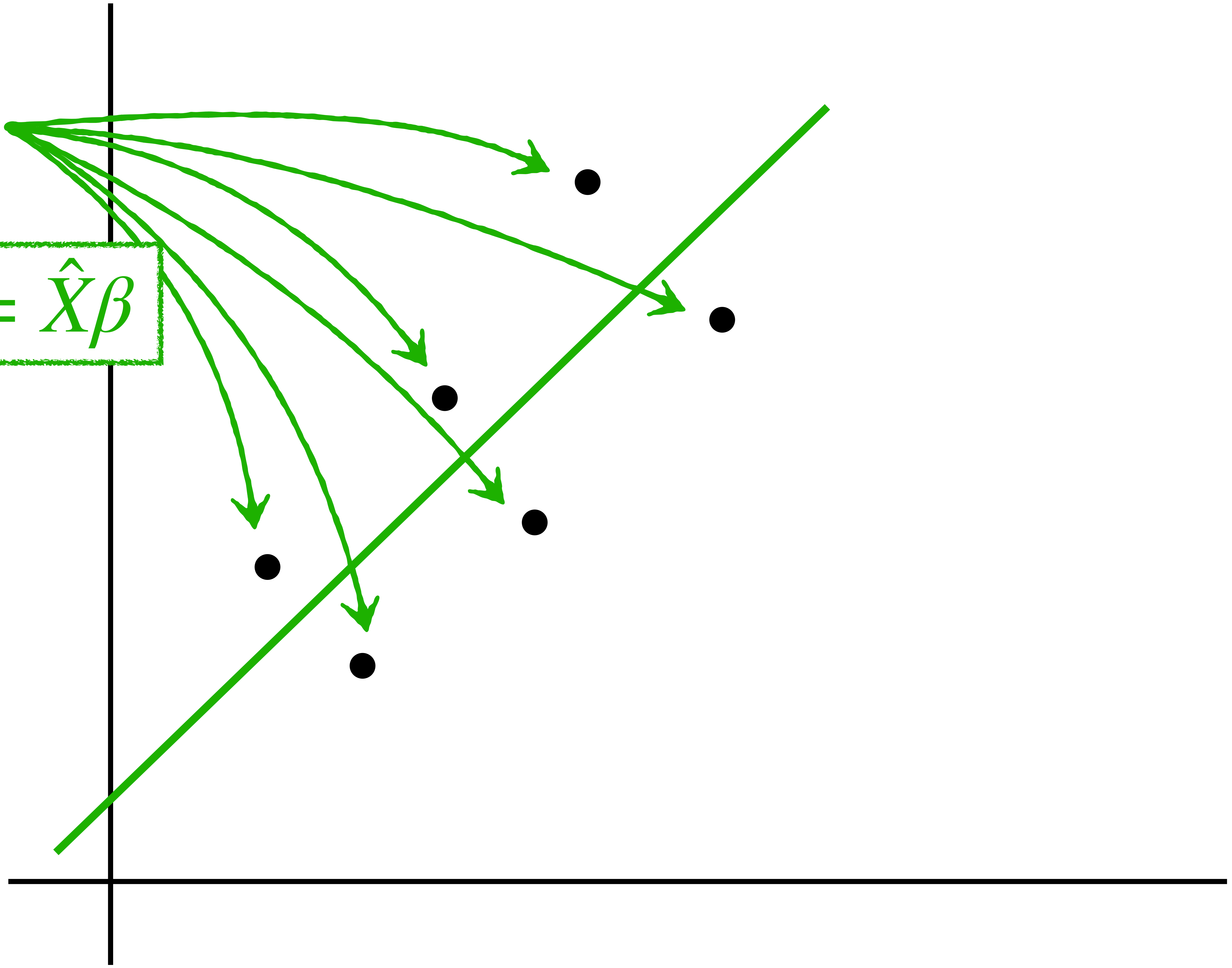
$$\hat{Y} = \hat{X}\beta$$

Given a matrix (Y) of observations

Substituting $\hat{Y} = \hat{X}\beta$

The **Mean Squared Error (MSE)**

$$\frac{1}{n} \| Y - \hat{Y} \|^2 = \frac{1}{n} \| Y - \hat{X}\beta \|^2$$



Data is synthetic and not plotted to scale

Multiple Regression

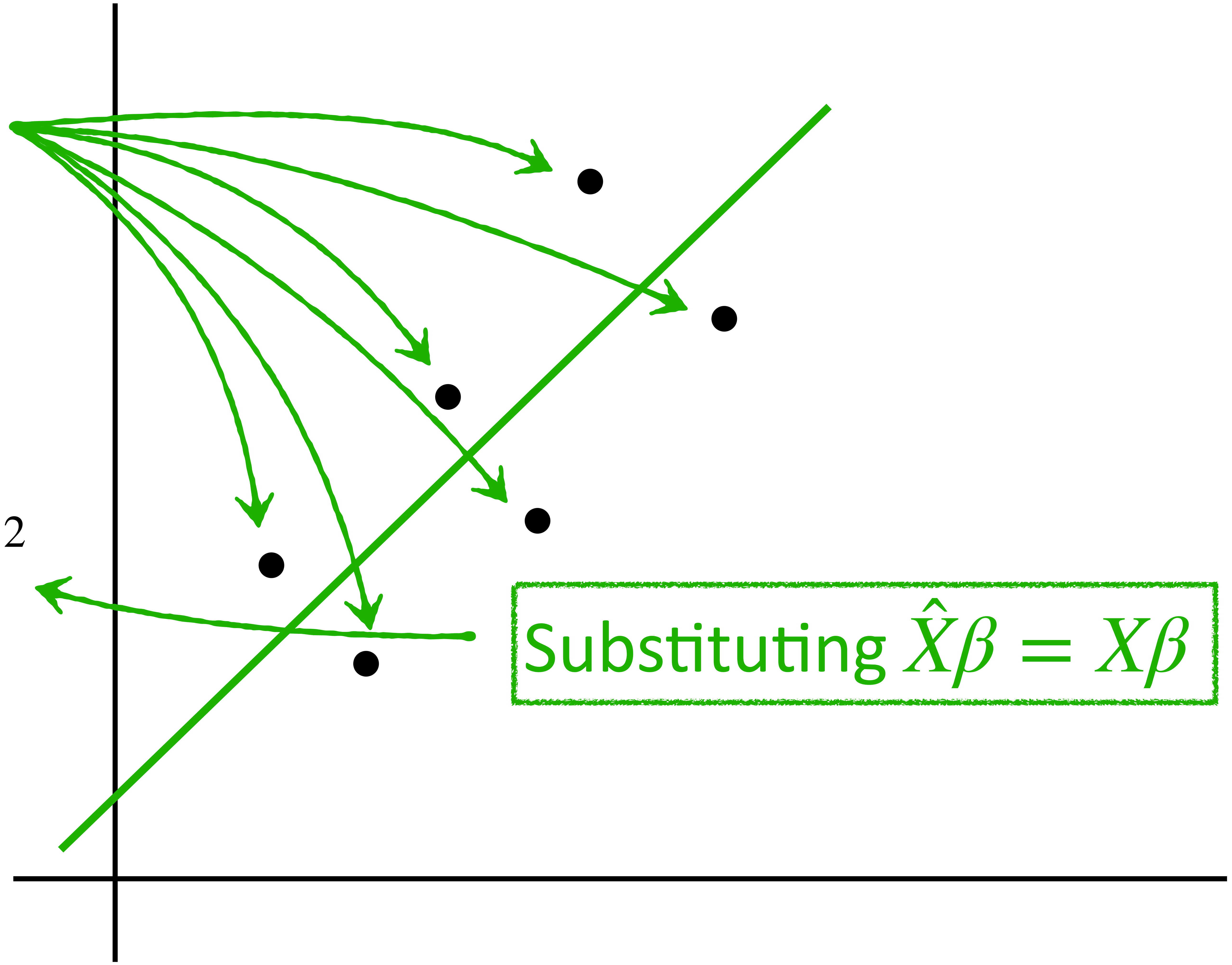
$$\hat{y} = \beta_0 + \beta_1 \hat{x}$$

Given a matrix (Y) of observations

$$\hat{Y} = \hat{X}\beta$$

The **Mean Squared Error (MSE)**

$$\frac{1}{n} \| Y - \hat{Y} \|^2 = \frac{1}{n} \| Y - \hat{X}\beta \|^2 = \frac{1}{n} \| Y - X\beta \|^2$$



Data is synthetic and not plotted to scale

Multiple Regression

Linear Model in $k+1$ Dimensions

$$\hat{Y} = \hat{X}\beta$$

The **Mean Squared Error (MSE):**

$$\frac{1}{n}\| Y - X\beta \|^2$$

The Problem Statement:

Multiple Regression: Compute the matrix β such that the Mean Squared Error (MSE) is minimized.

Data is synthetic and not plotted to scale

The Problem Statement:

Multiple Regression: Compute the matrix β such that the Mean Squared Error (MSE) is minimized.

$$\frac{1}{n} \| Y - X\beta \|^2$$

This is the cost function (aka loss function) that we must minimize.

$$\frac{\partial}{\partial \beta} \frac{1}{n} \| Y - X\beta \|^2 = 0$$

We take the partial derivate and set it = 0

$$\beta = (X^T X)^{-1} X^T Y$$

Solving for β

For the details of the derivation see the tutorial on Derivation of the Matrix Form for Multiple Regression

The Problem Statement:

Multiple Regression: Compute the matrix β such that the Mean Squared Error (MSE) is minimized.

Solution:

$$\beta = (X^T X)^{-1} X^T Y$$

This is the **Closed form solution** for **Multiple Regression**. However this requires inverting a matrix which is not always possible.

For more details on the reasons why see the tutorial on Vectors & Matrices

Related Tutorials & Textbooks

[Simple Linear Regression]

A statistical technique of making predictions from data. The tutorial introduces a linear model in two dimensions and uses that model to predict the value of one dependent variable given one independent variable.

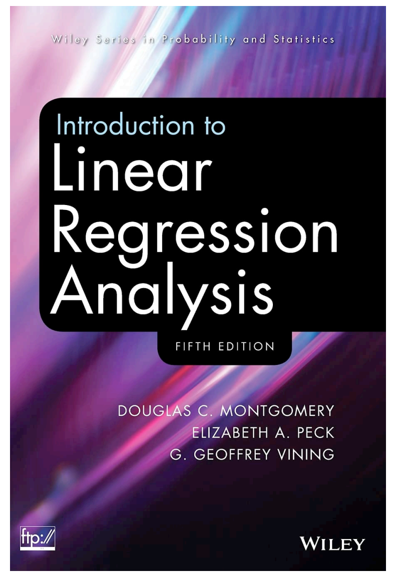
Multiple Regression: Deriving the Matrix Form

A proof of the closed form matrix representation of the multiple regression model. This closed form represents a linear model with k + 1 parameters and solves for the matrix β. This requires a matrix inverse that is not always possible.

Gradient Descent for Multiple Regression

Gradient Descent algorithm for multiple regression and how it can be used to optimize k + 1 parameters for a Linear model in multiple dimensions.

Recommended Textbooks



Introduction to Linear Regression Analysis

by Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining

For a complete list of tutorials see:
https://arrsingh.com/ai-tutorials