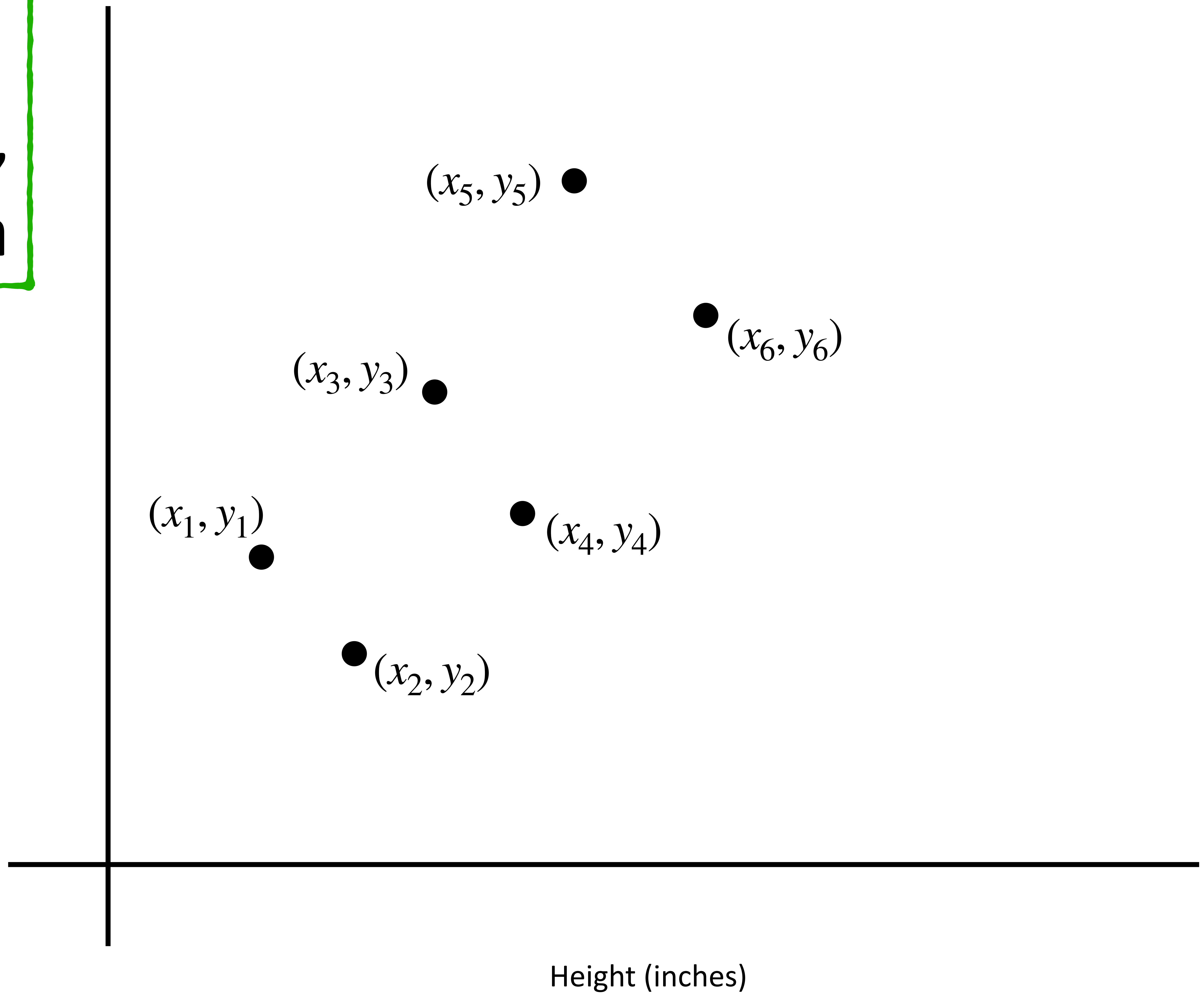
Gradient Descent

Simple Linear Regression using Gradient Descent

Rahul Singh
rsingh@arrsingh.com

Simple Linear Regression

Problem Statement: Given a set of data points in $\mathbf{R}^2$,
$(x_1, y_1), (x_2, y_2), (x_0, y_0) \ldots (x_n, y_n)$,
find the line that **best fits the data**

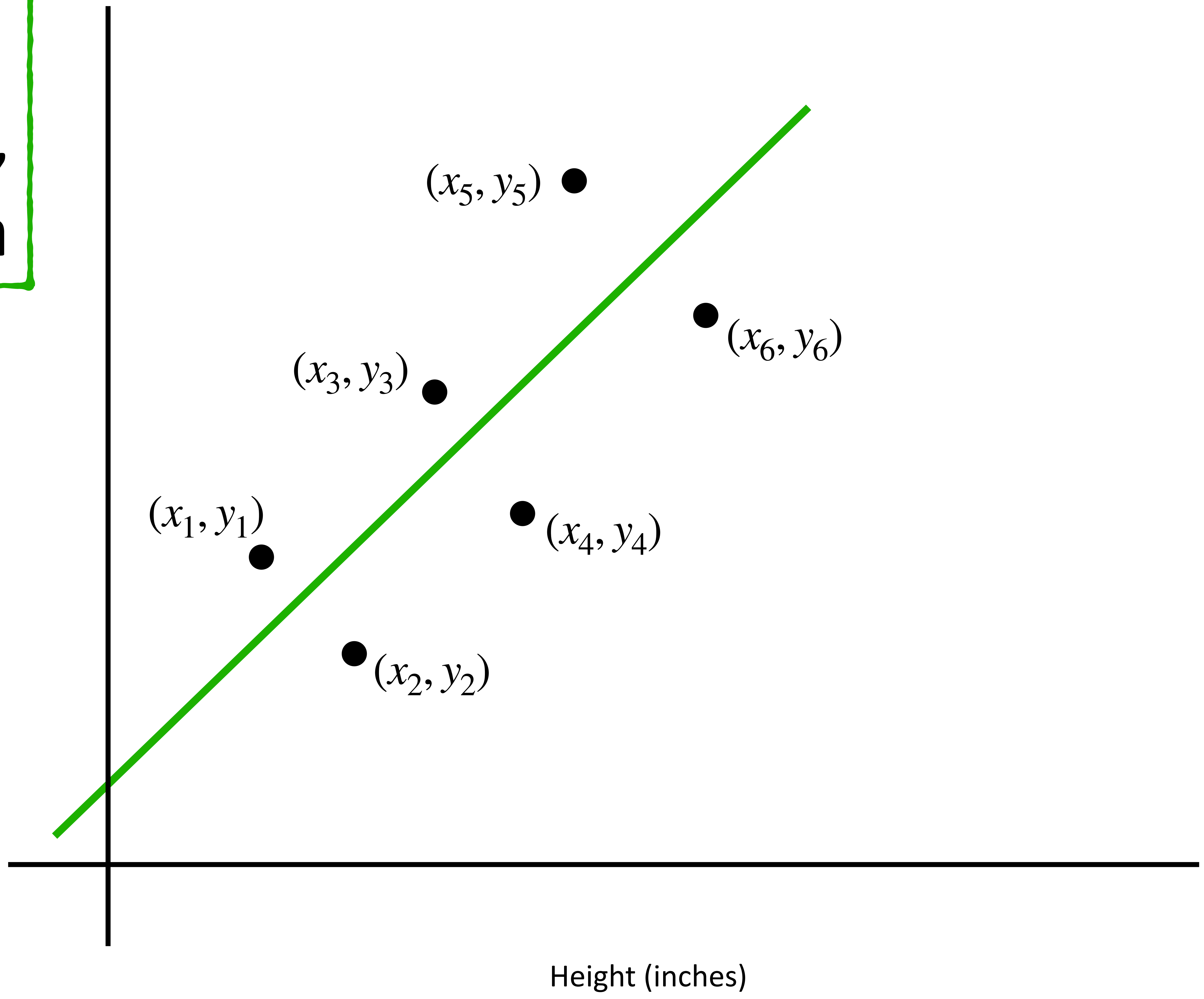


Data is synthetic and not plotted to scale

Simple Linear Regression

Problem Statement: Given a set of data points in $\mathbf{R}^2$, $(x_1, y_1), (x_2, y_2), (x_0, y_0) \ldots (x_n, y_n)$, find the line that **best fits the data**

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



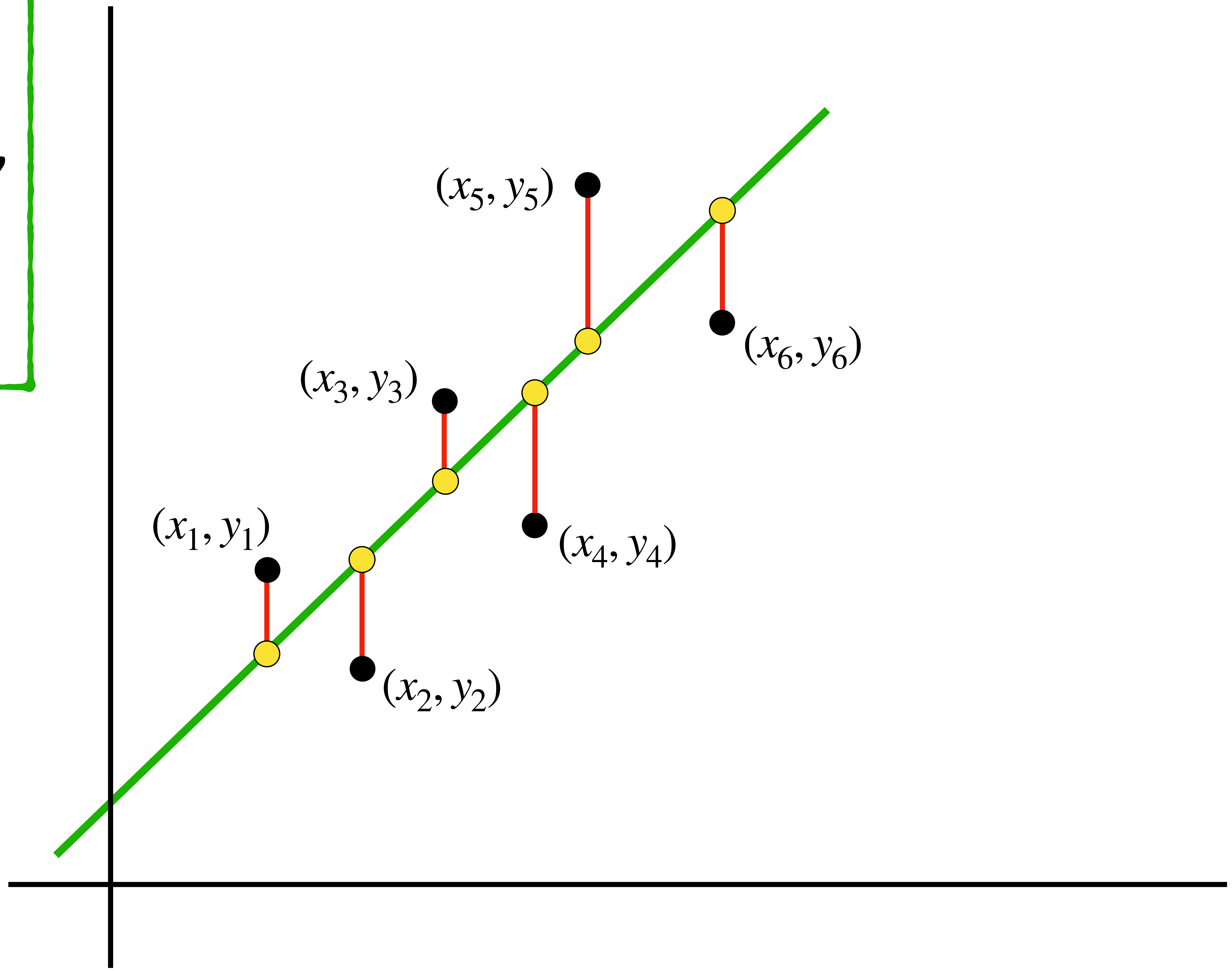
Data is synthetic and not plotted to scale

Simple Linear Regression

Problem Statement: Given a set of data points in $\mathbf{R}^2$,

$(x_1, y_1), (x_2, y_2), (x_0, y_0) \ldots (x_n, y_n)$,

find the line that minimizes the **Mean Squared Error (MSE)**



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

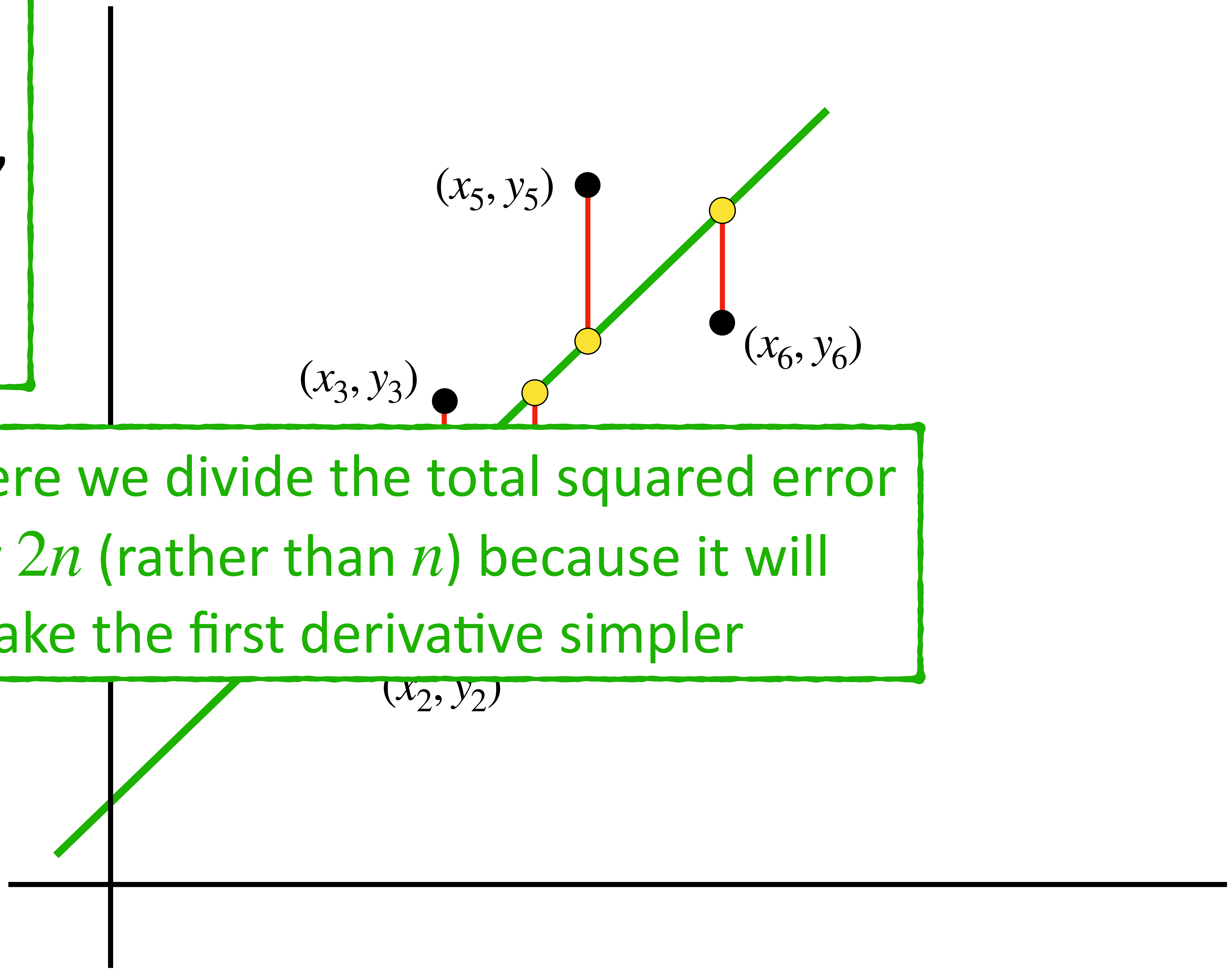
Mean Squared Error (MSE)

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

Data is synthetic and not plotted to scale

Simple Linear Regression

Problem Statement: Given a set of data points in $\mathbf{R}^2$, $(x_1, y_1), (x_2, y_2), (x_0, y_0) \ldots (x_n, y_n)$, find the line that minimizes the **Mean Squared Error (MSE)**



Here we divide the total squared error by $2n$ (rather than n) because it will make the first derivative simpler

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

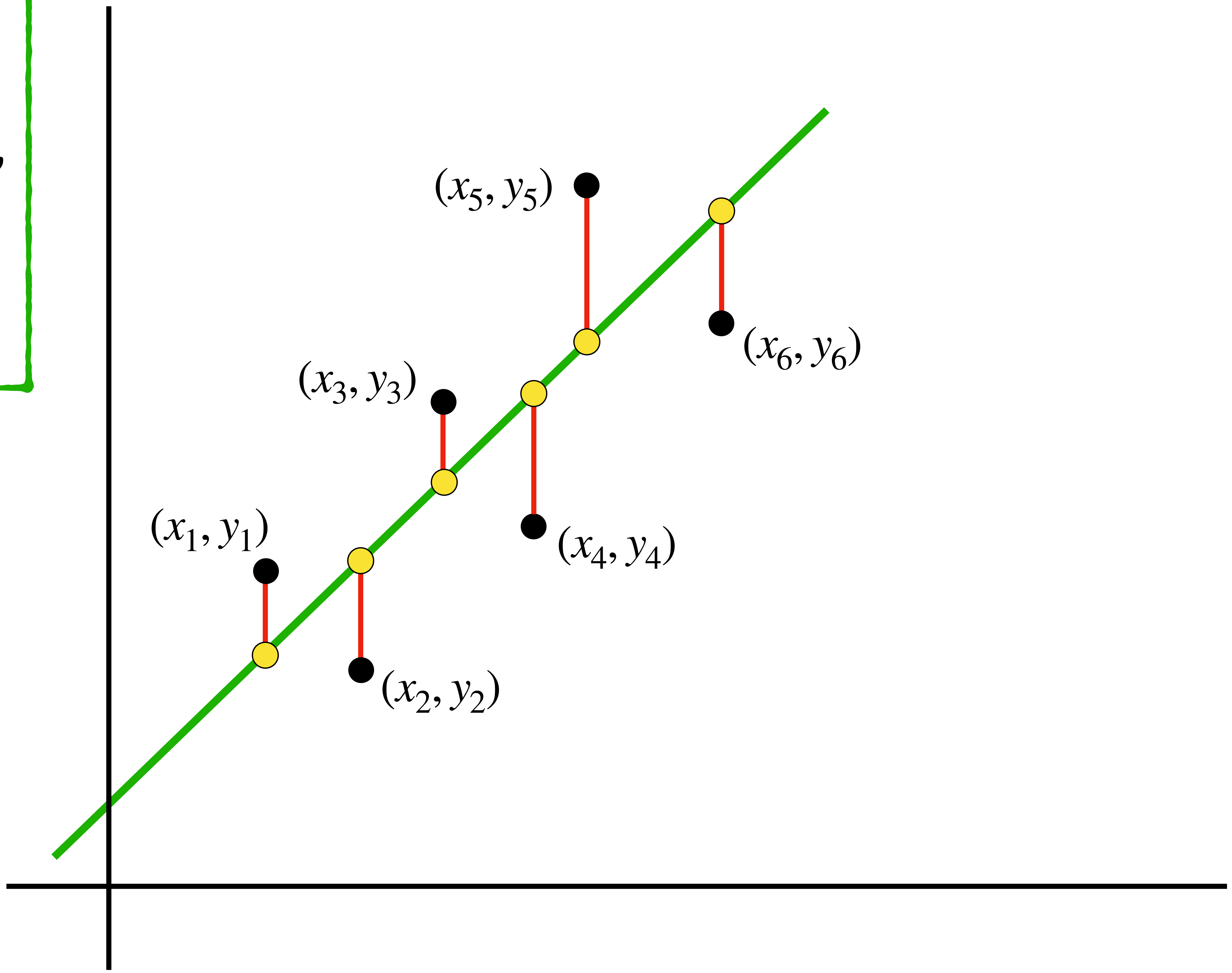
Mean Squared Error (MSE)

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

Data is synthetic and not plotted to scale

Simple Linear Regression

Problem Statement: Given a set of data points in $\mathbf{R}^2$, $(x_1, y_1), (x_2, y_2), (x_0, y_0) \ldots (x_n, y_n)$, find the line that minimizes the **Mean Squared Error (MSE)**



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

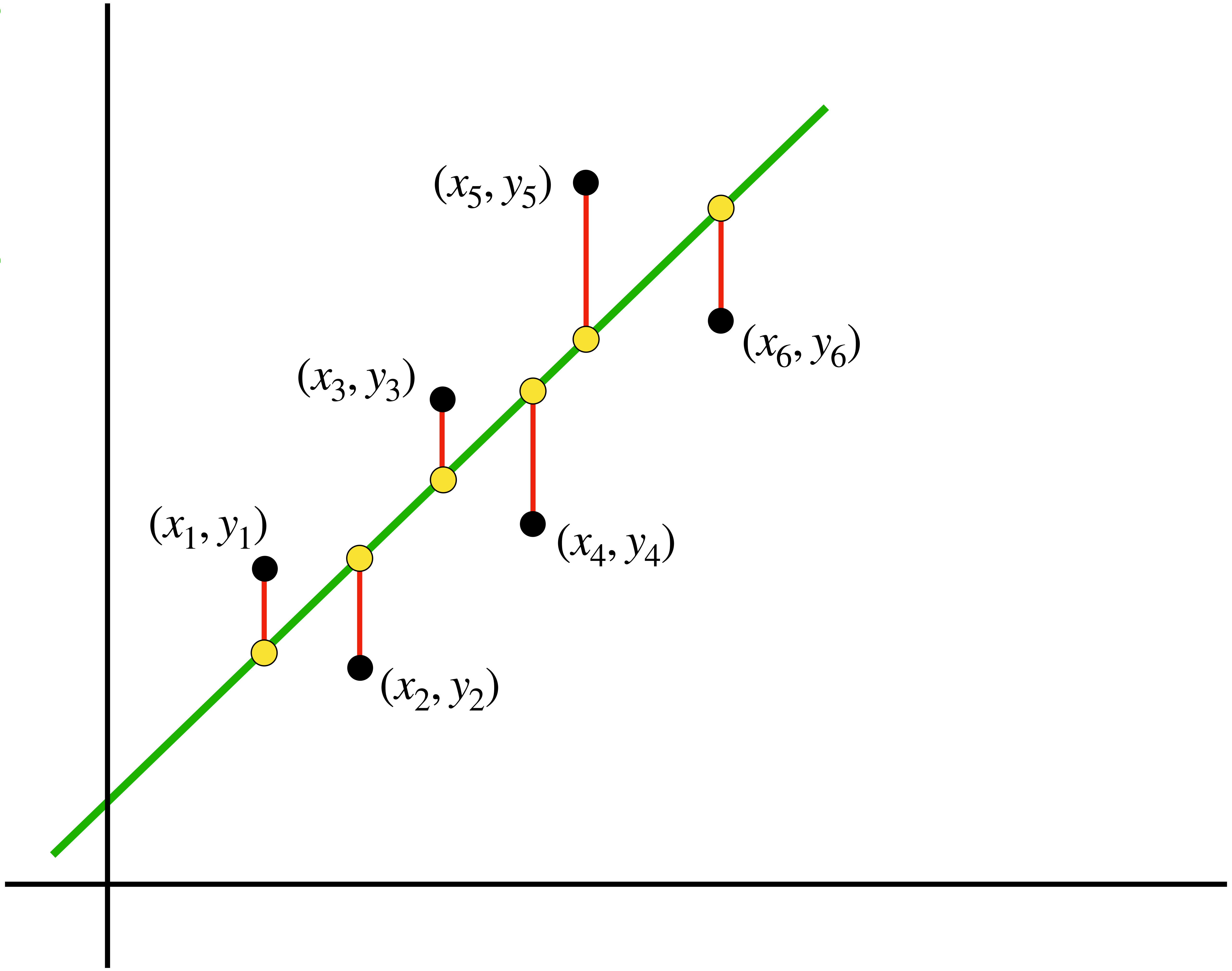
Mean Squared Error (MSE)

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

Data is synthetic and not plotted to scale

The Problem Statement:

Simple Linear Regression: Find the values of β_0 and β_1 such that the **Mean Squared Error (MSE)** is minimized.



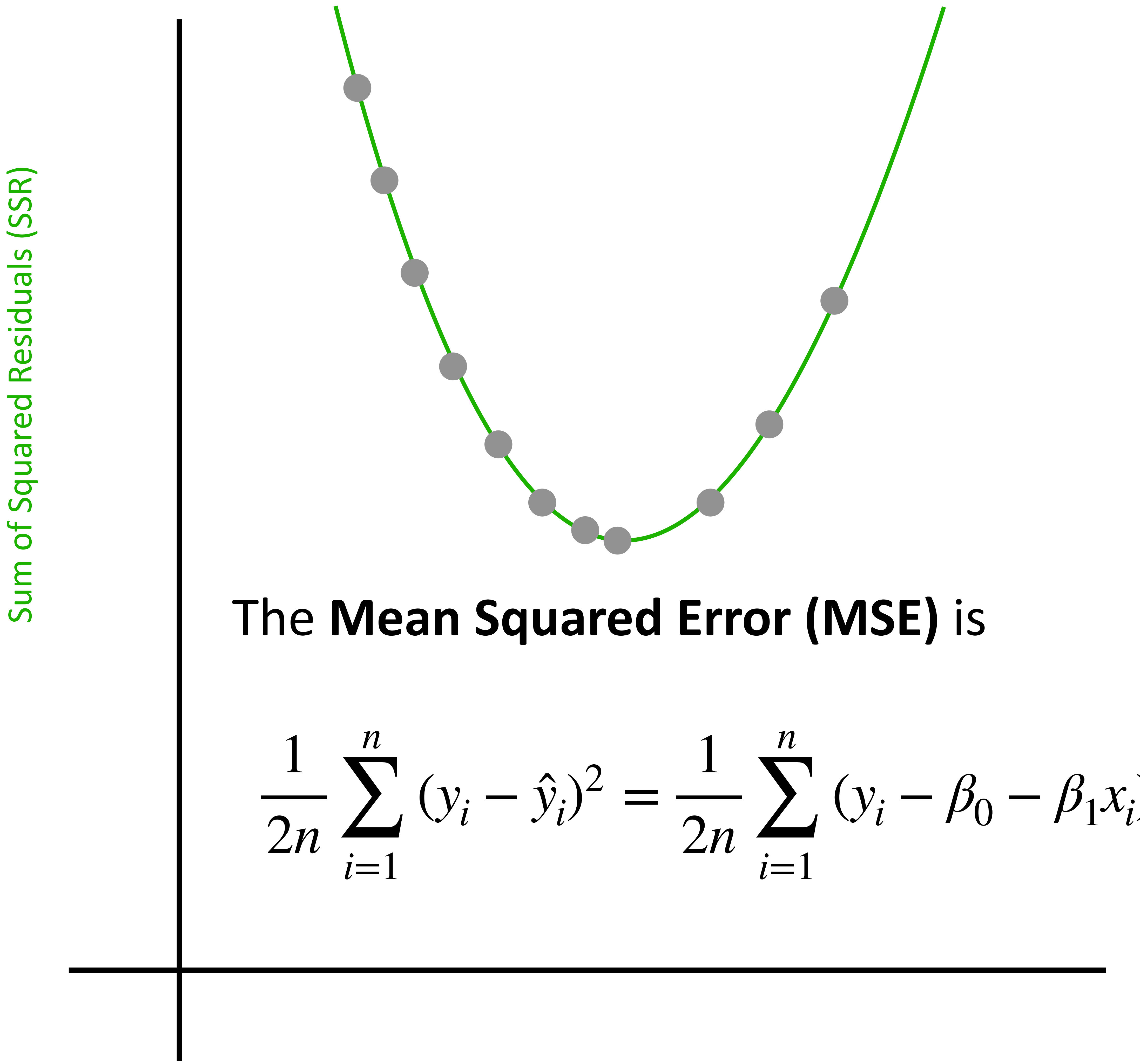
The line of best fit is $\hat{y} = \beta_0 + \beta_1 x$

Mean Squared Error (MSE)

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

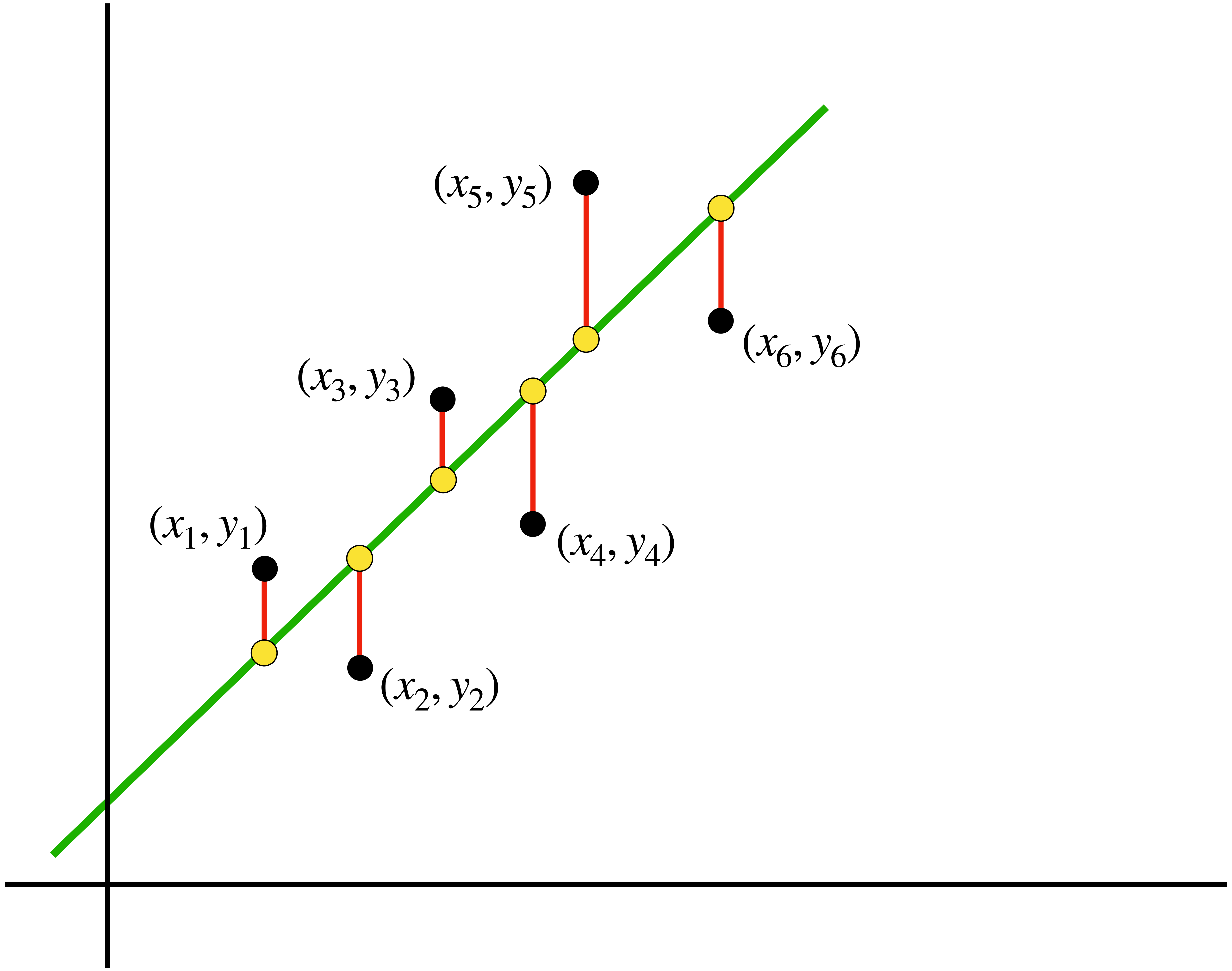
Data is synthetic and not plotted to scale

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

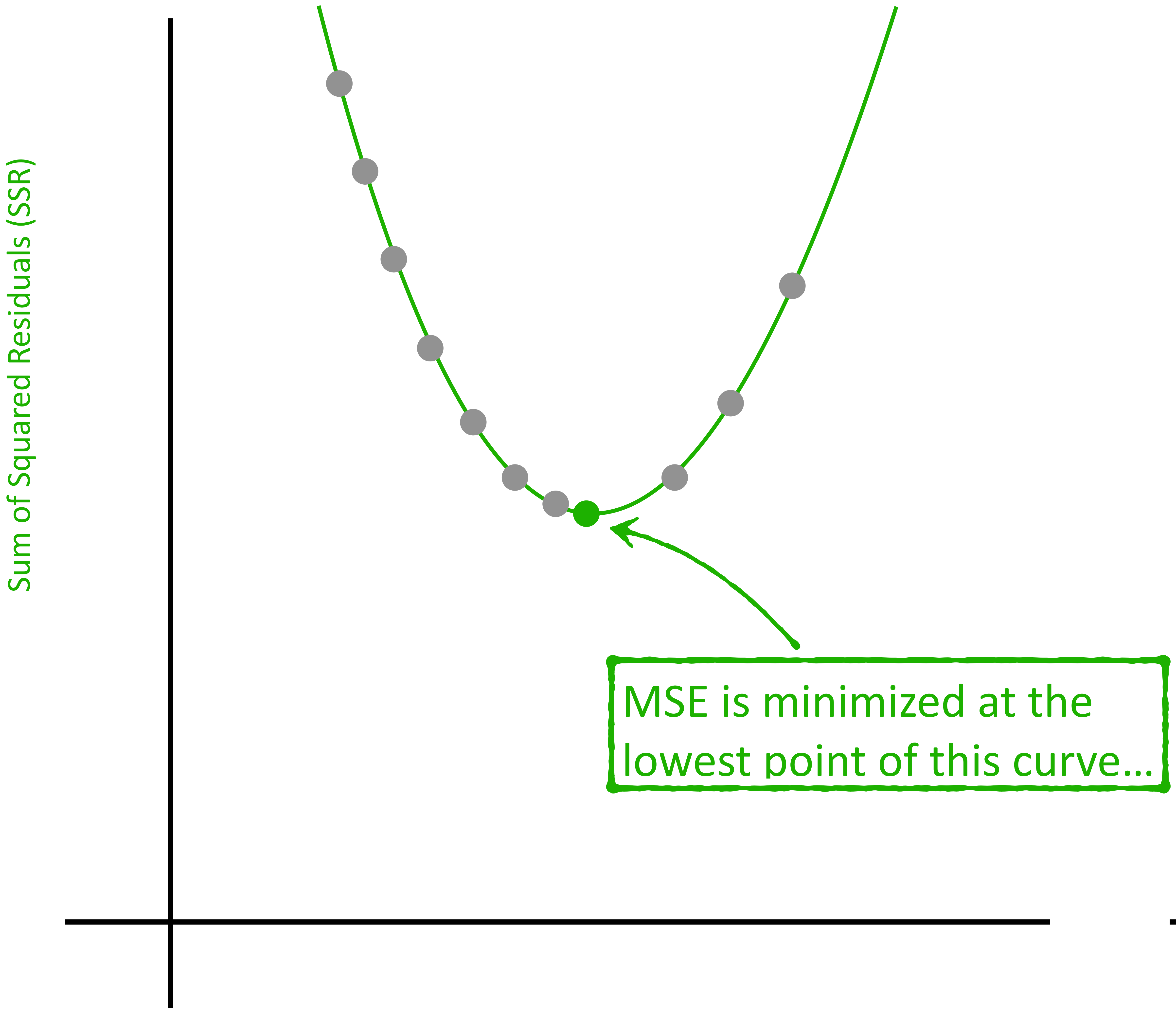


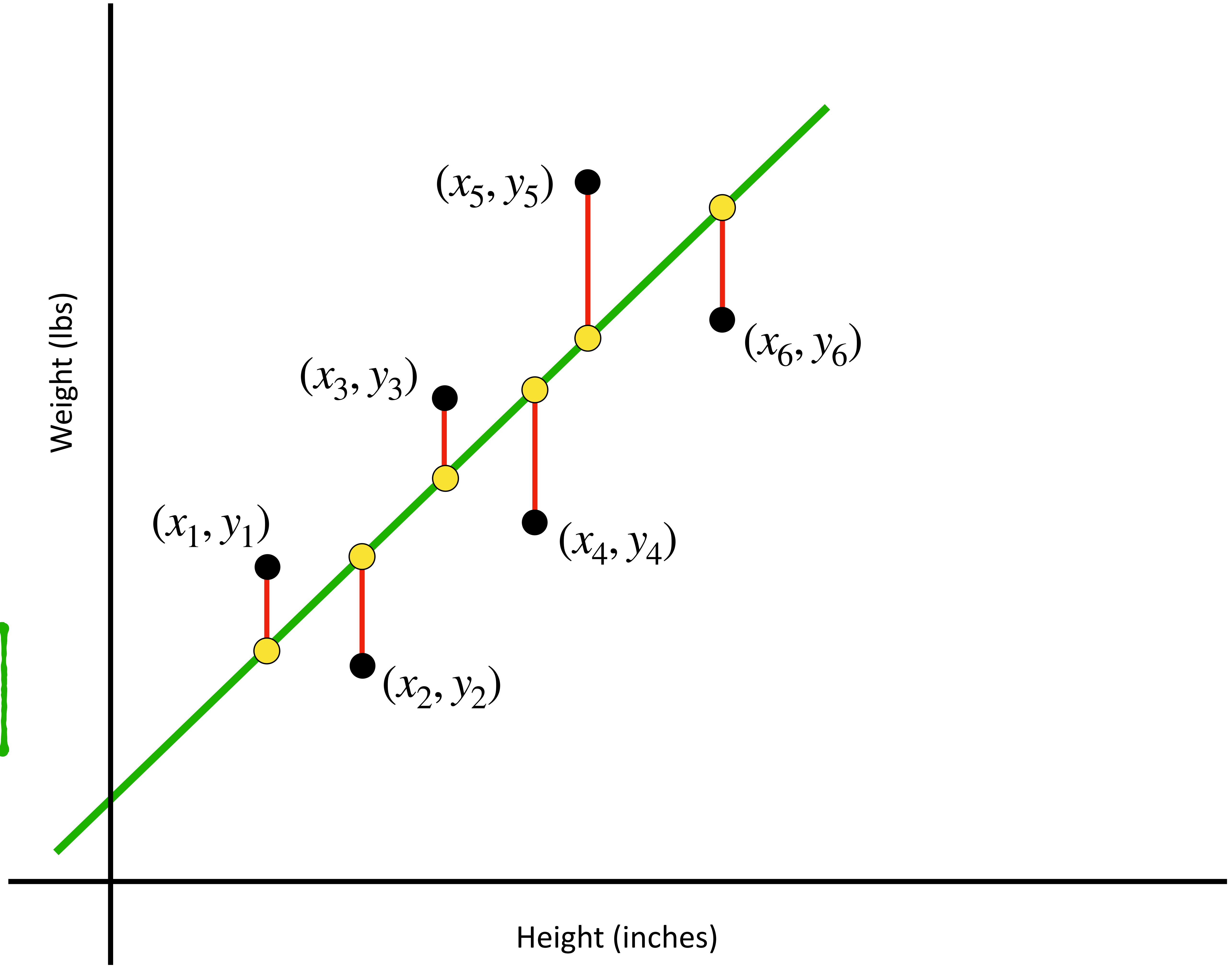
The **Mean Squared Error (MSE)** is

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



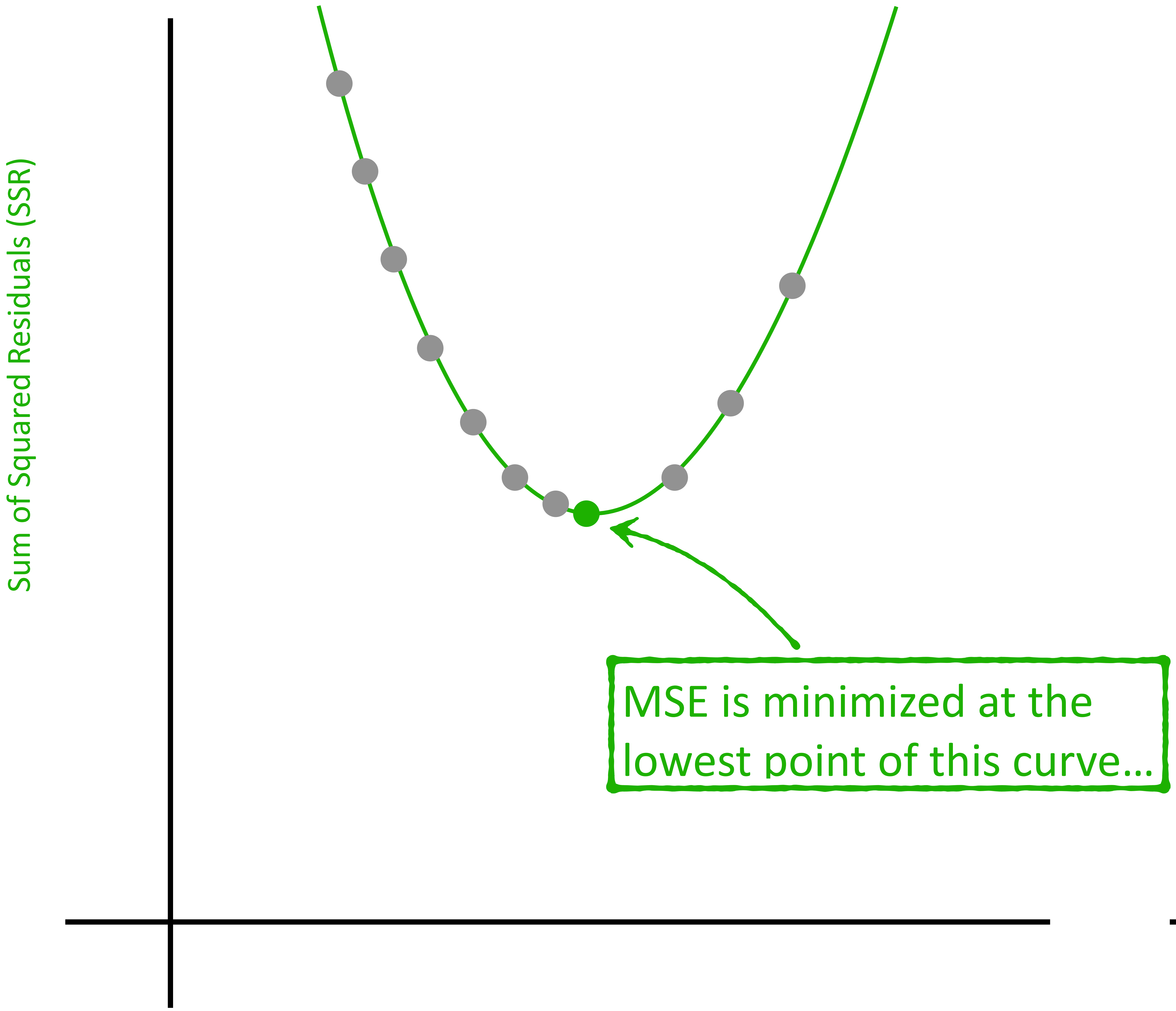
Mean Squared Error (MSE) for various values of β_0 and β_1 follows this curve

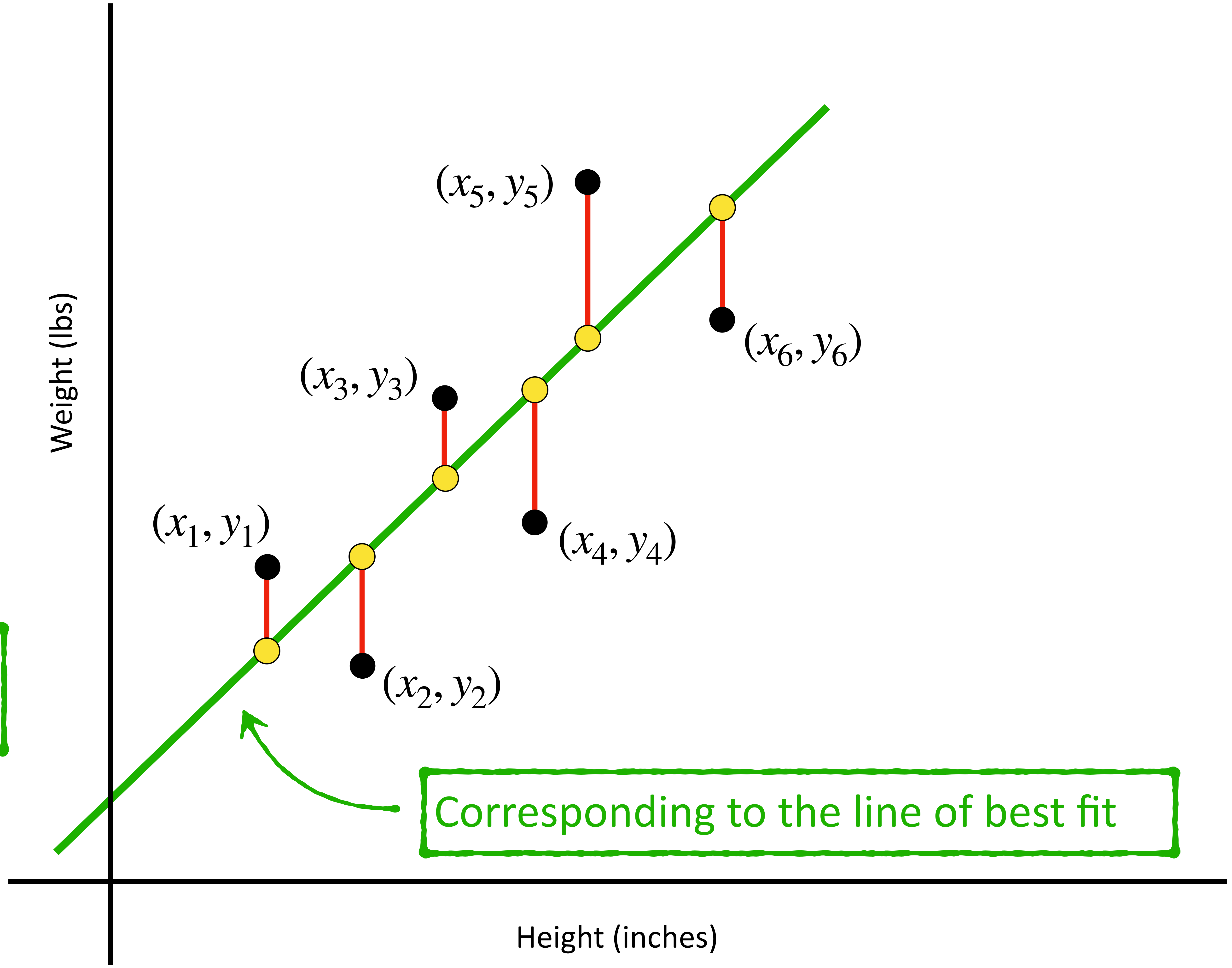




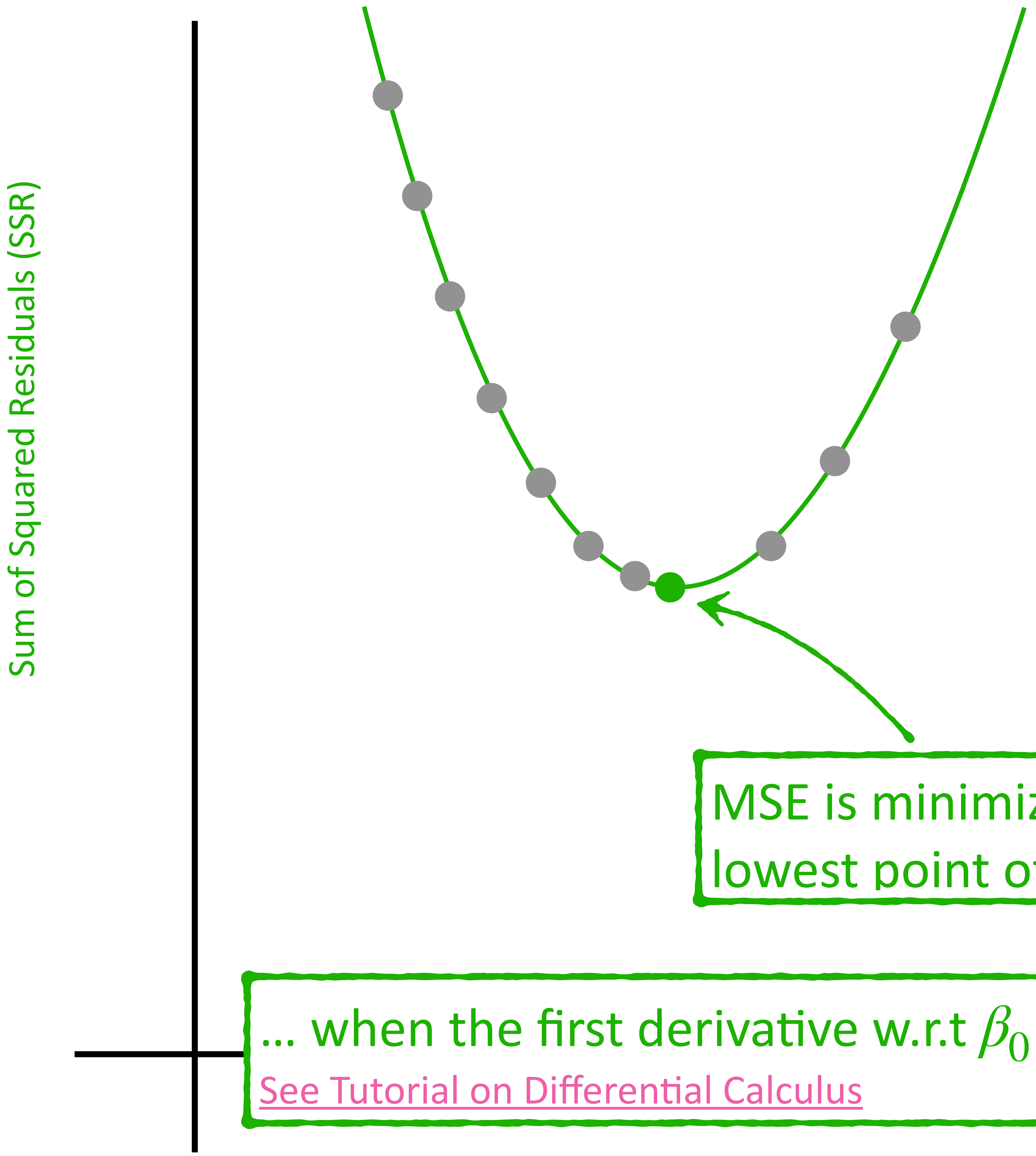
Simple Linear Regression

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





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



MSE is minimized at the lowest point of this curve…

… when the first derivative w.r.t β_0 and β_1 equals 0

See Tutorial on Differential Calculus

The **Mean Squared Error (MSE)** is…

$$\frac{1}{2n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\frac{1}{2n}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1x_i)^2$$

The **first derivative w.r.t** β_0 **and** β_1 is…

$$\frac{\partial}{\partial\beta_0}\frac{1}{2n}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1x_i)^2=-\frac{1}{n}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1x_i)$$

$$\frac{\partial}{\partial\beta_1}\frac{1}{2n}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1x_i)^2=-\frac{1}{n}\sum_{i=1}^{n}(x_iy_i-\beta_0x_i-\beta_1x_i^2)$$

Simple Linear Regression

We can find the optimal values for β_0 and β_1 by solving these two equations…

… as we did in the tutorial on Simple Linear Regression

The **Mean Squared Error (MSE)** is…

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

The **first derivative w.r.t** β_0 **and** β_1 is…

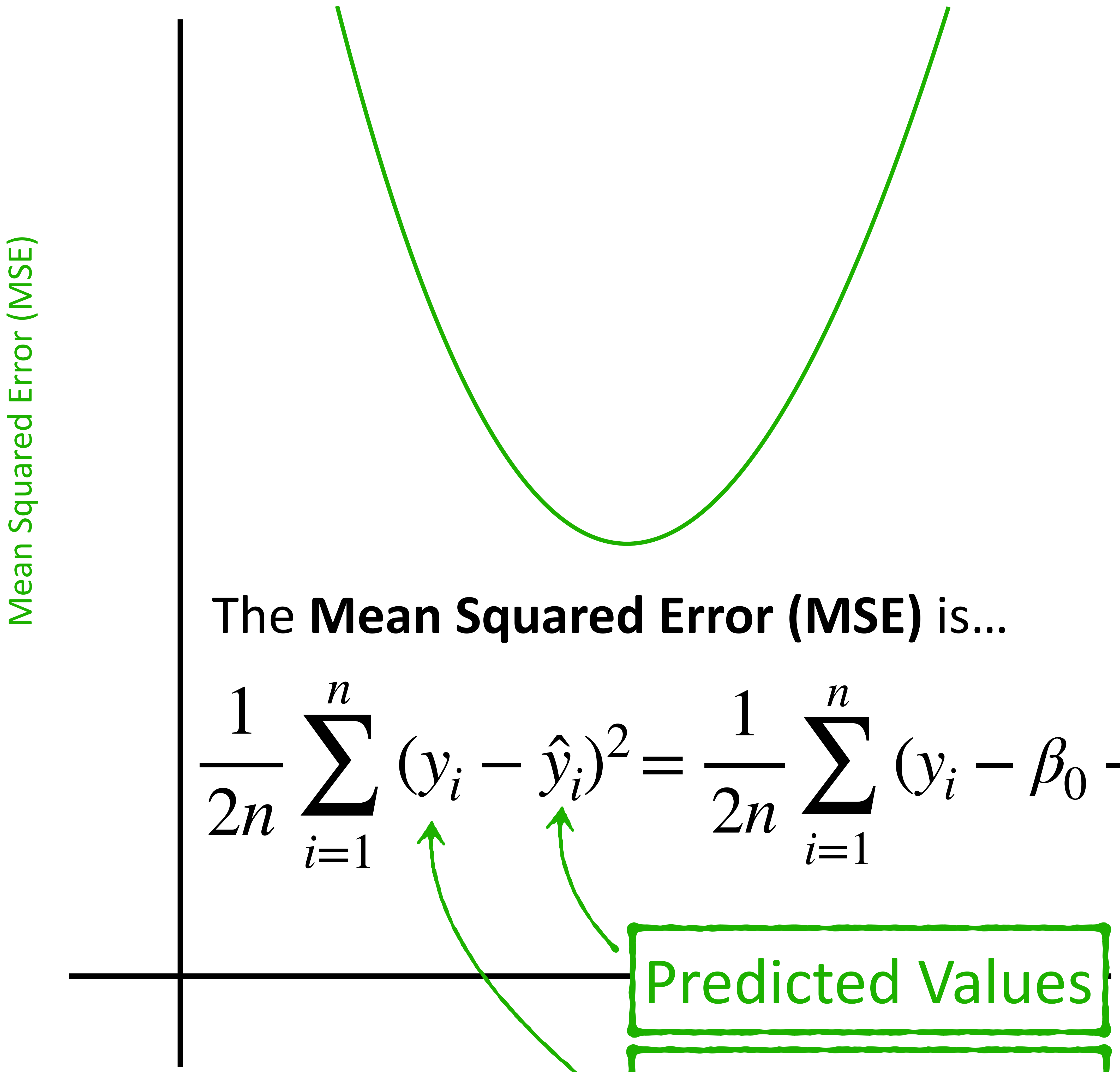
$$\frac{\partial}{\partial\beta_0}\frac{1}{2n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = -\frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)$$

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

This time we will use **Gradient Descent** to find the optimal values of β_0 and β_1

Gradient Descent: Basic Concepts

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



The **Mean Squared Error (MSE)** is…

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

Predicted Values

Observed Values

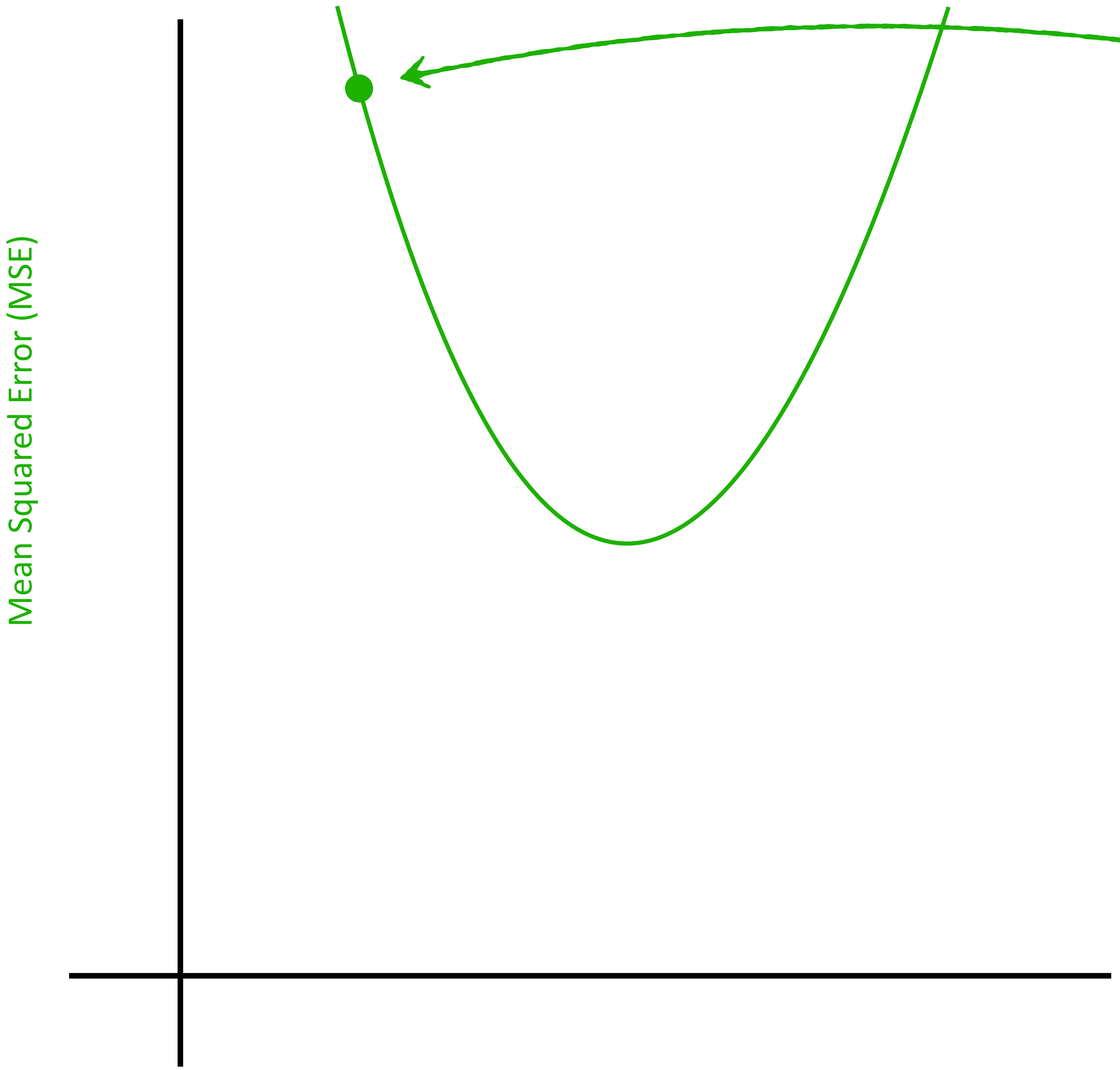
Every point on this curve is the MSE for different values of β_0 and β_1

For any given point on this curve, we can calculate the slope…

… the slope is the first derivative w.r.t β_0 and β_1

Gradient Descent

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



Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

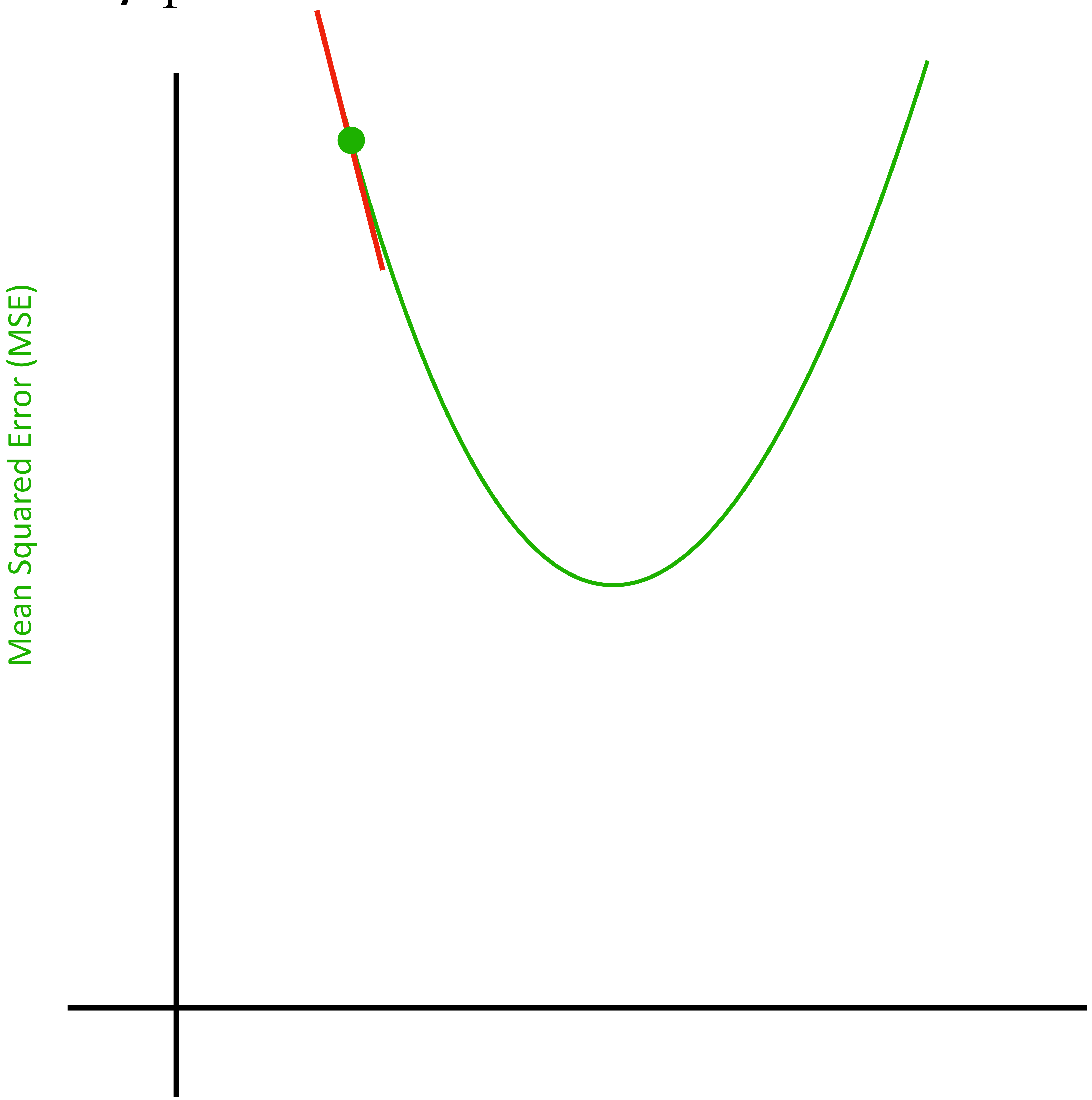
Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

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



Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

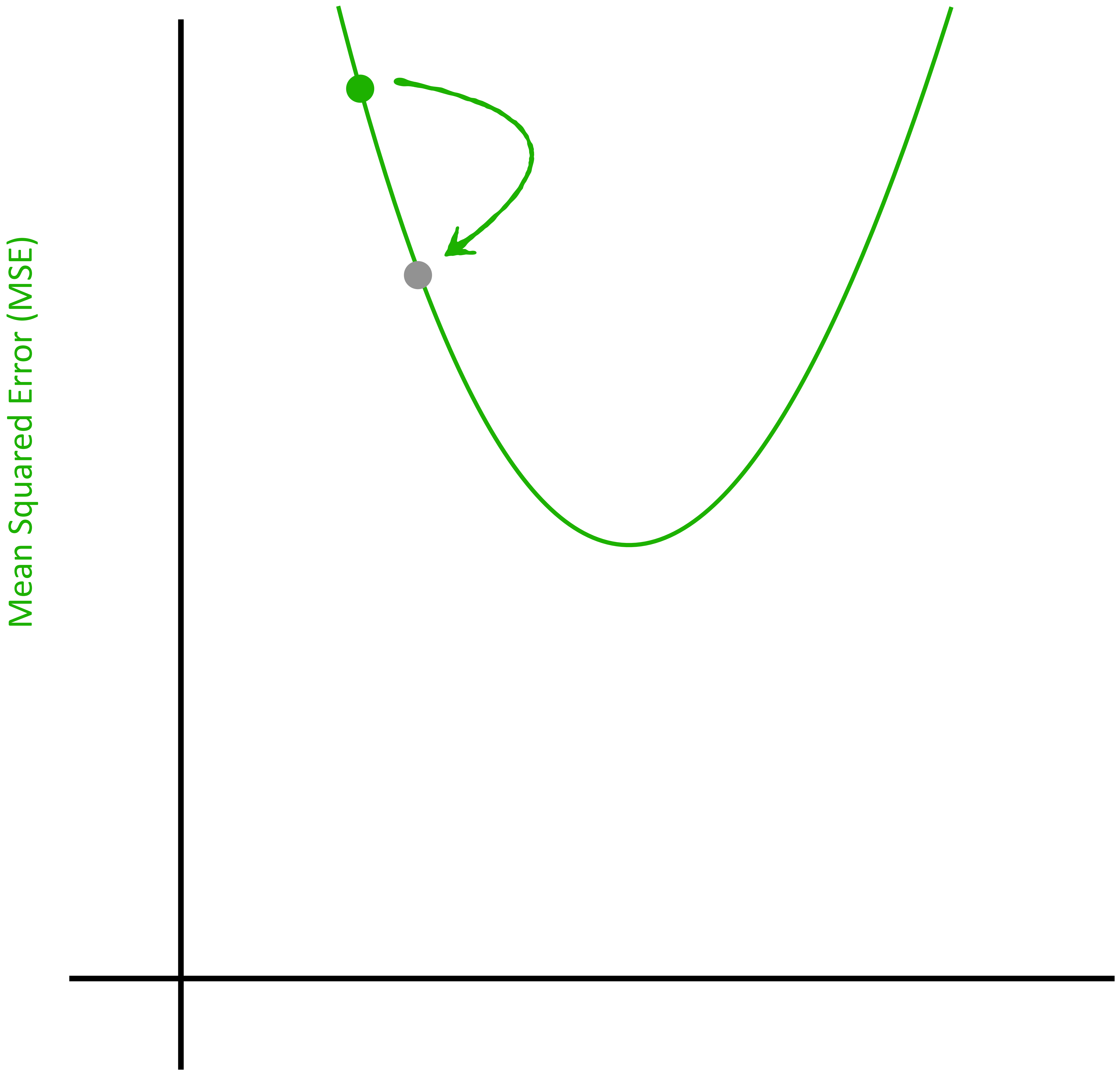
Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

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



Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

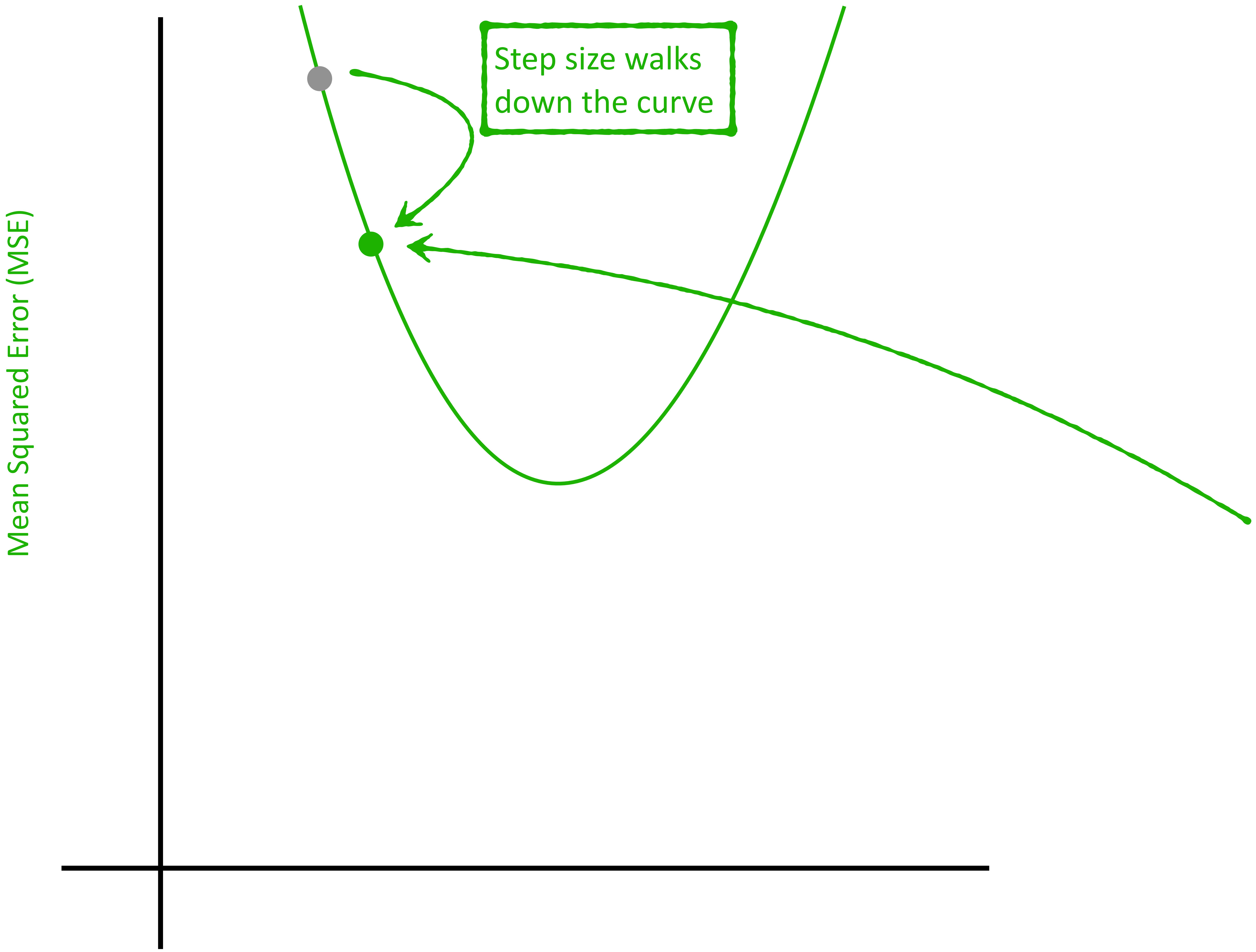
Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

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



Gradient Descent

Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

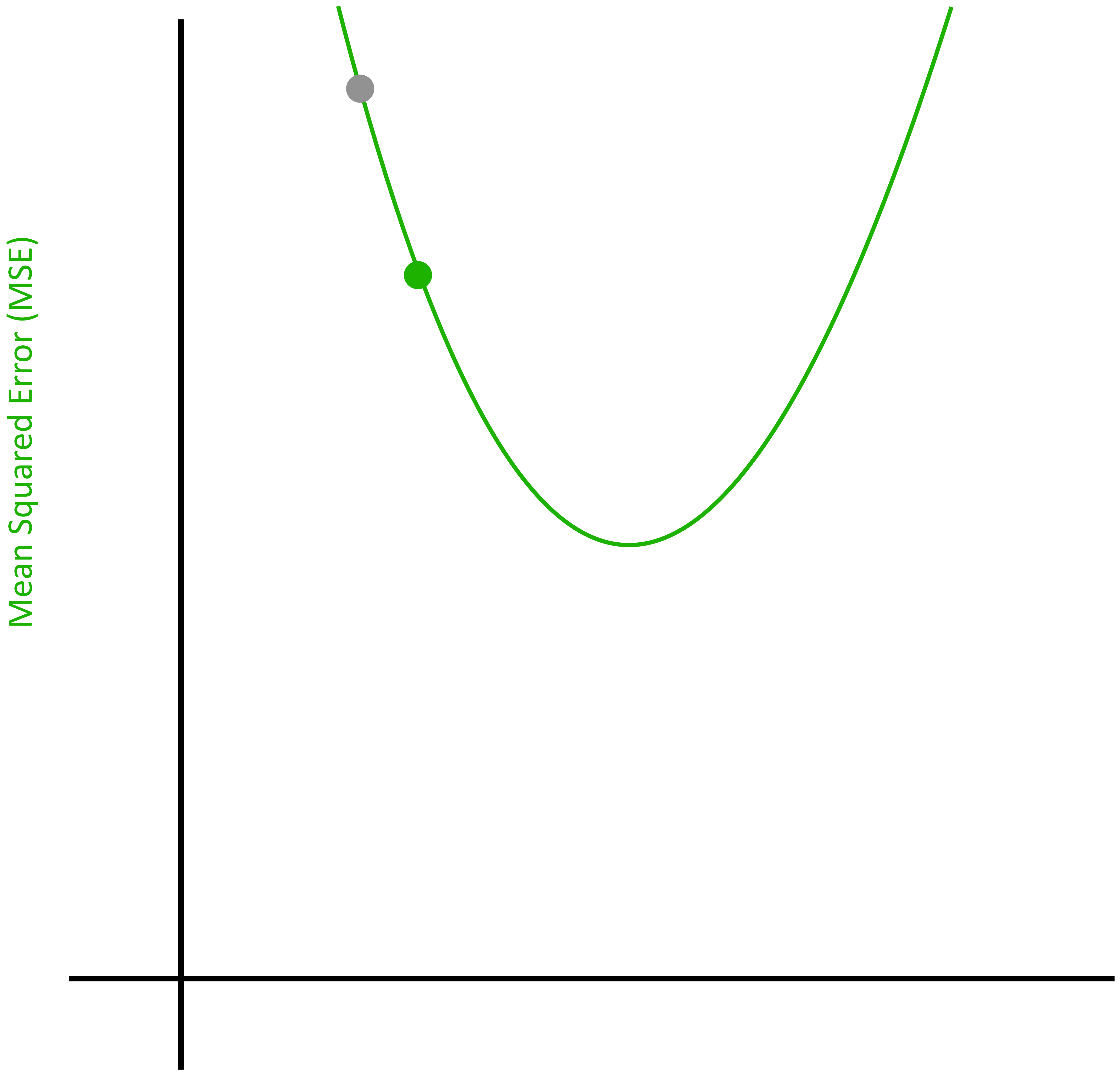
Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

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



Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

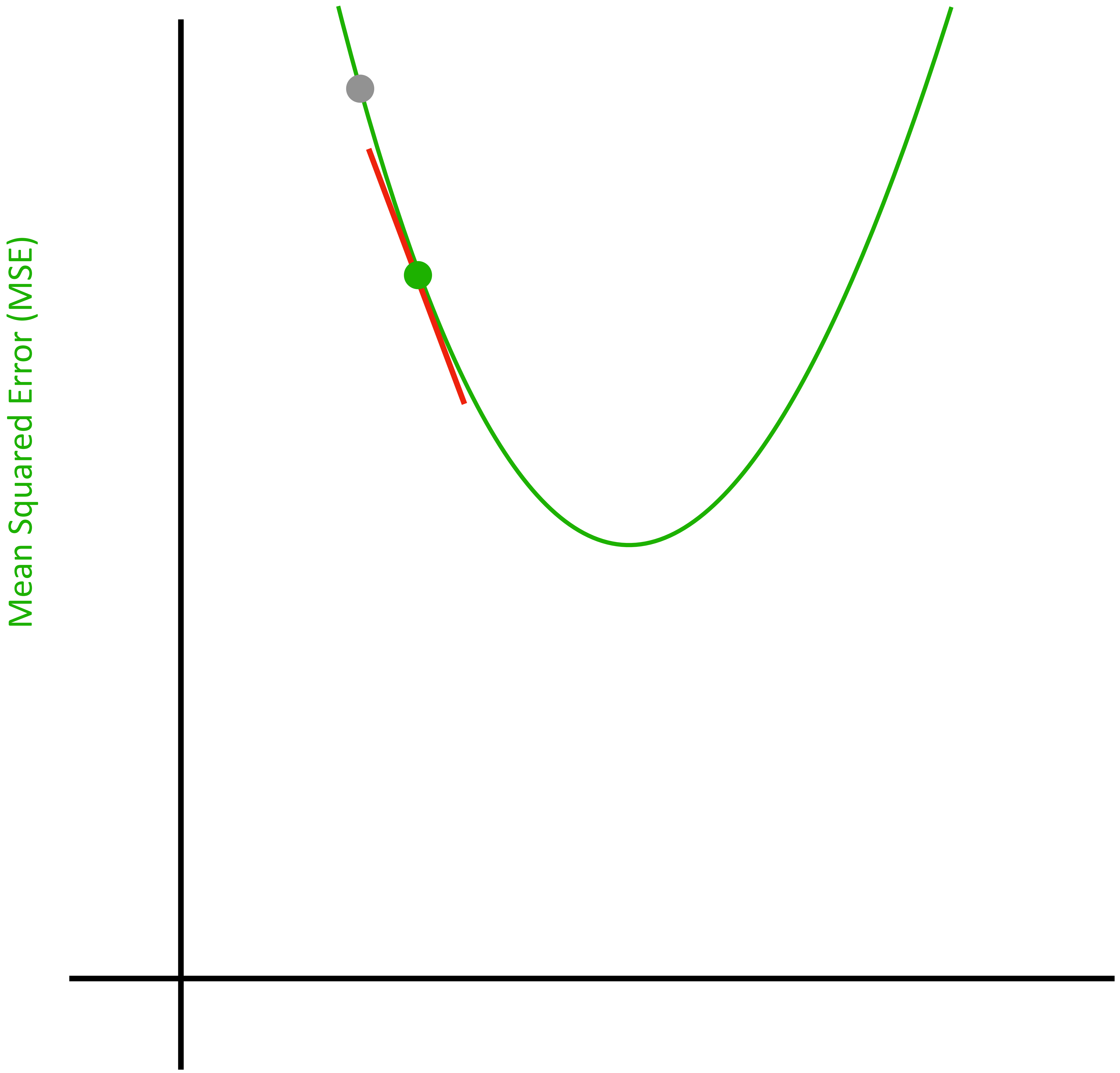
Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

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



Gradient Descent

Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

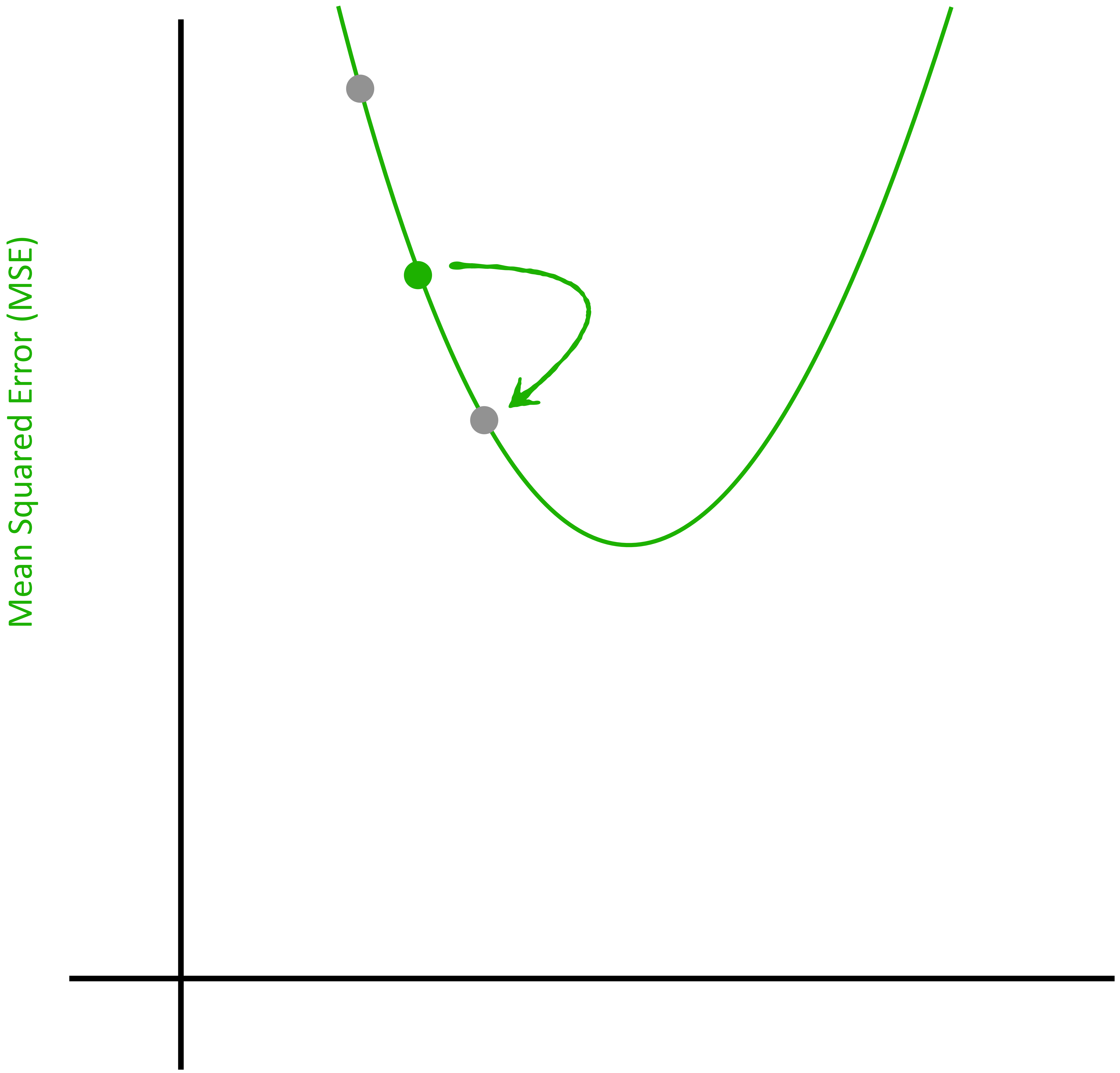
Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

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



Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

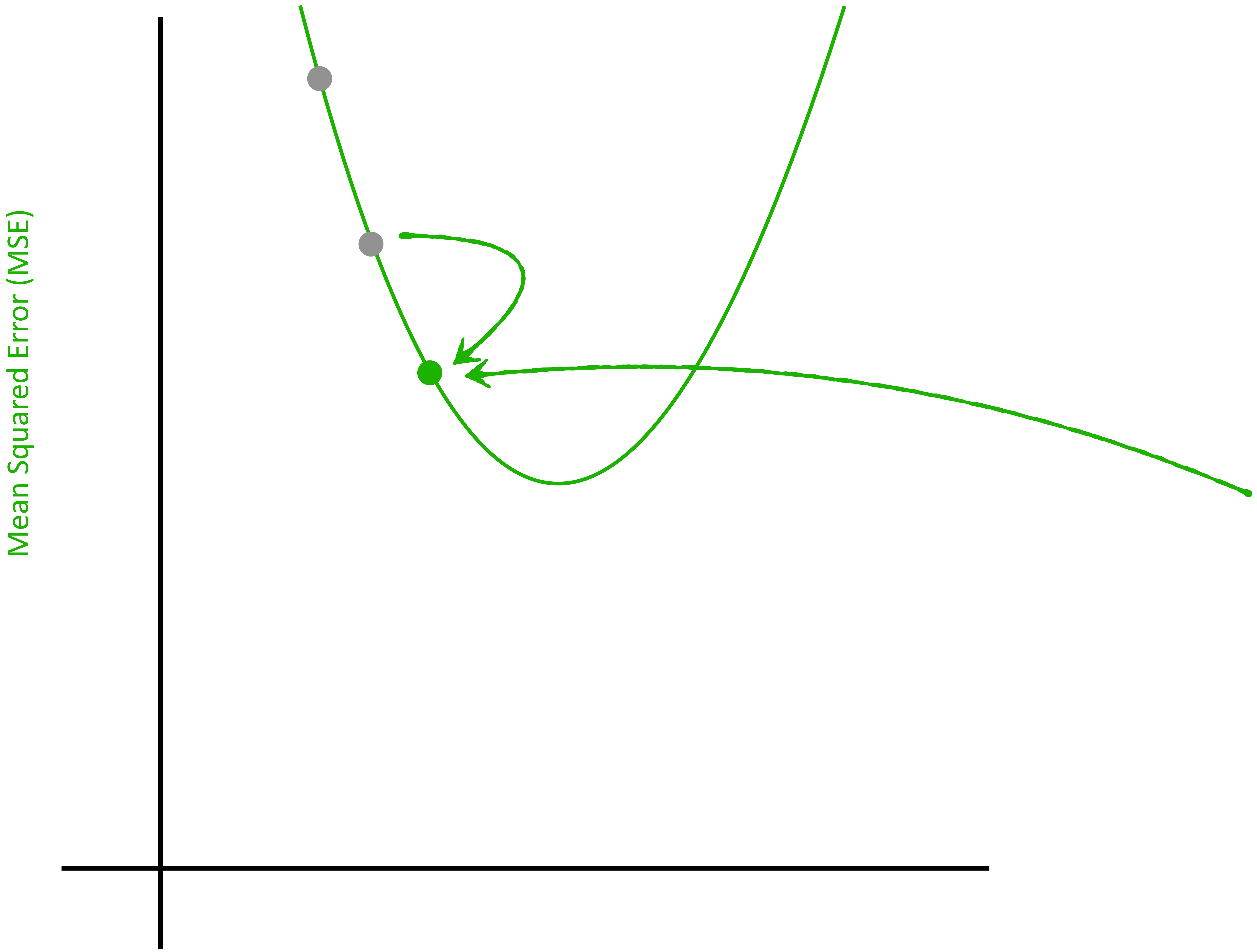
Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

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

Gradient Descent



Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

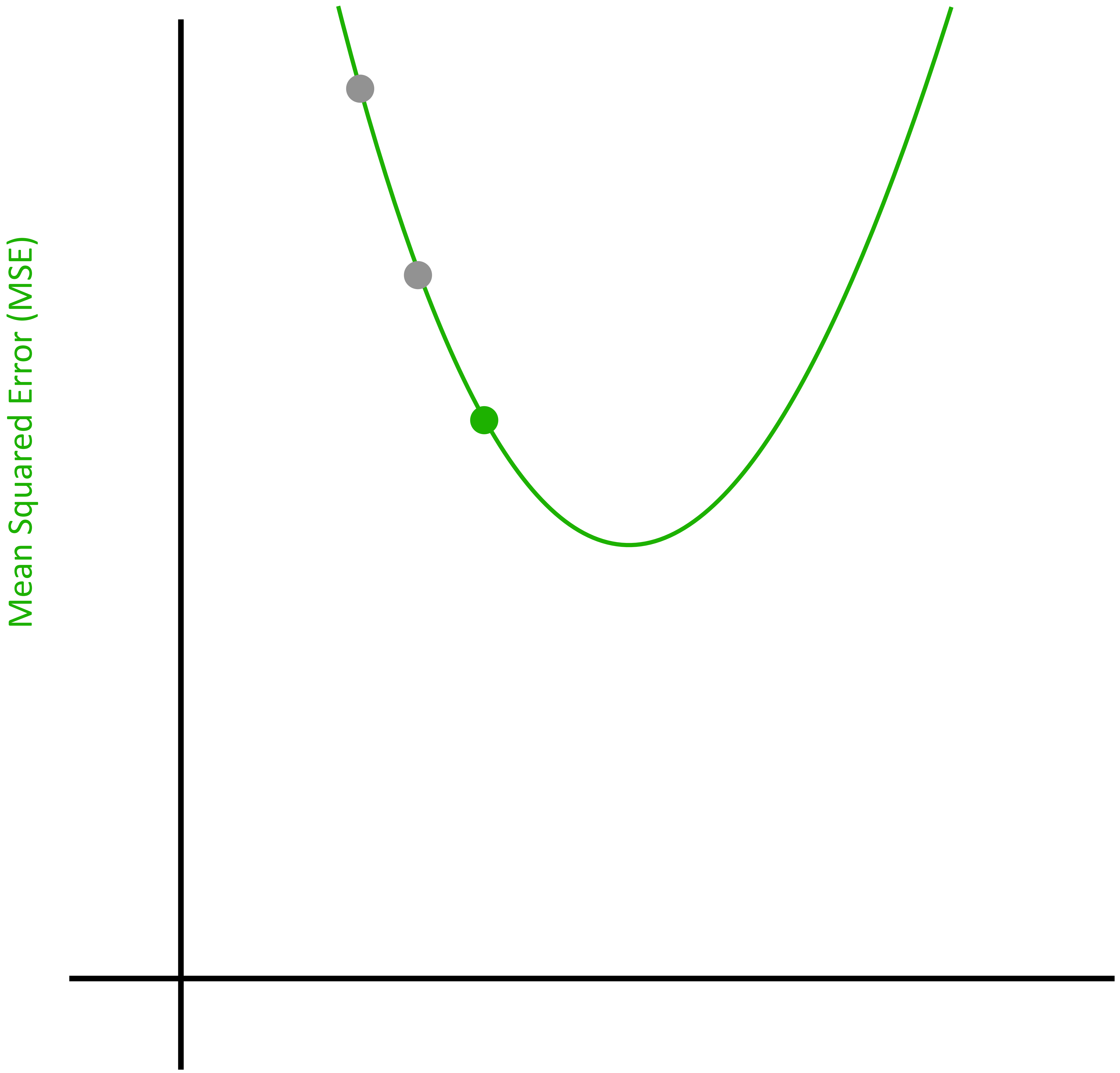
Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

Gradient Descent

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



Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

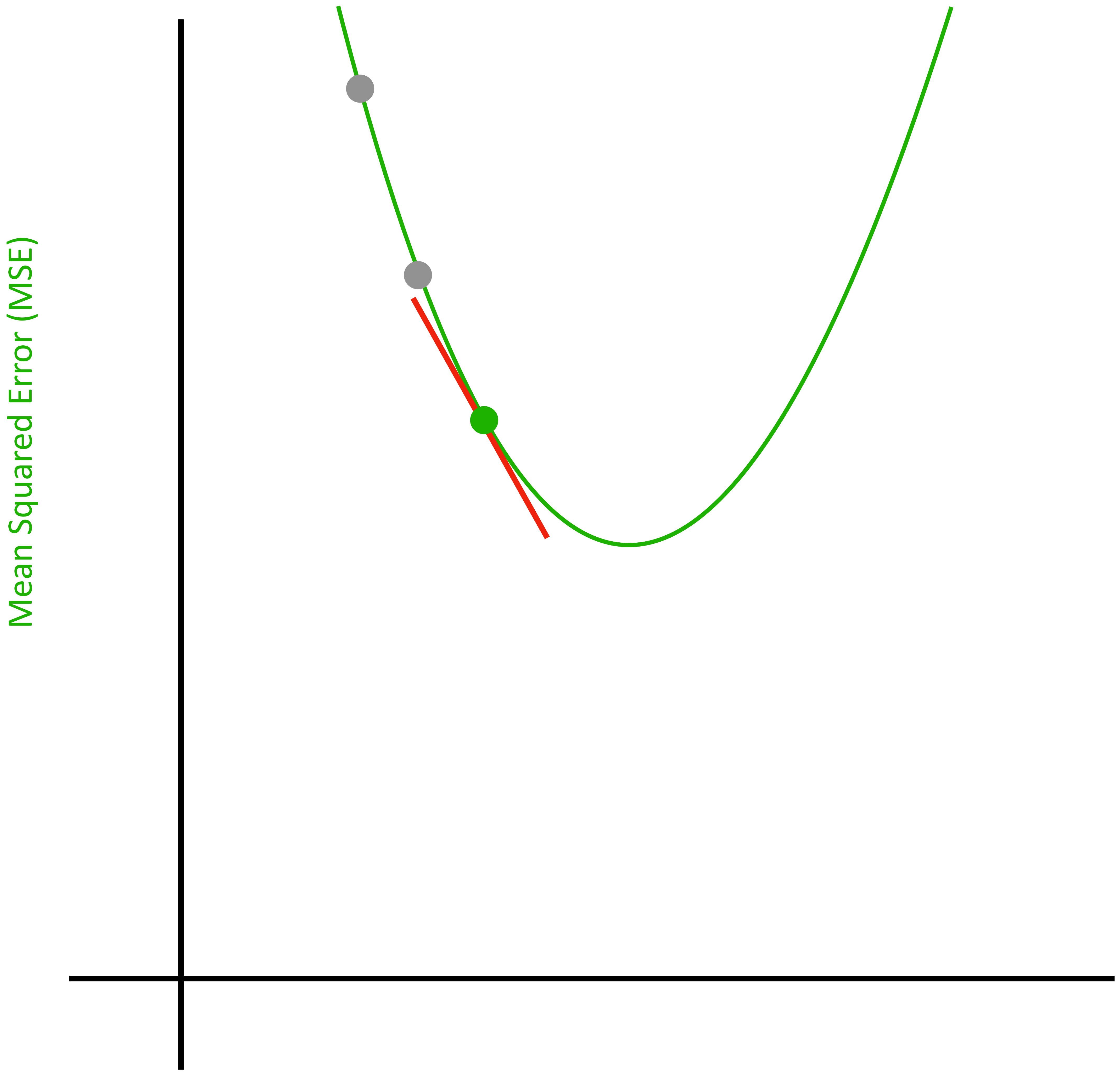
Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

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

Gradient Descent



Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

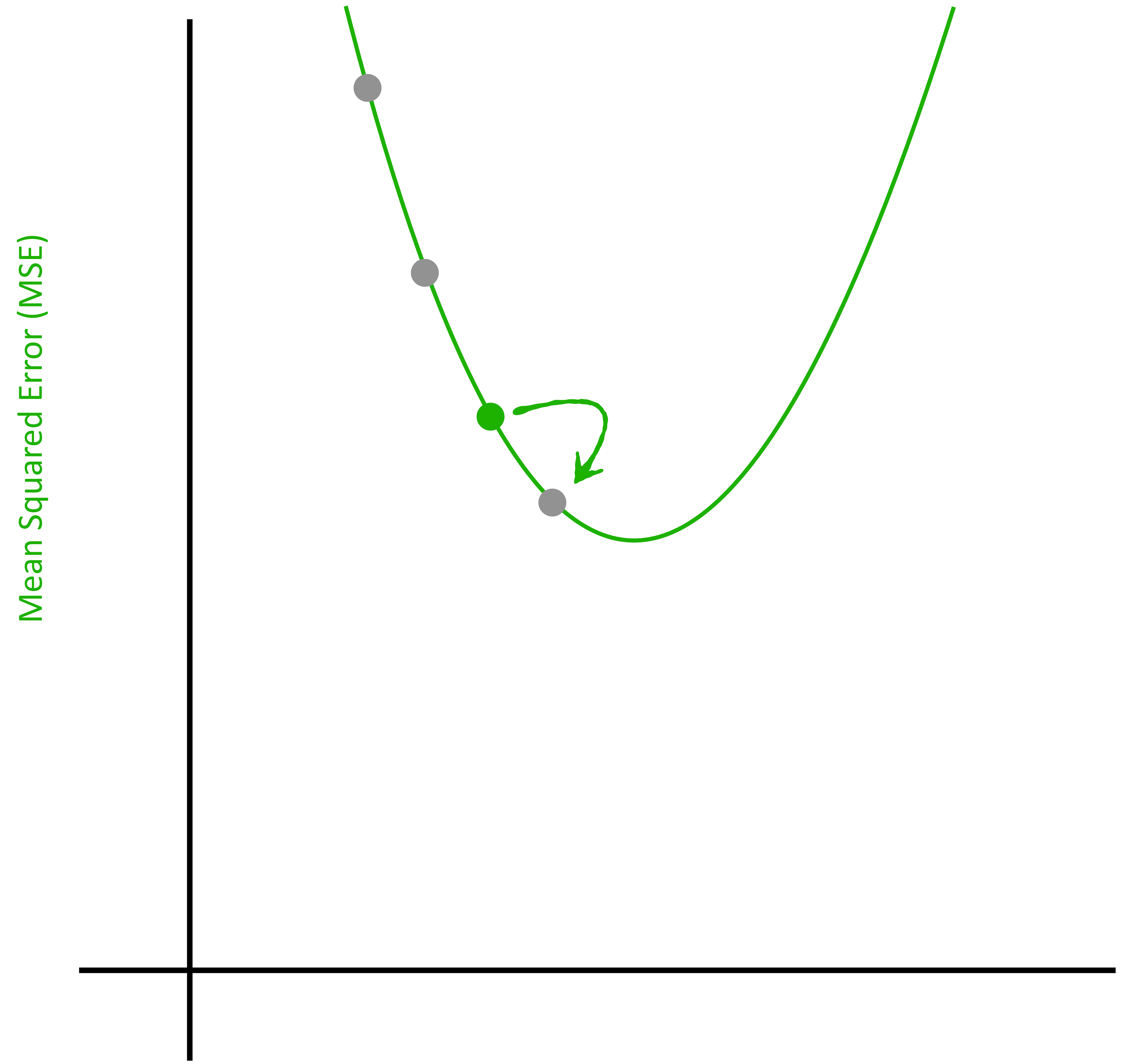
Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

Gradient Descent

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



Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

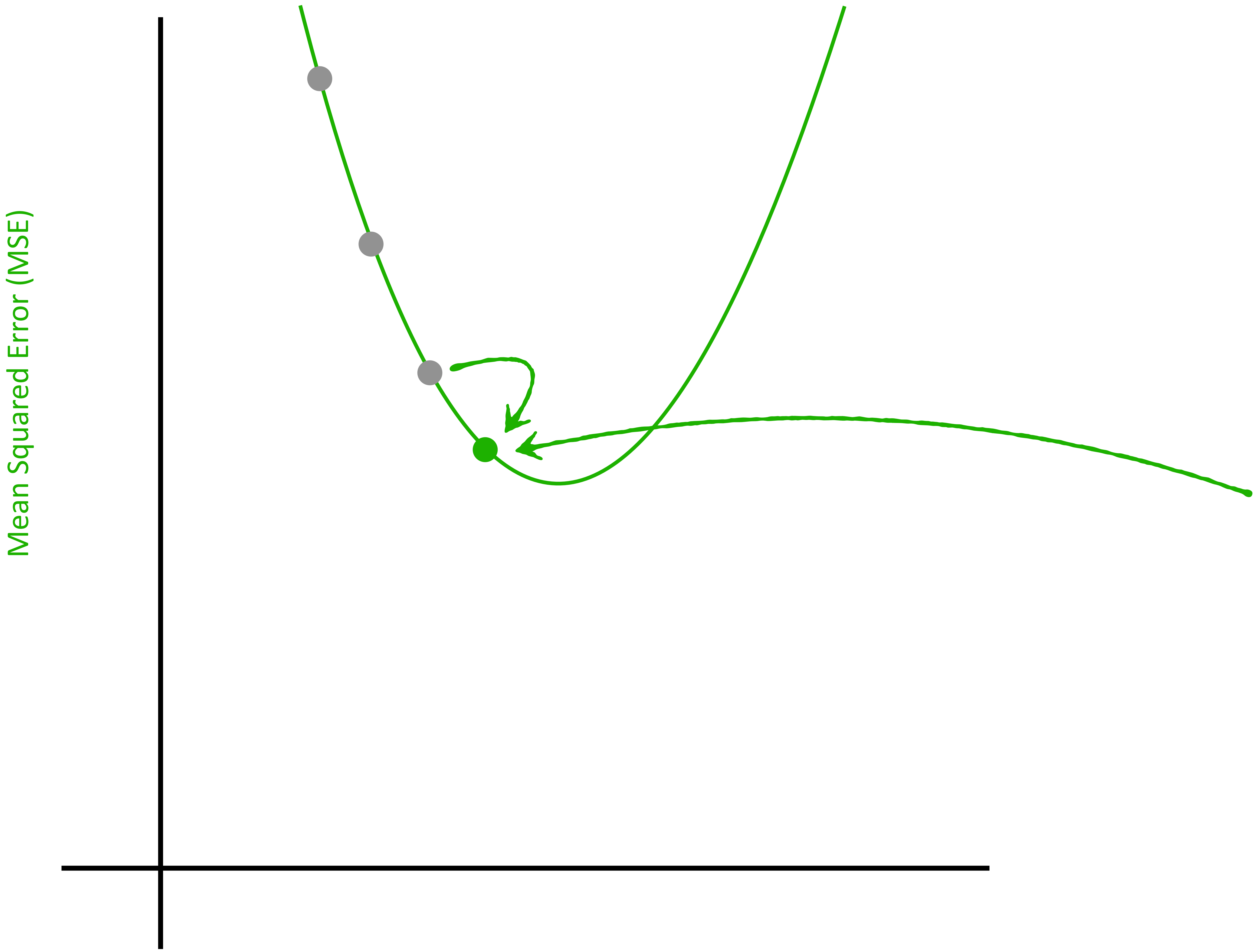
Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

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

Gradient Descent



Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

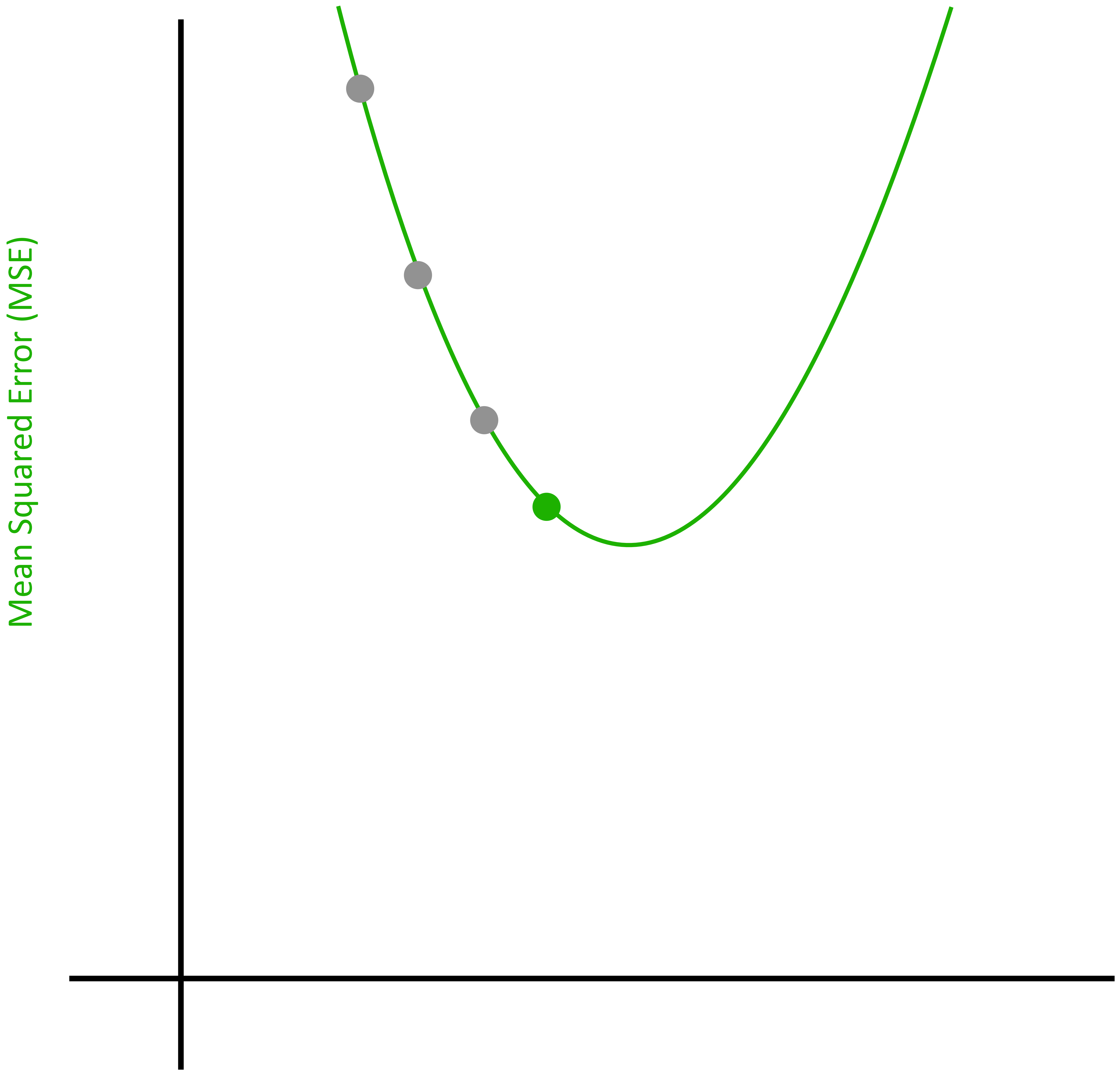
Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

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



Gradient Descent

Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

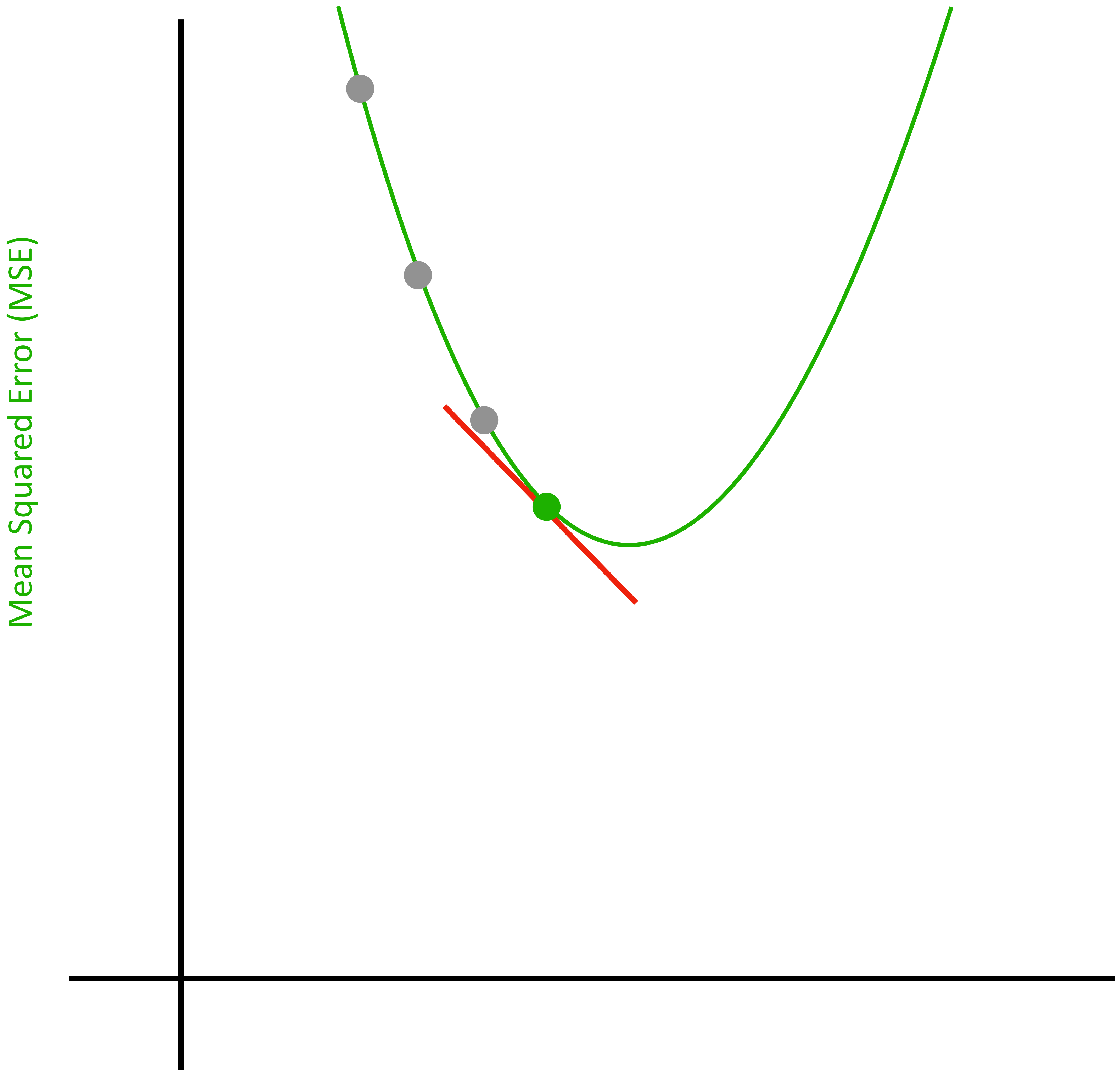
Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

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



Gradient Descent

Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

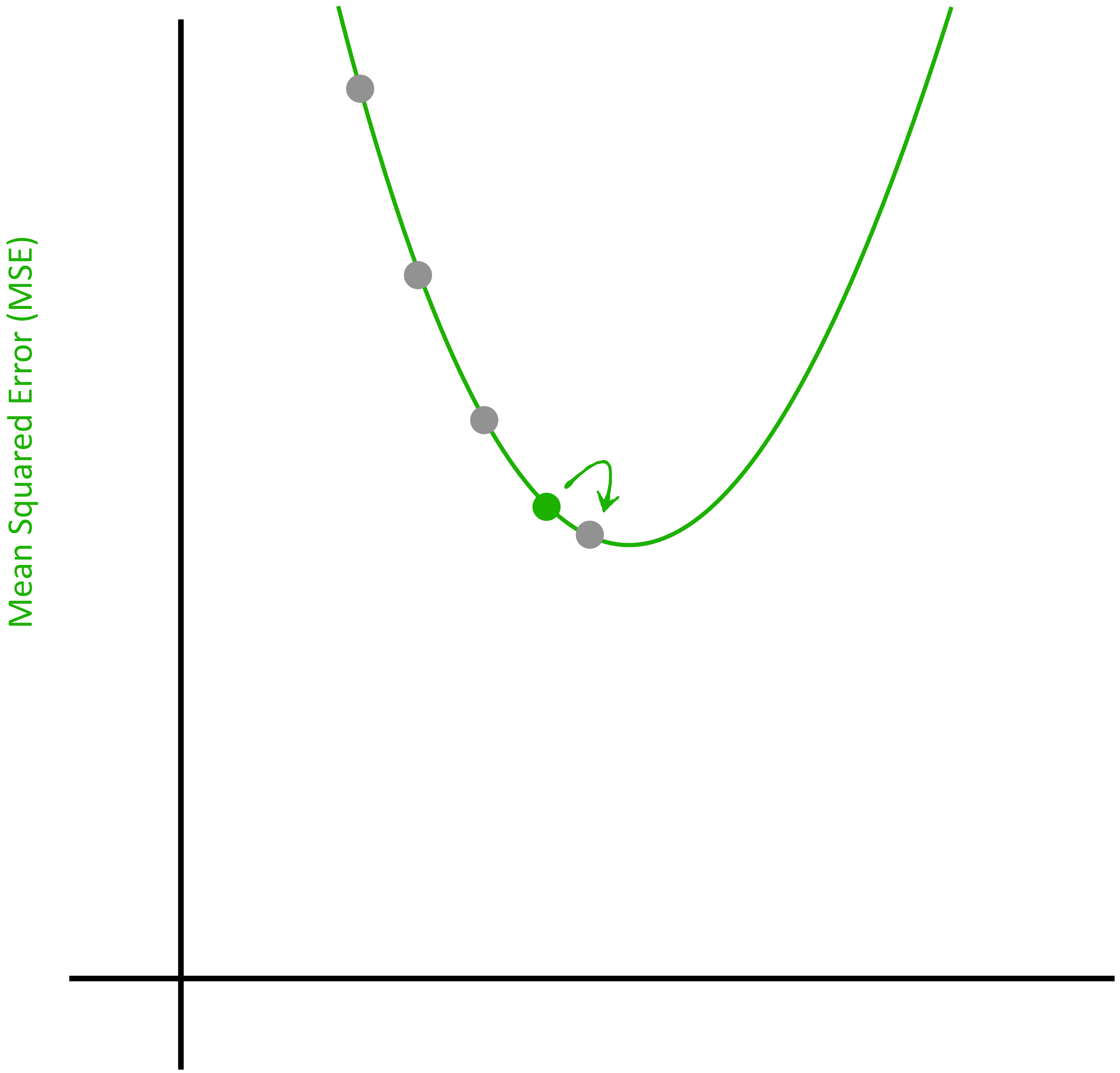
Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

Gradient Descent

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



Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

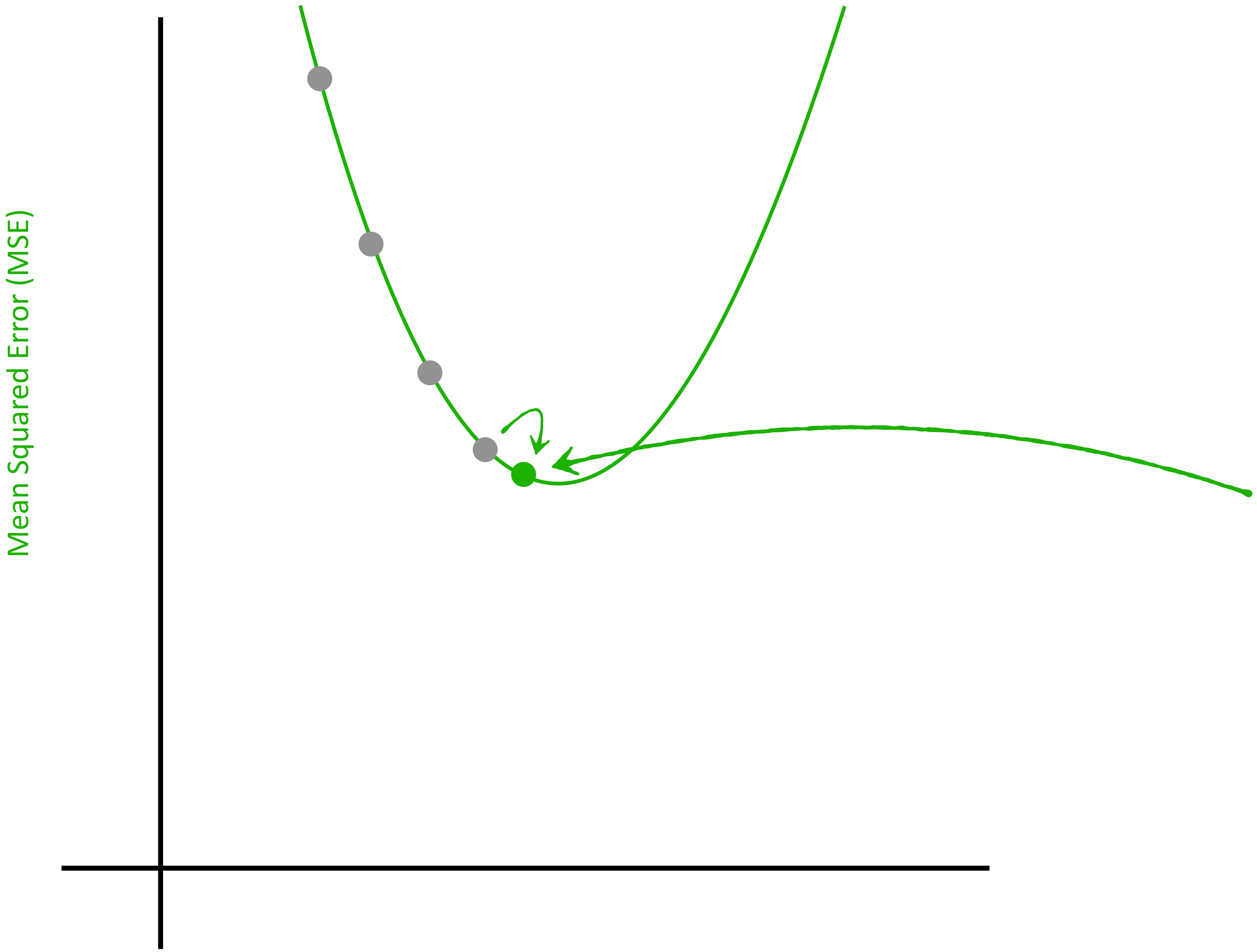
Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

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

Gradient Descent



Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

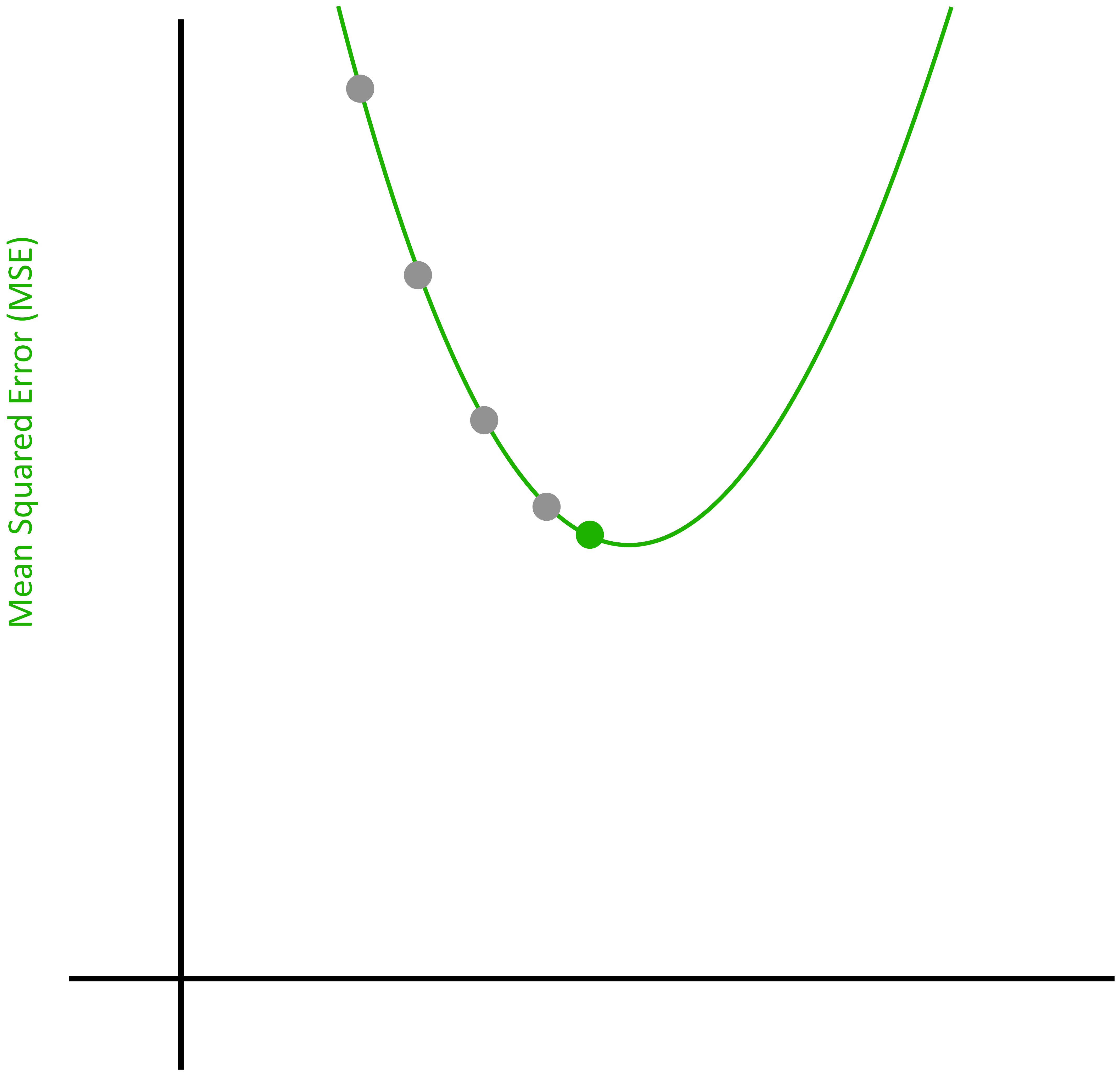
Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

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

Gradient Descent



Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

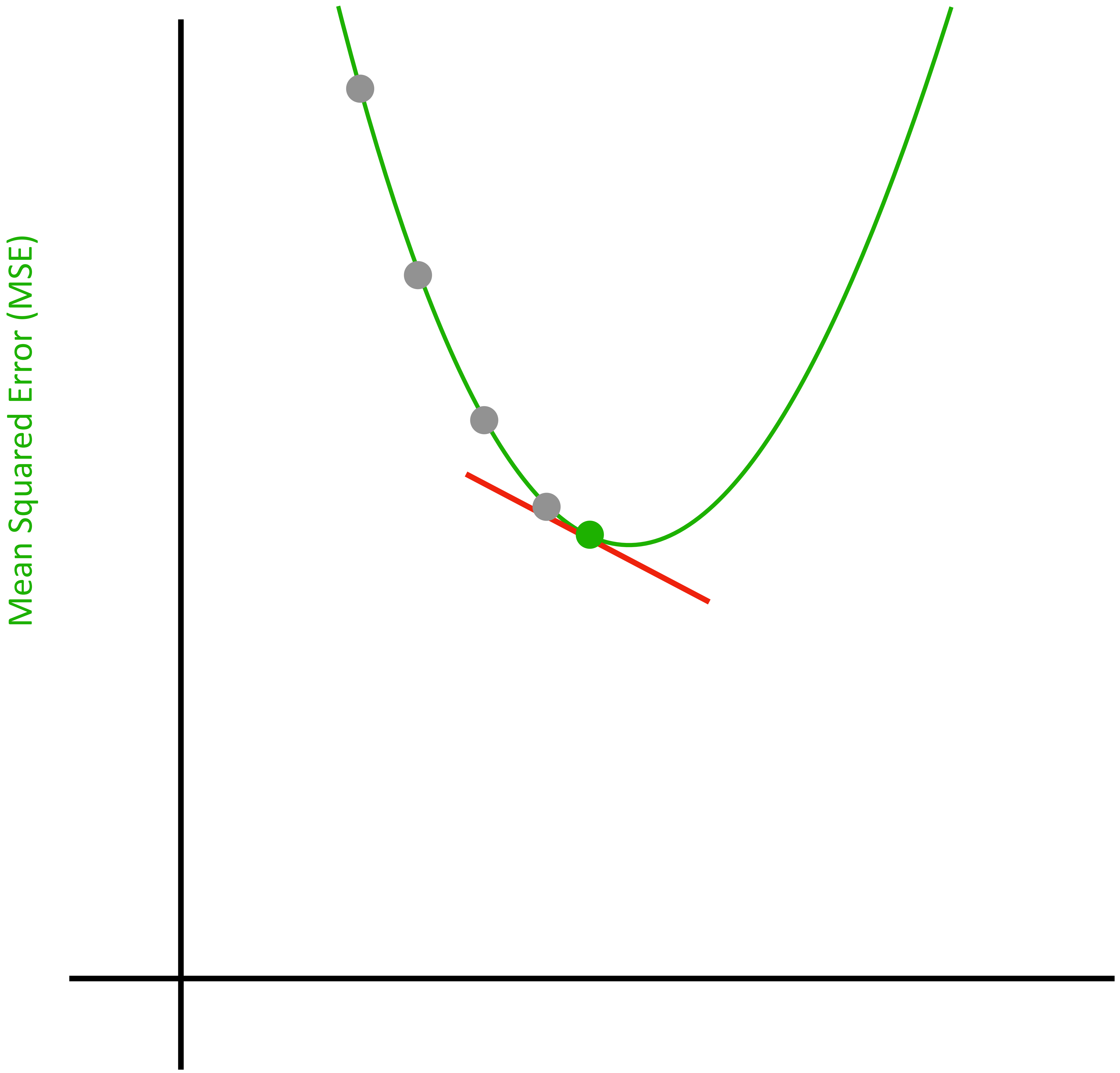
Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

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



Gradient Descent

Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

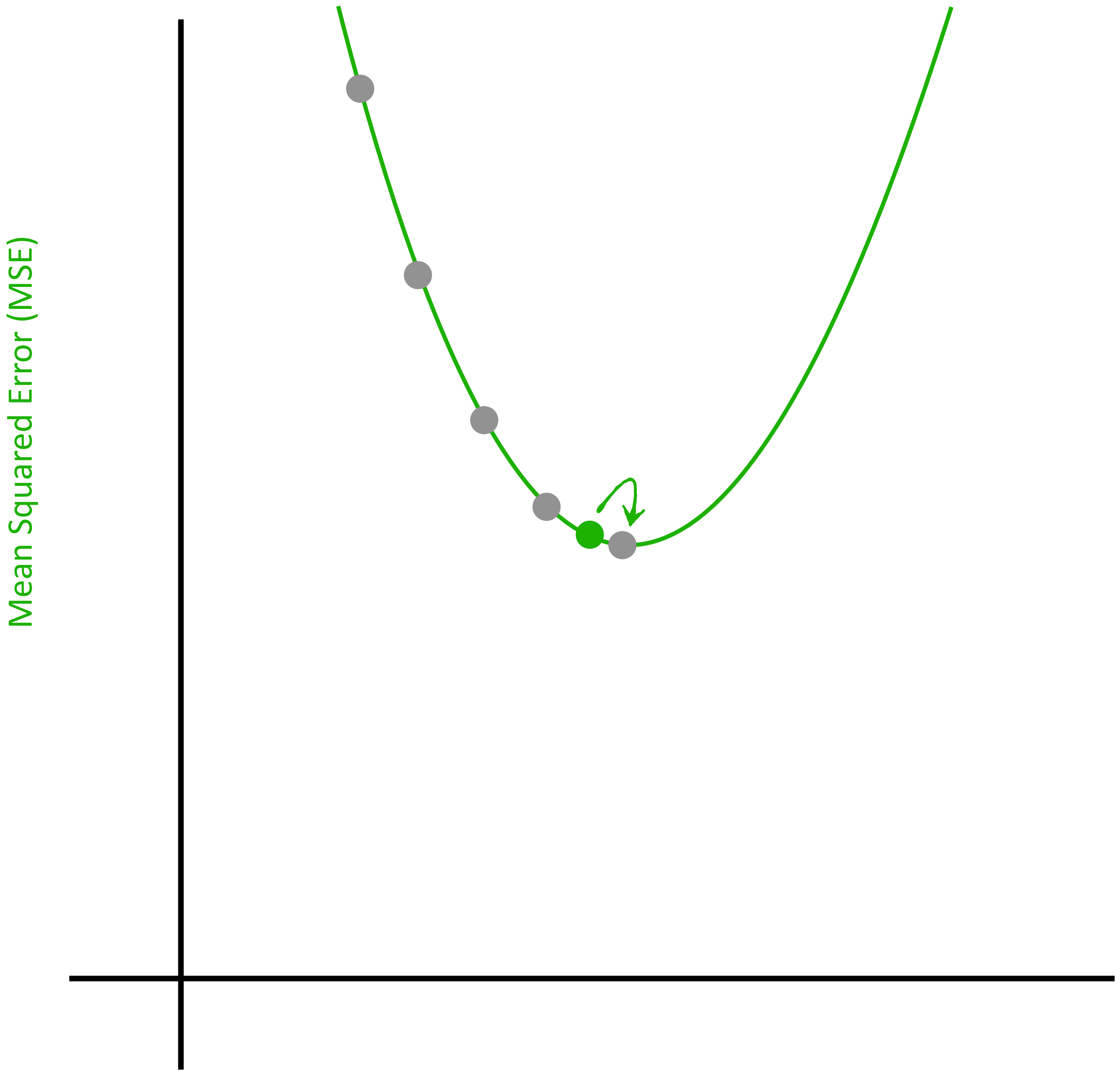
Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

Gradient Descent

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



Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

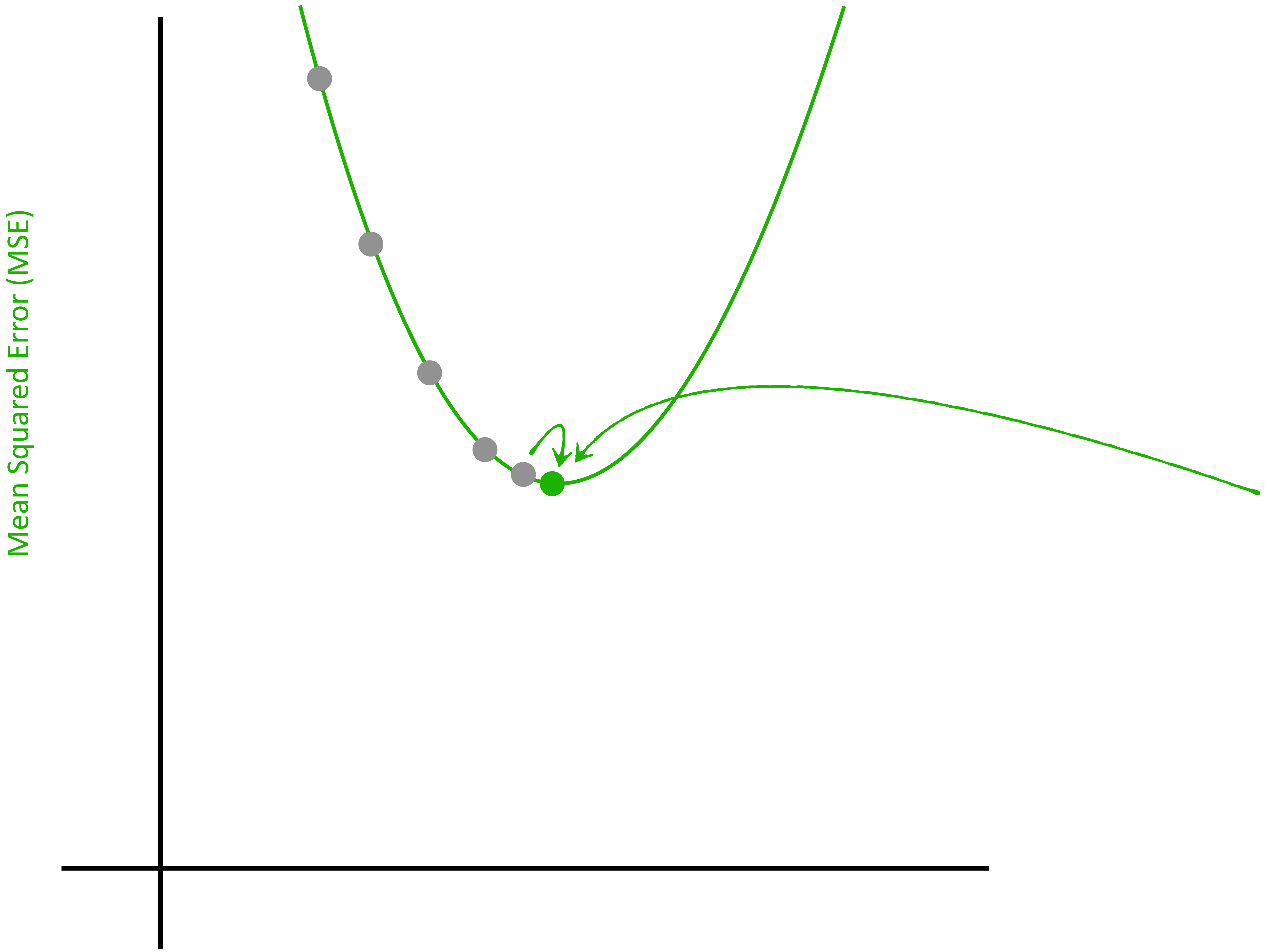
Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

Gradient Descent

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



Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

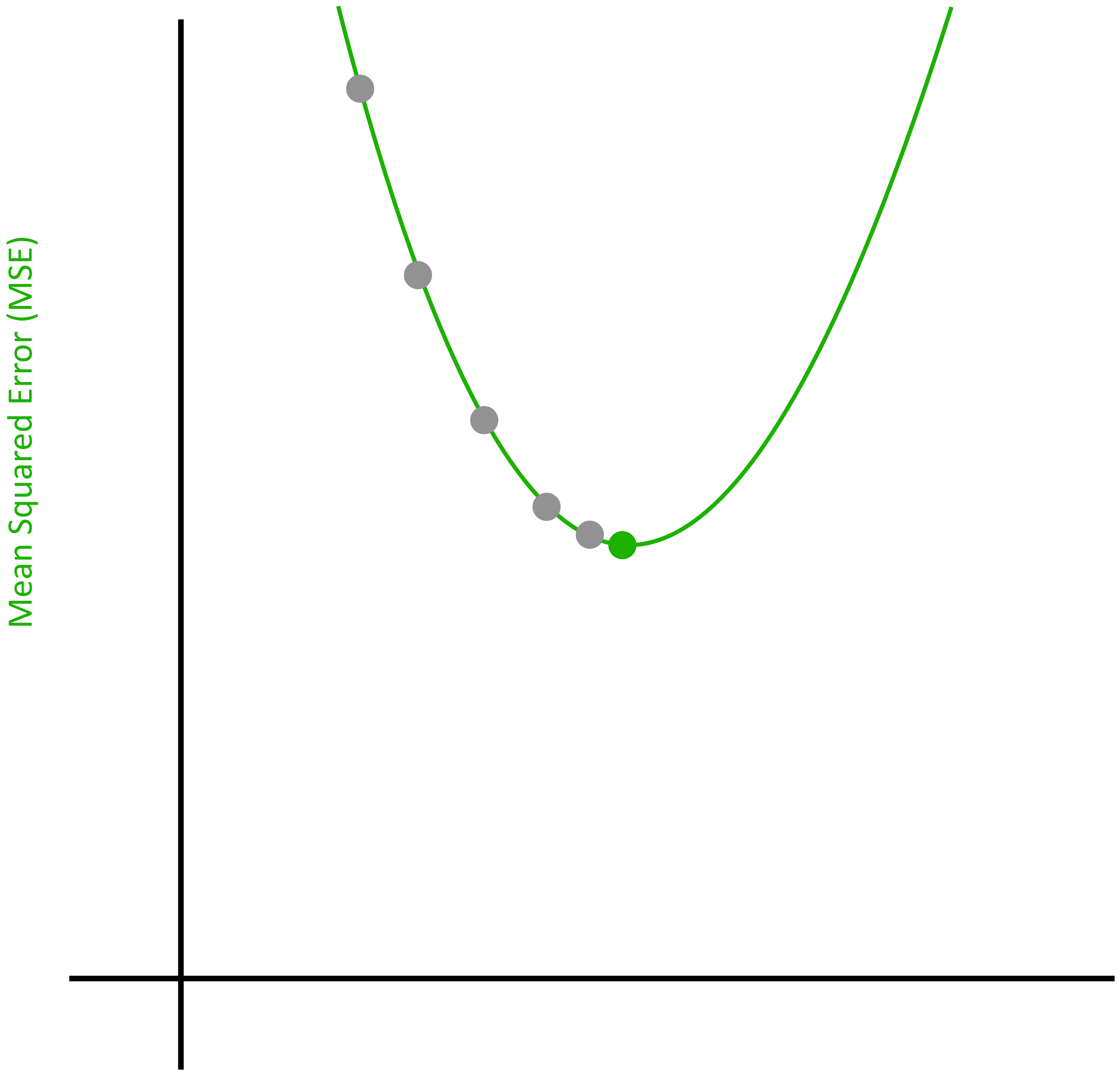
Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

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



Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

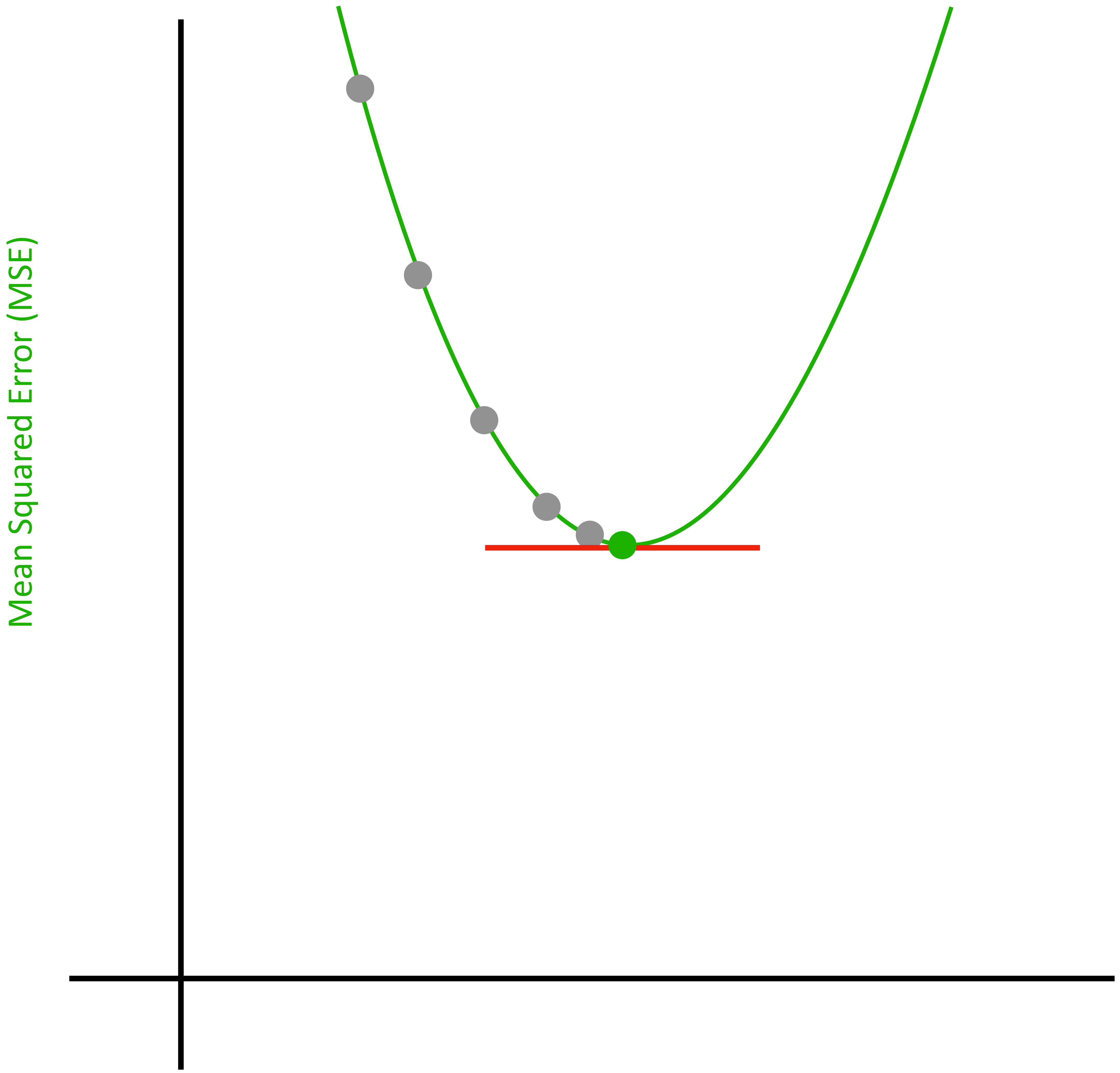
Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

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



Gradient Descent

Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

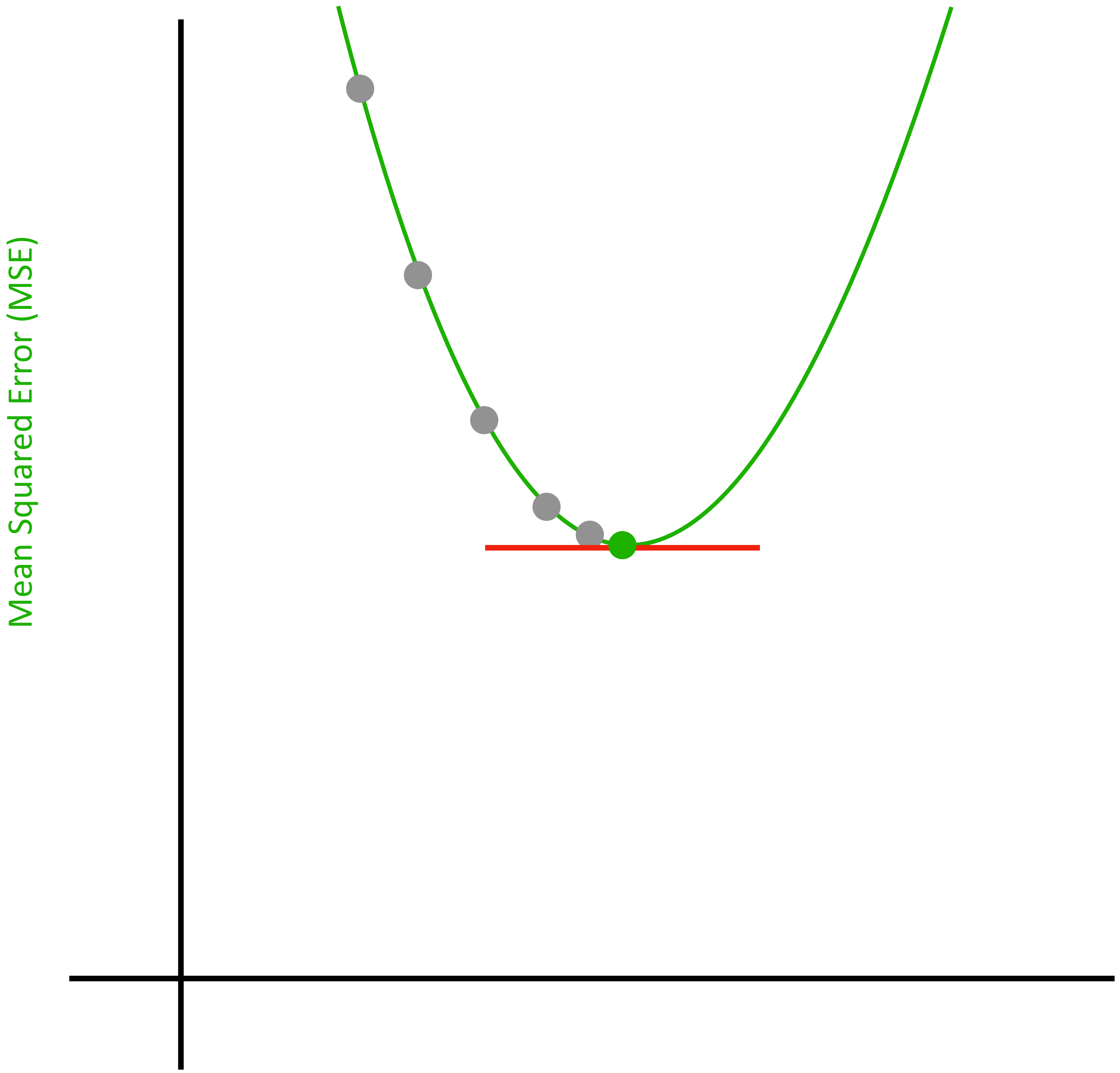
Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

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



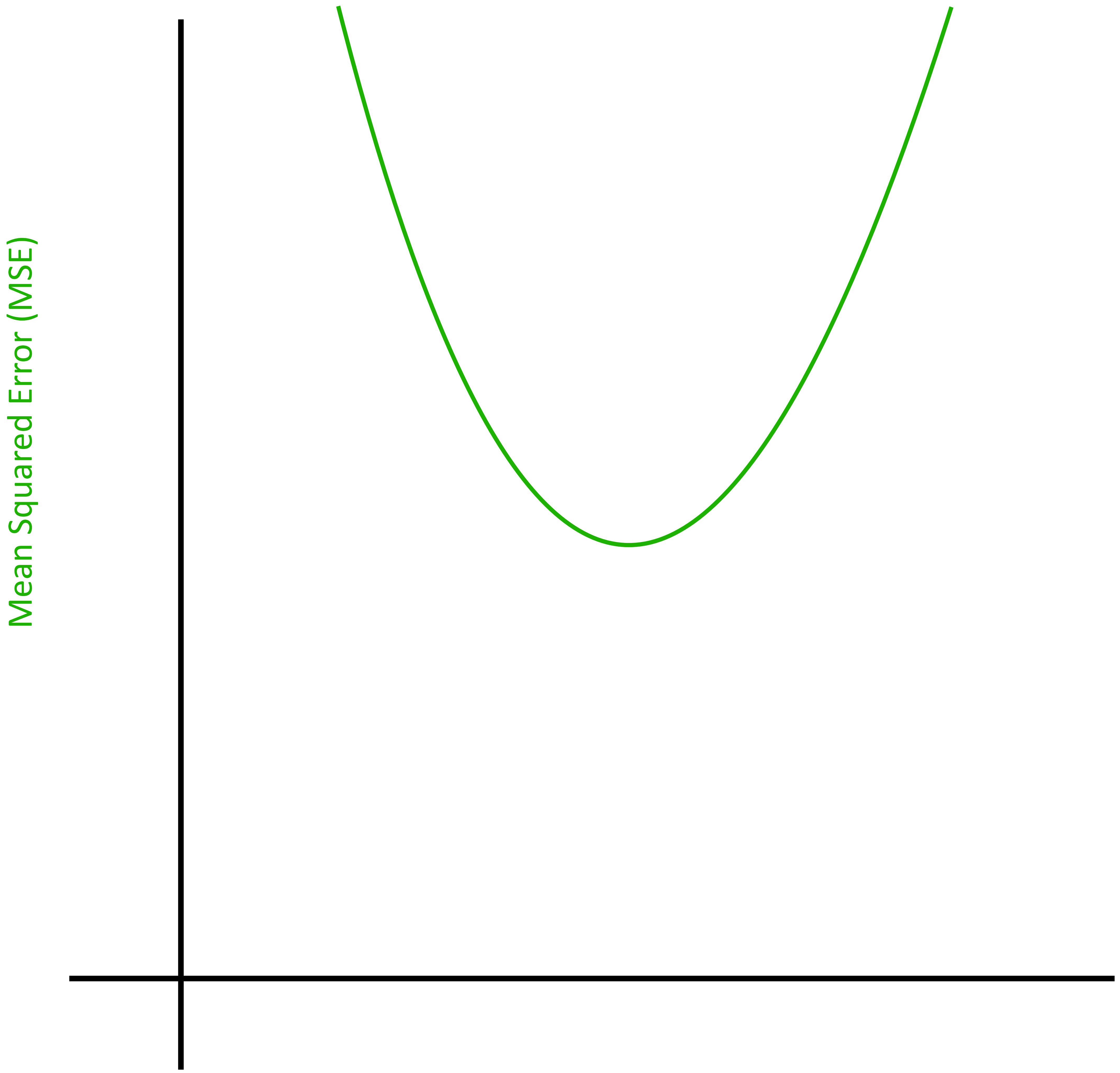
Gradient Descent

Gradient Descent: Basic Concept

Gradient Descent continues in this manner until the step size is close to zero or a fixed number of iterations

Gradient Descent: Lets walk through the algorithm

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



Gradient Descent: Basic Concept

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

Mean Squared Error (MSE)

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

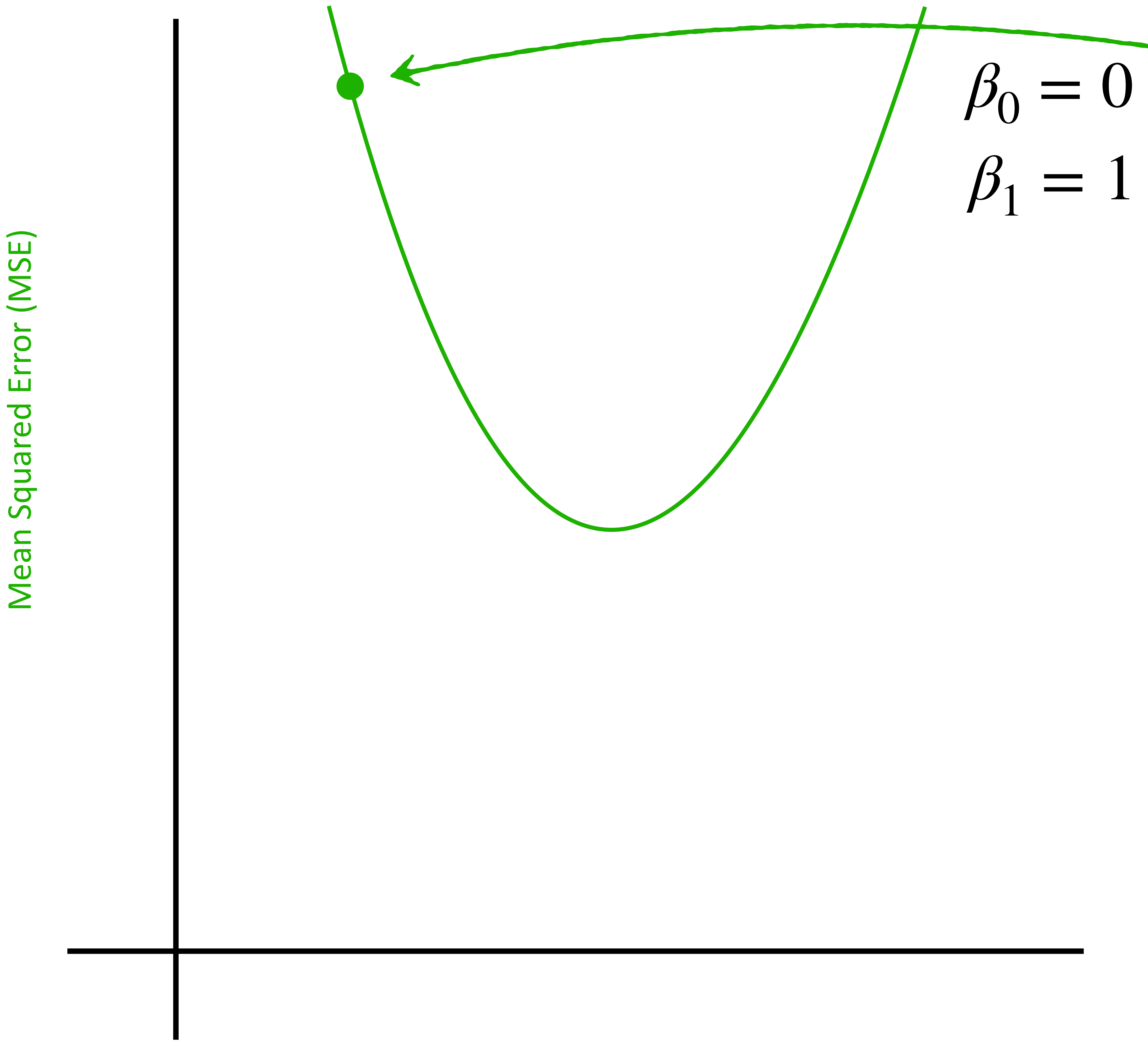
The **first derivative w.r.t** β_0 **and** β_1 is…

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

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

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

Gradient Descent



Gradient Descent: Algorithm

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

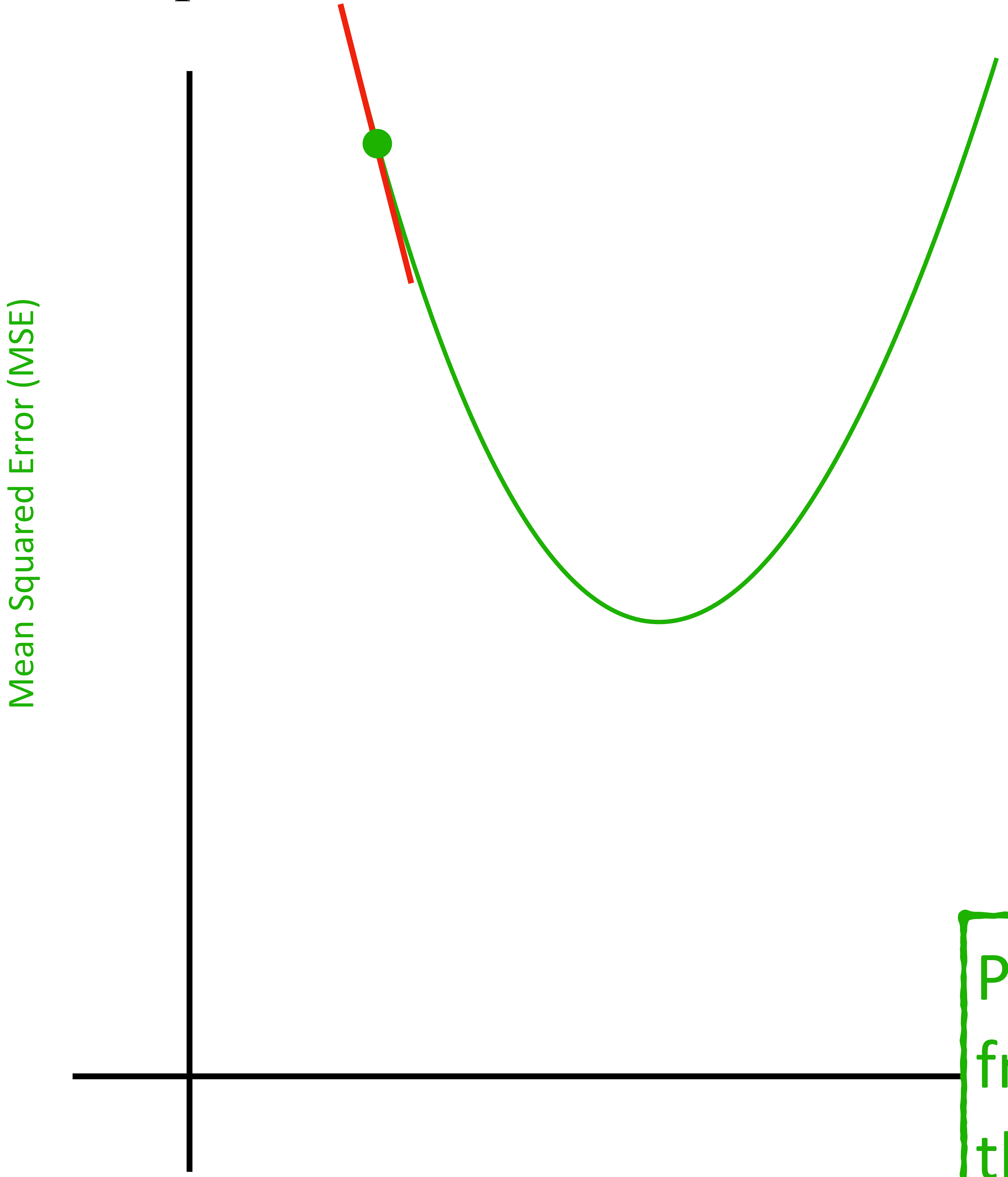
Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

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

Gradient Descent



$\beta_0 = 0$

$\beta_1 = 1$

Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

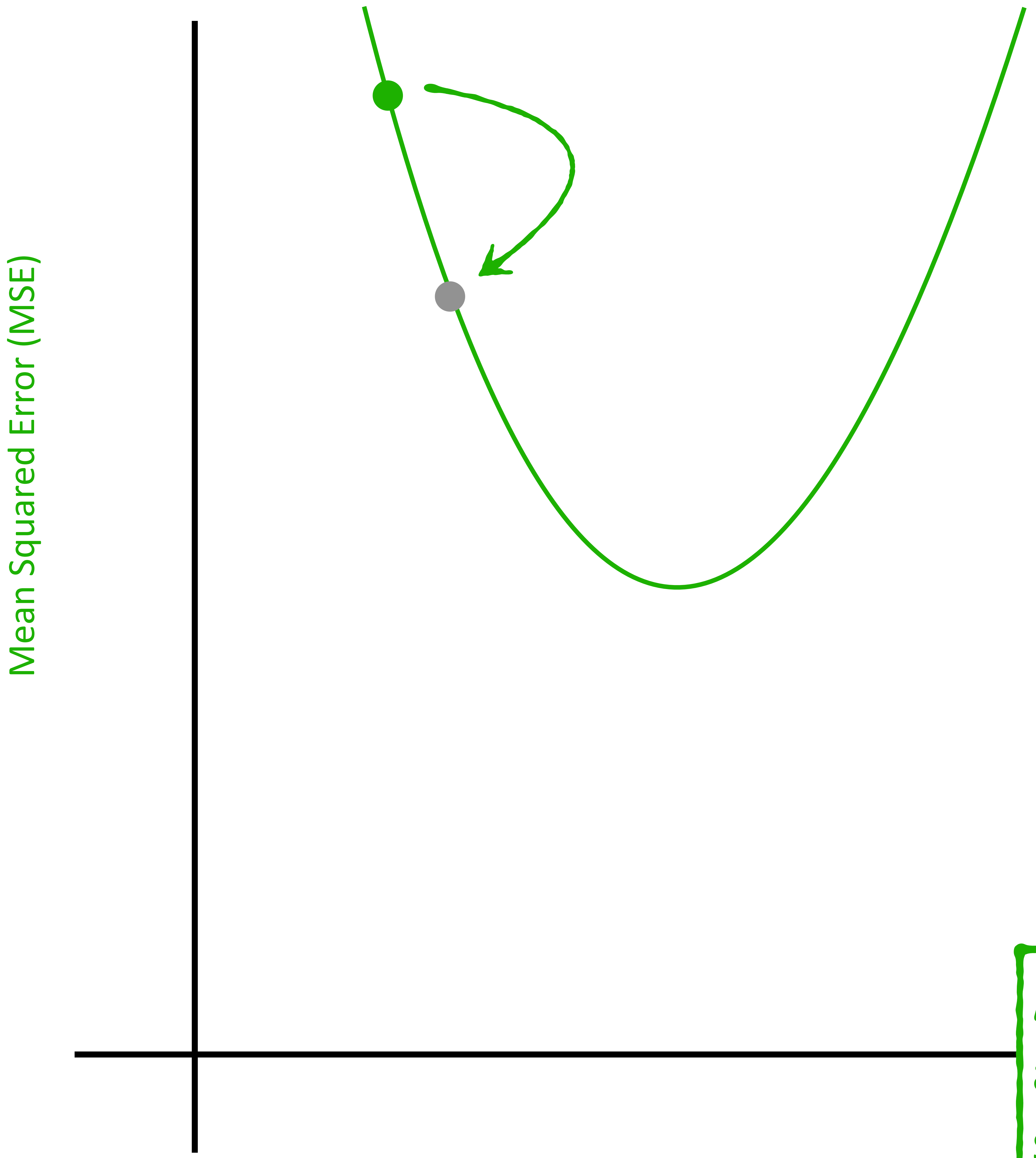
$$\frac{\partial}{\beta_0}MSE = \frac{\partial}{\partial \beta_0}\frac{1}{2n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = -\frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)$$

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

Plugin the values of x_i and y_i from the observations and the values of β_0 and β_1

i	Height (in)	Weight (lbs)
1	62	138
2	55	178
3	44	123
4	75	200
5	65	229
6	50	102

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



$\beta_0 = 0$
$\beta_1 = 1$

Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

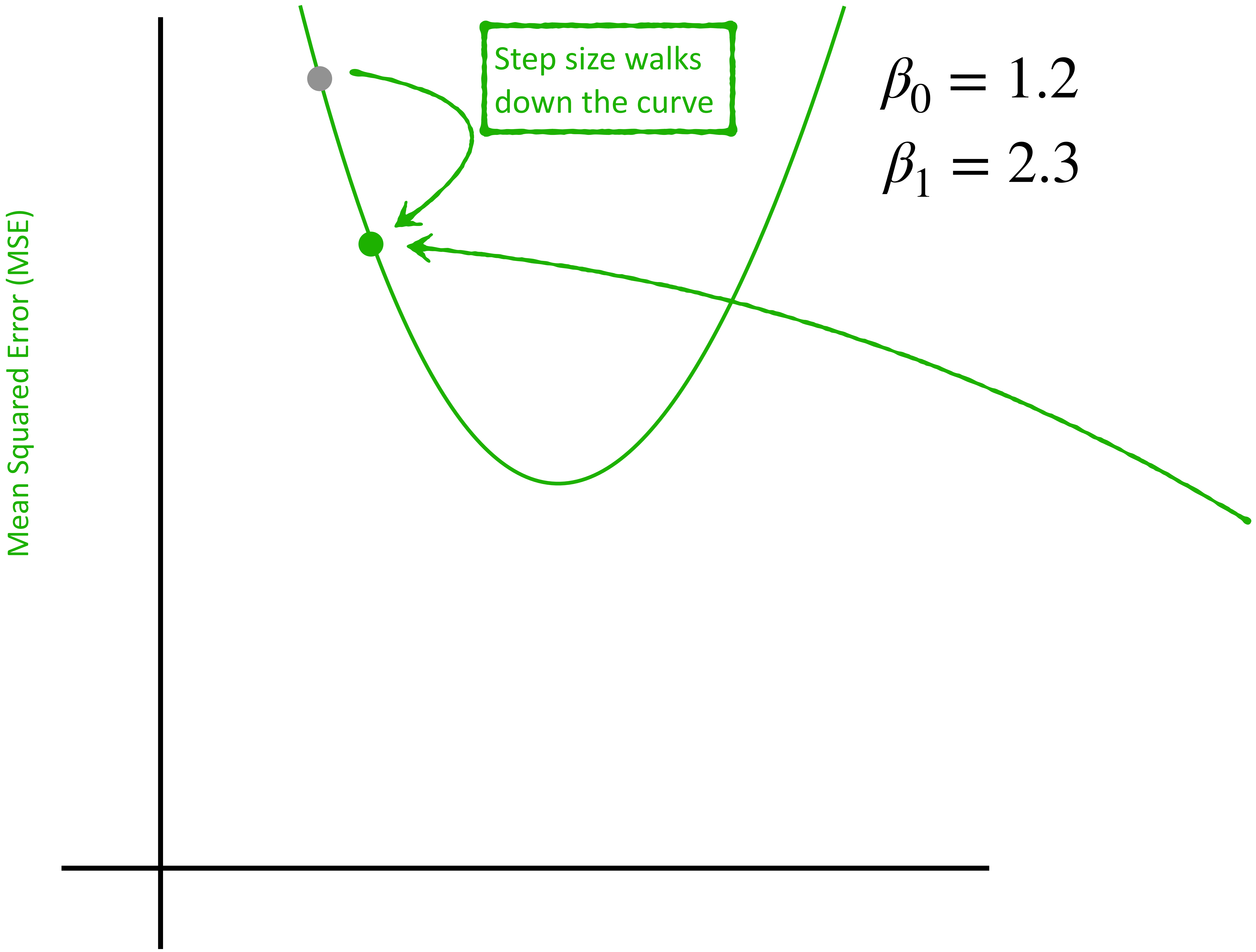
Step 3: Calculate a step size that is proportional to the slope

$$step\_size_{\beta_0} = \frac{\partial}{\beta_0} MSE \times learning\_rate$$

$$step\_size_{\beta_1} = \frac{\partial}{\beta_1} MSE \times learning\_rate$$

$learning\_rate$ is a small value that determines how the algorithm adjusts the parameters on each iteration. Too large and it will take big steps and fail to converge. Too small and it will take many small steps and take too long to converge.

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



Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

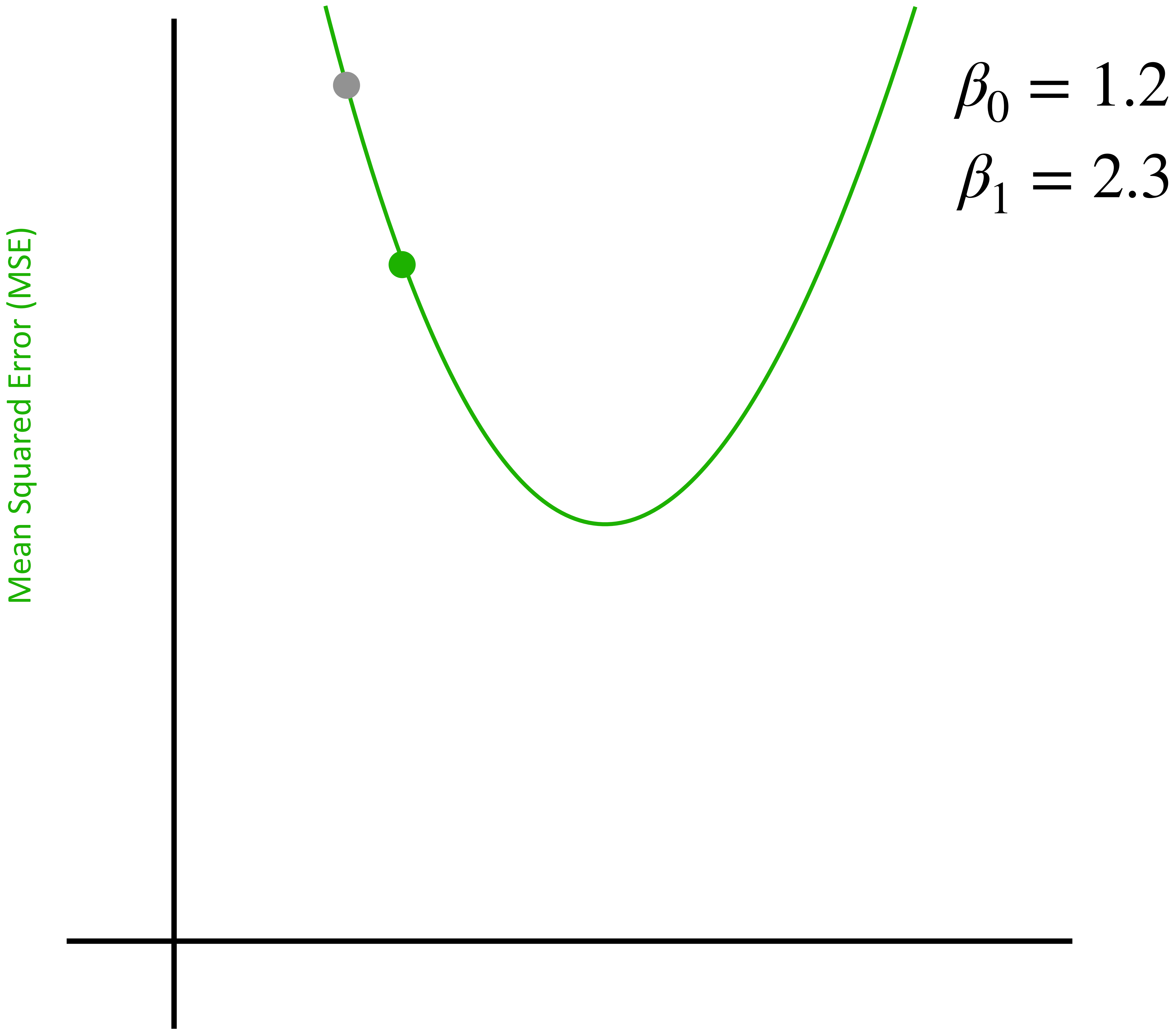
Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

$$\beta_0 = \beta_0 - step\_size_{\beta_0}$$

$$\beta_1 = \beta_1 - step\_size_{\beta_1}$$

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



Gradient Descent

Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

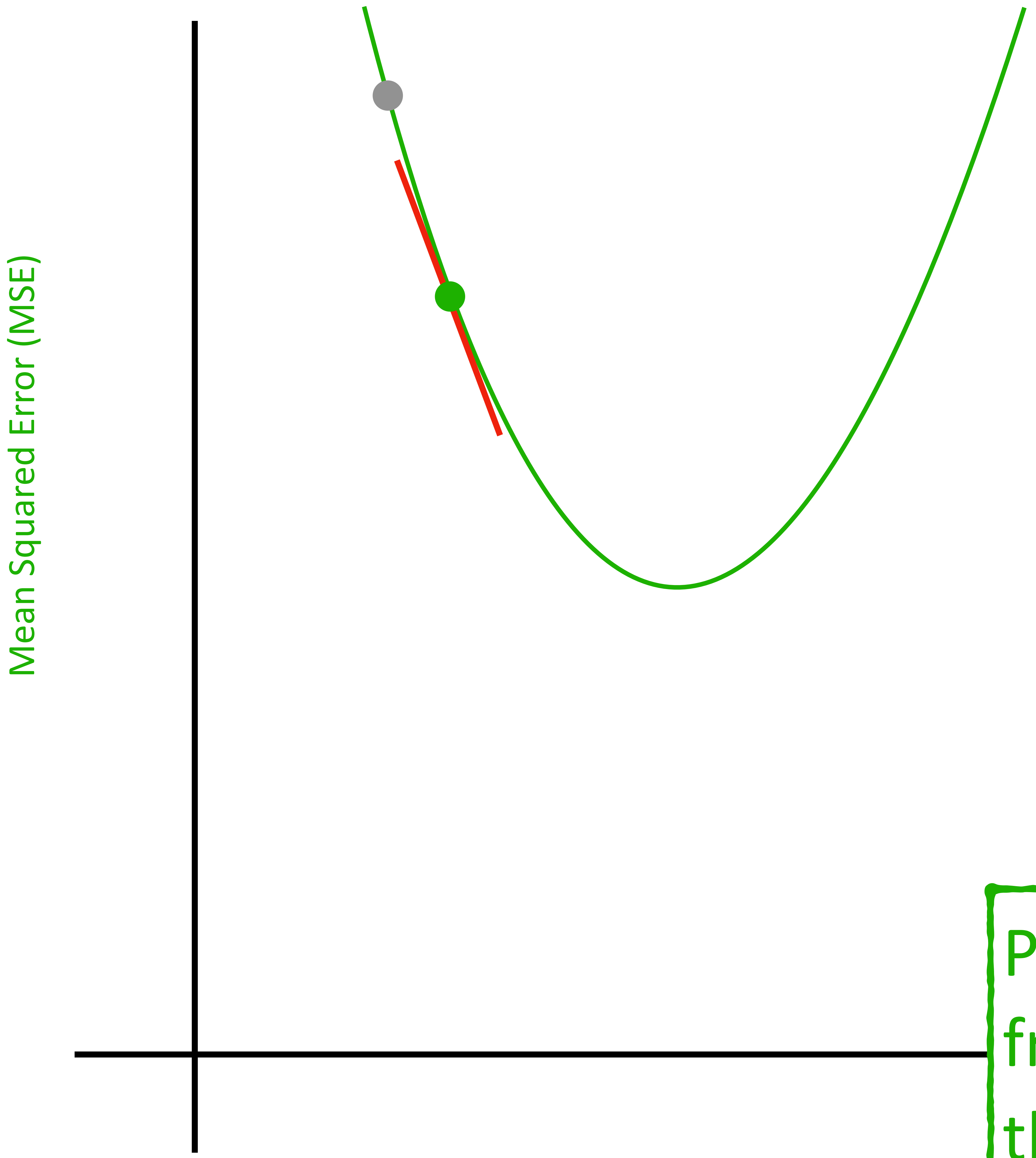
Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

Gradient Descent

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



$$\beta_0 = 1.2$$
$$\beta_1 = 2.3$$

Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

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

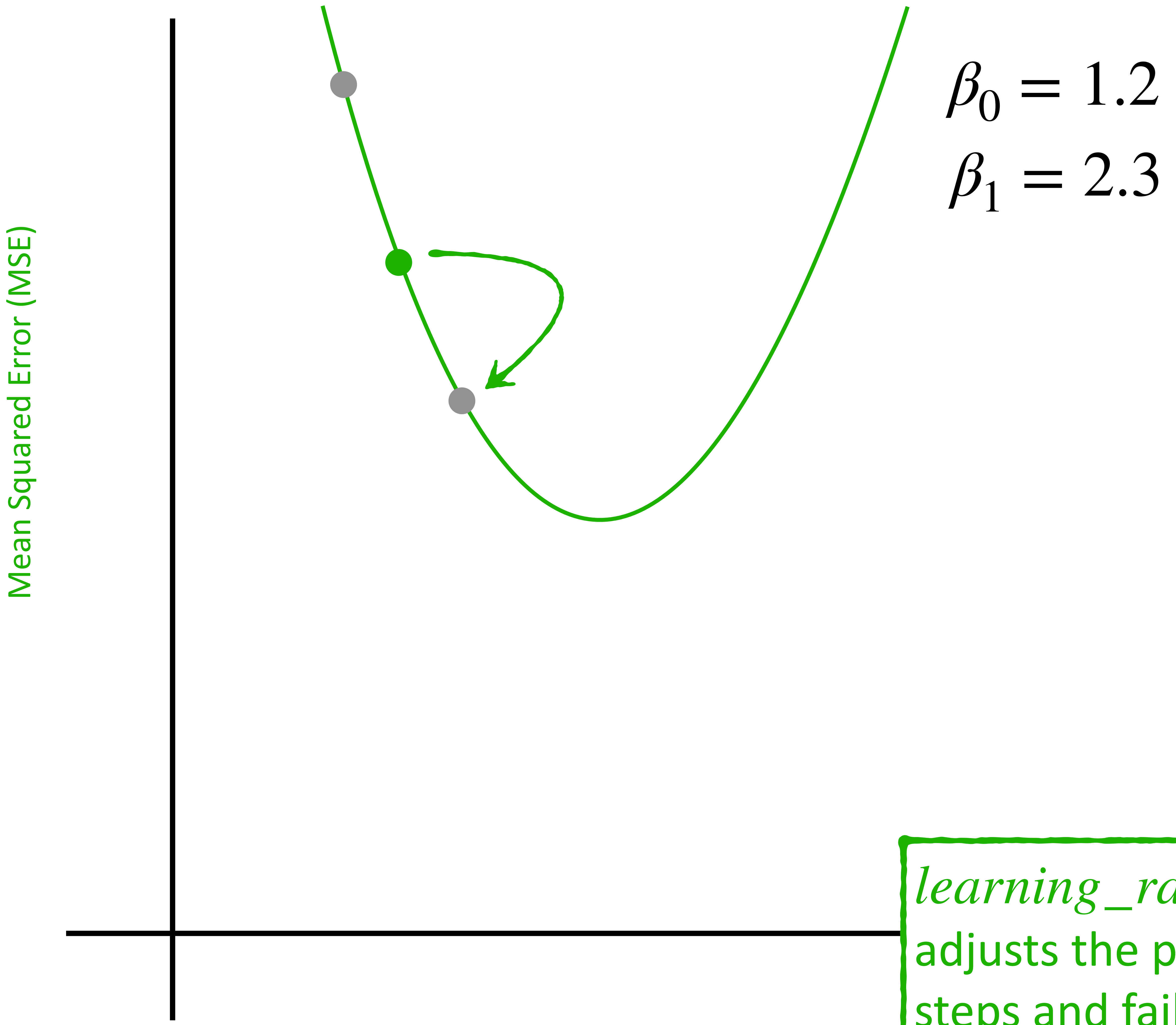
$$\frac{\partial}{\beta_1}MSE = \frac{\partial}{\partial\beta_1}\frac{1}{2n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = -\frac{1}{n}\sum_{i=1}^{n}(x_i y_i - \beta_0 x_i - \beta_1 x_i^2)$$

Plugin the values of x_i and y_i from the observations and the values of β_0 and β_1

i	Height (in)	Weight (lbs)
1	62	138
2	55	178
3	44	123
4	75	200
5	65	229
6	50	102

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

Gradient Descent



Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

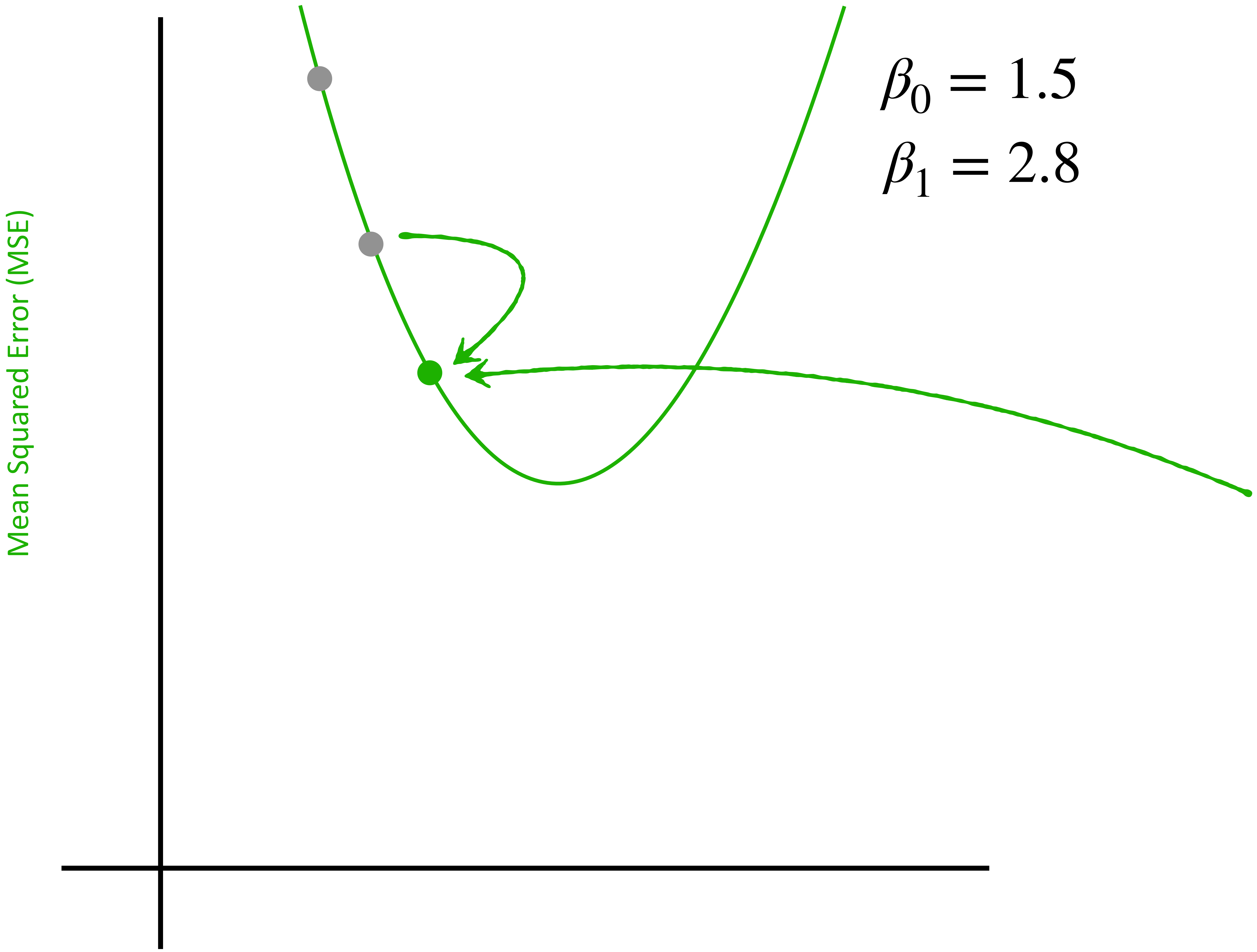
Step 3: Calculate a step size that is proportional to the slope

$$step\_size_{\beta_0} = \frac{\partial}{\beta_0} MSE \times learning\_rate$$

$$step\_size_{\beta_1} = \frac{\partial}{\beta_1} MSE \times learning\_rate$$

$learning\_rate$ is a small value that determines how the algorithm adjusts the parameters on each iteration. Too large and it will take big steps and fail to converge. Too small and it will take many small steps and take too long to converge.

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



Gradient Descent

Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

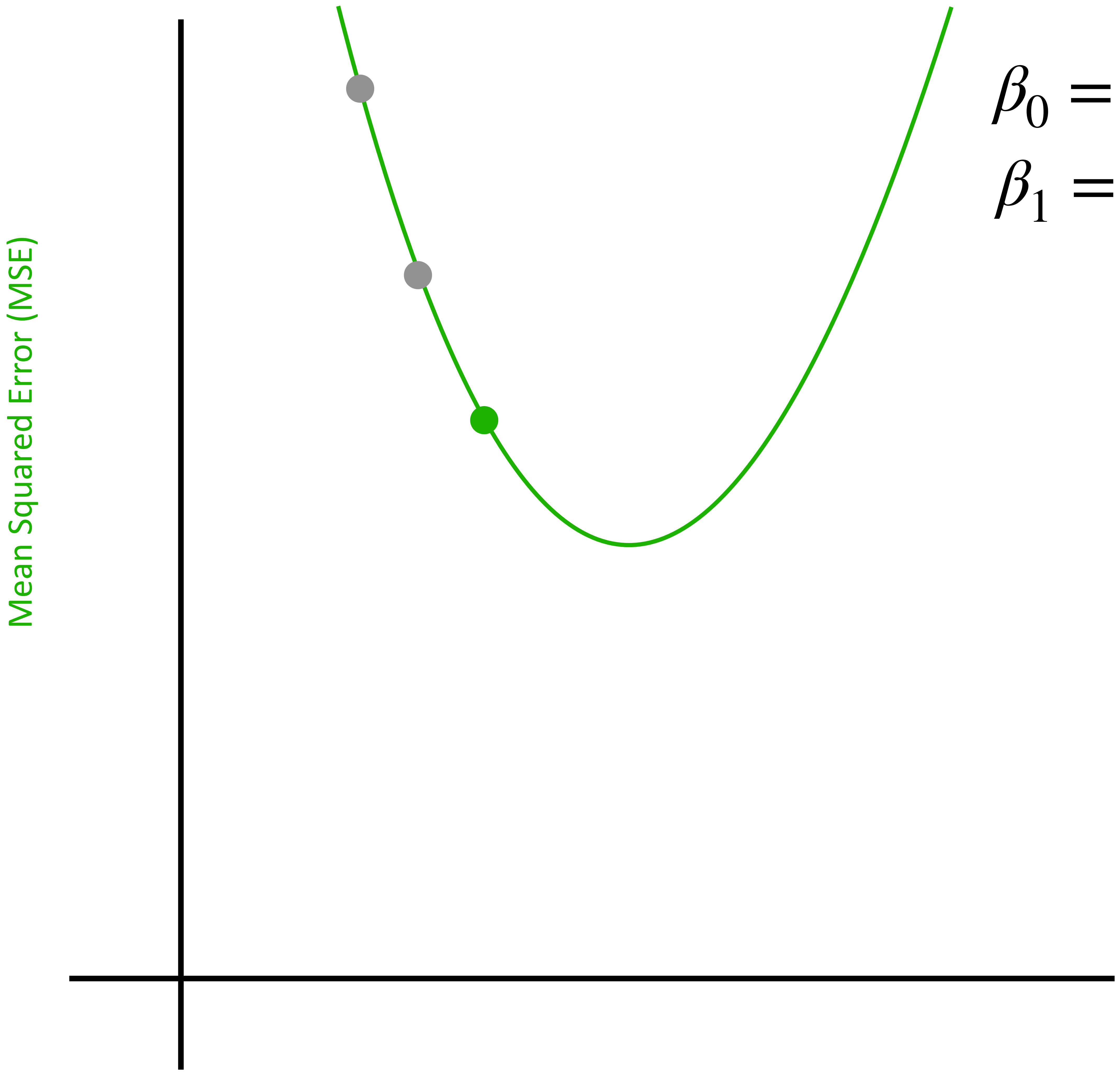
Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

$$\beta_0 = \beta_0 - step\_size_{\beta_0}$$

$$\beta_1 = \beta_1 - step\_size_{\beta_1}$$

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



$\beta_0 = 1.5$

$\beta_1 = 2.8$

Gradient Descent

Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

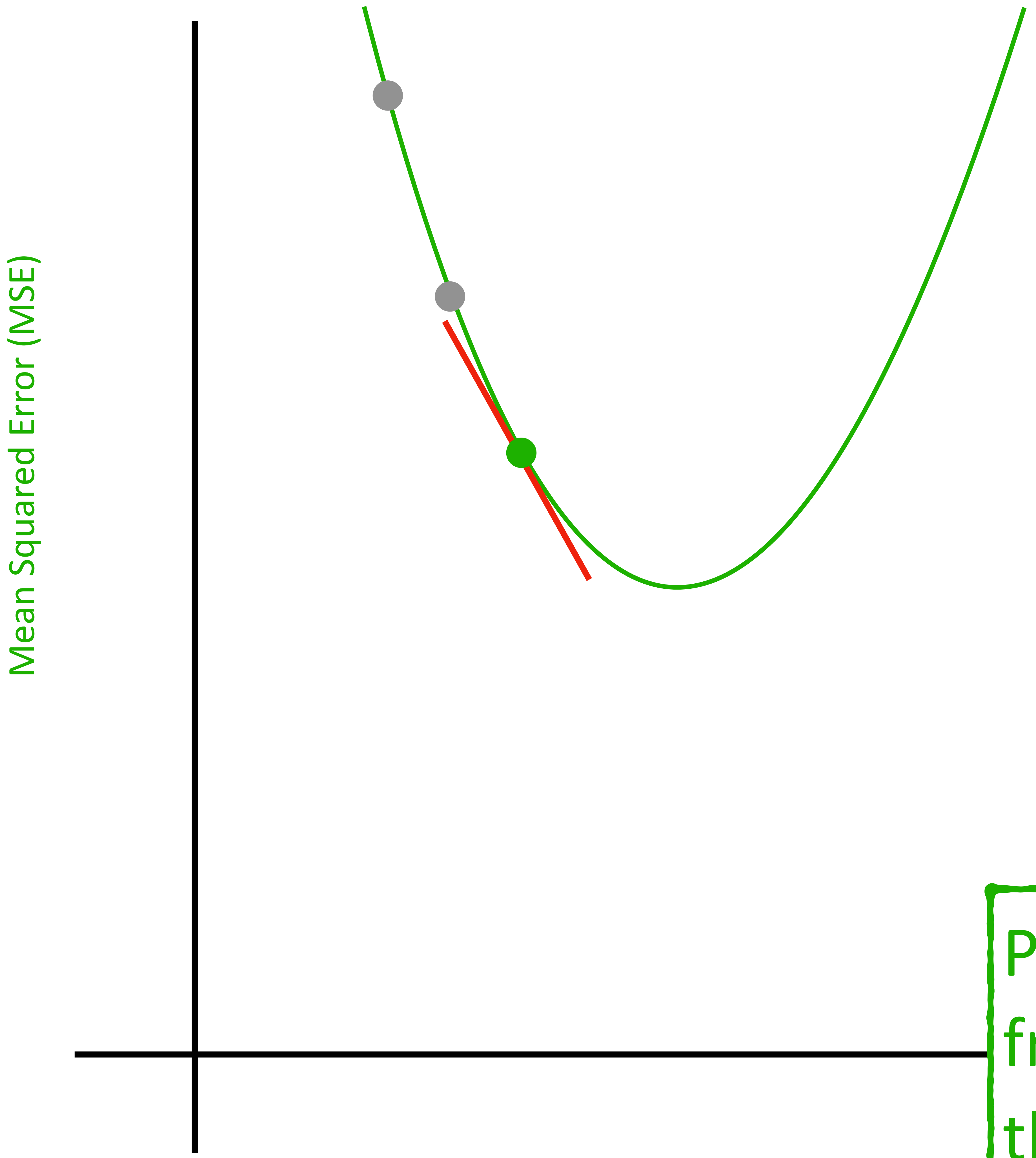
Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

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

Gradient Descent



$$\beta_0 = 1.5$$
$$\beta_1 = 2.8$$

Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

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

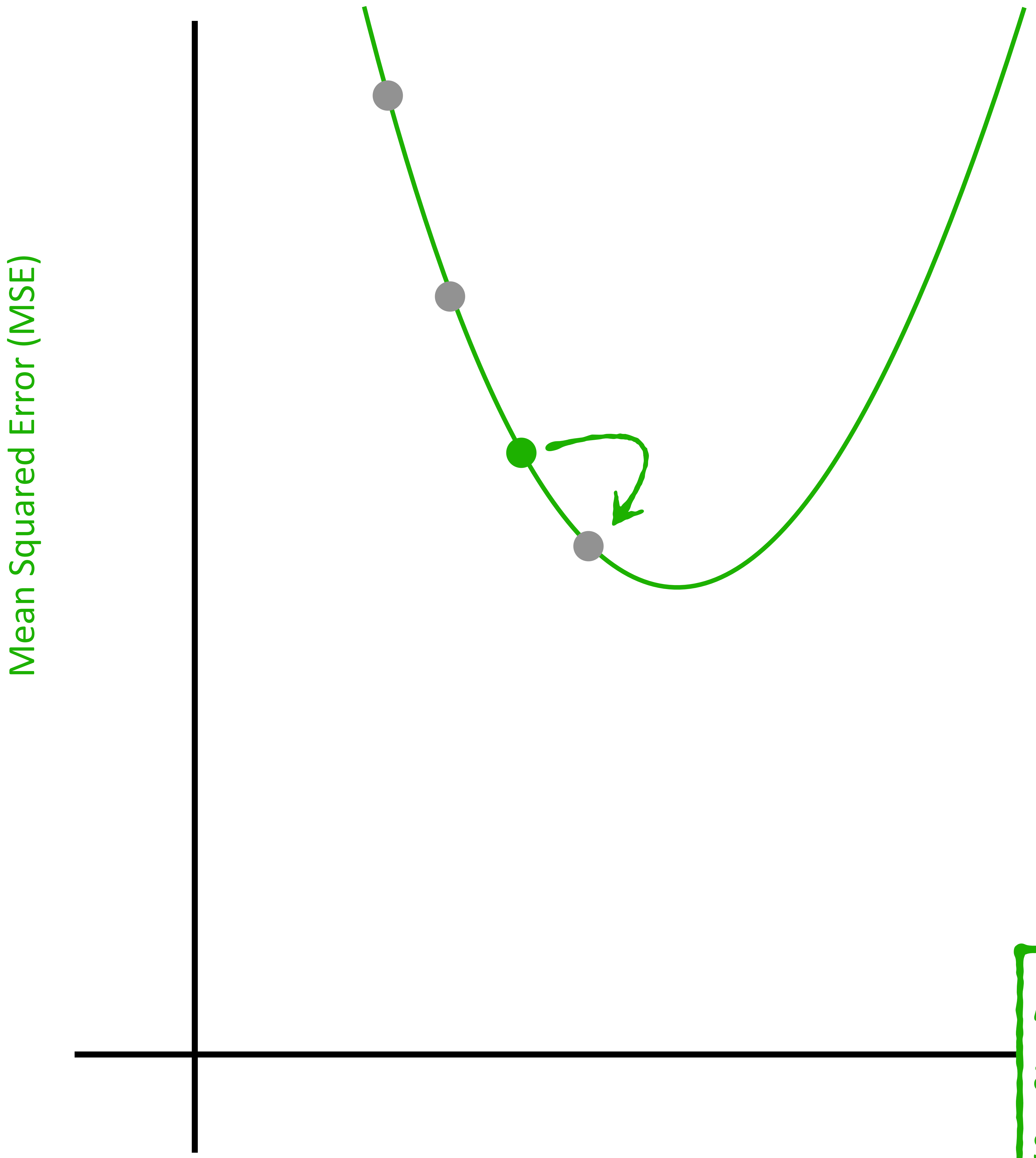
$$\frac{\partial}{\beta_1}MSE = \frac{\partial}{\partial\beta_1}\frac{1}{2n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = -\frac{1}{n}\sum_{i=1}^{n}(x_i y_i - \beta_0 x_i - \beta_1 x_i^2)$$

Plugin the values of x_i and y_i from the observations and the values of β_0 and β_1

i	Height (in)	Weight (lbs)
1	62	138
2	55	178
3	44	123
4	75	200
5	65	229
6	50	102

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

Gradient Descent



$$\beta_0 = 1.5$$
$$\beta_1 = 2.8$$

Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

Step 3: Calculate a step size that is proportional to the slope

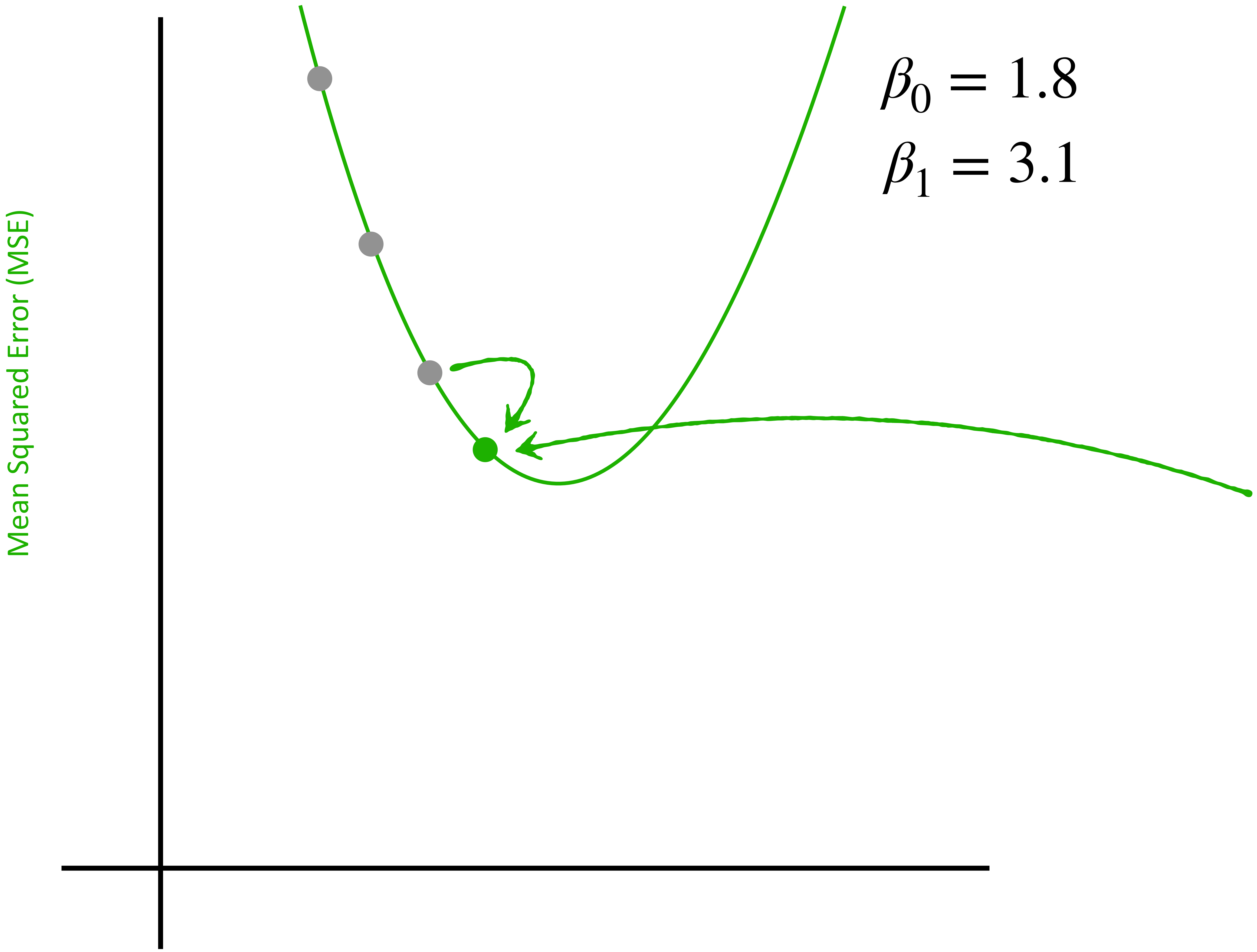
$$step\_size_{\beta_0} = \frac{\partial}{\beta_0} MSE \times learning\_rate$$

$$step\_size_{\beta_1} = \frac{\partial}{\beta_1} MSE \times learning\_rate$$

$learning\_rate$ is a small value that determines how the algorithm adjusts the parameters on each iteration. Too large and it will take big steps and fail to converge. Too small and it will take many small steps and take too long to converge.

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

Gradient Descent



Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

Step 3: Calculate a step size that is proportional to the slope

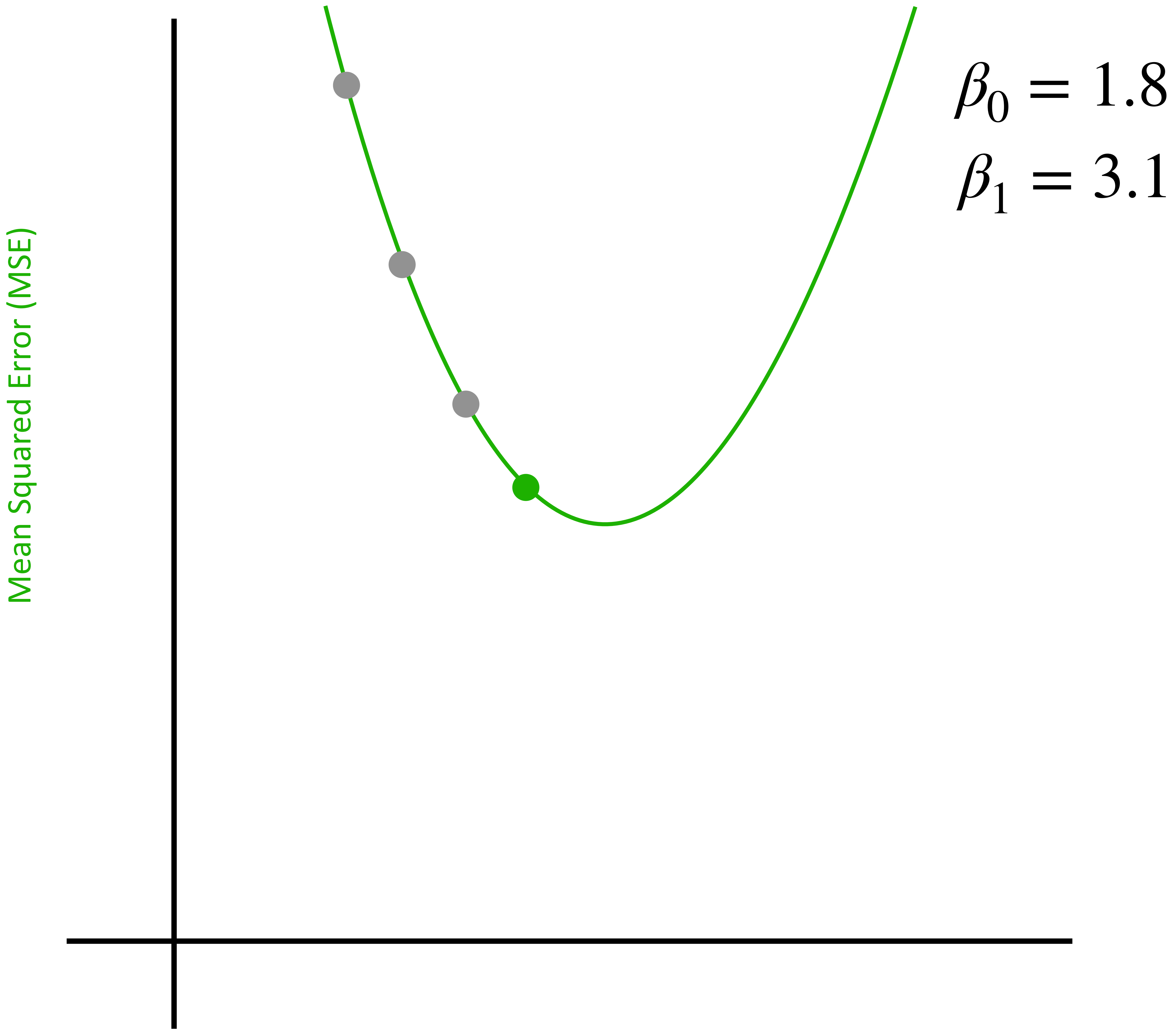
Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

$$\beta_0 = \beta_0 - step\_size_{\beta_0}$$

$$\beta_1 = \beta_1 - step\_size_{\beta_1}$$

Gradient Descent

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



Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

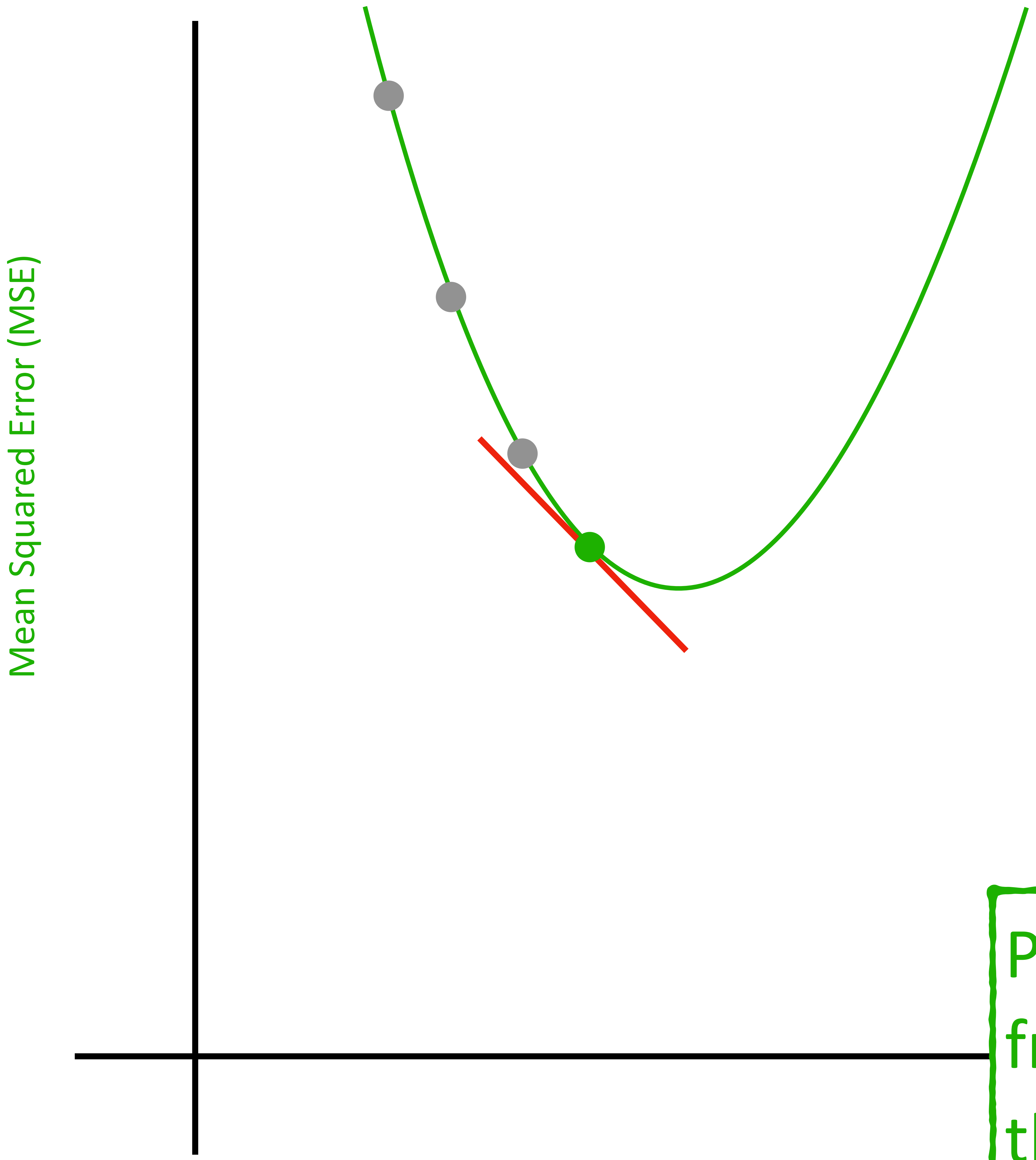
Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

Gradient Descent

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



$$\beta_0 = 1.8$$
$$\beta_1 = 3.1$$

Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

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

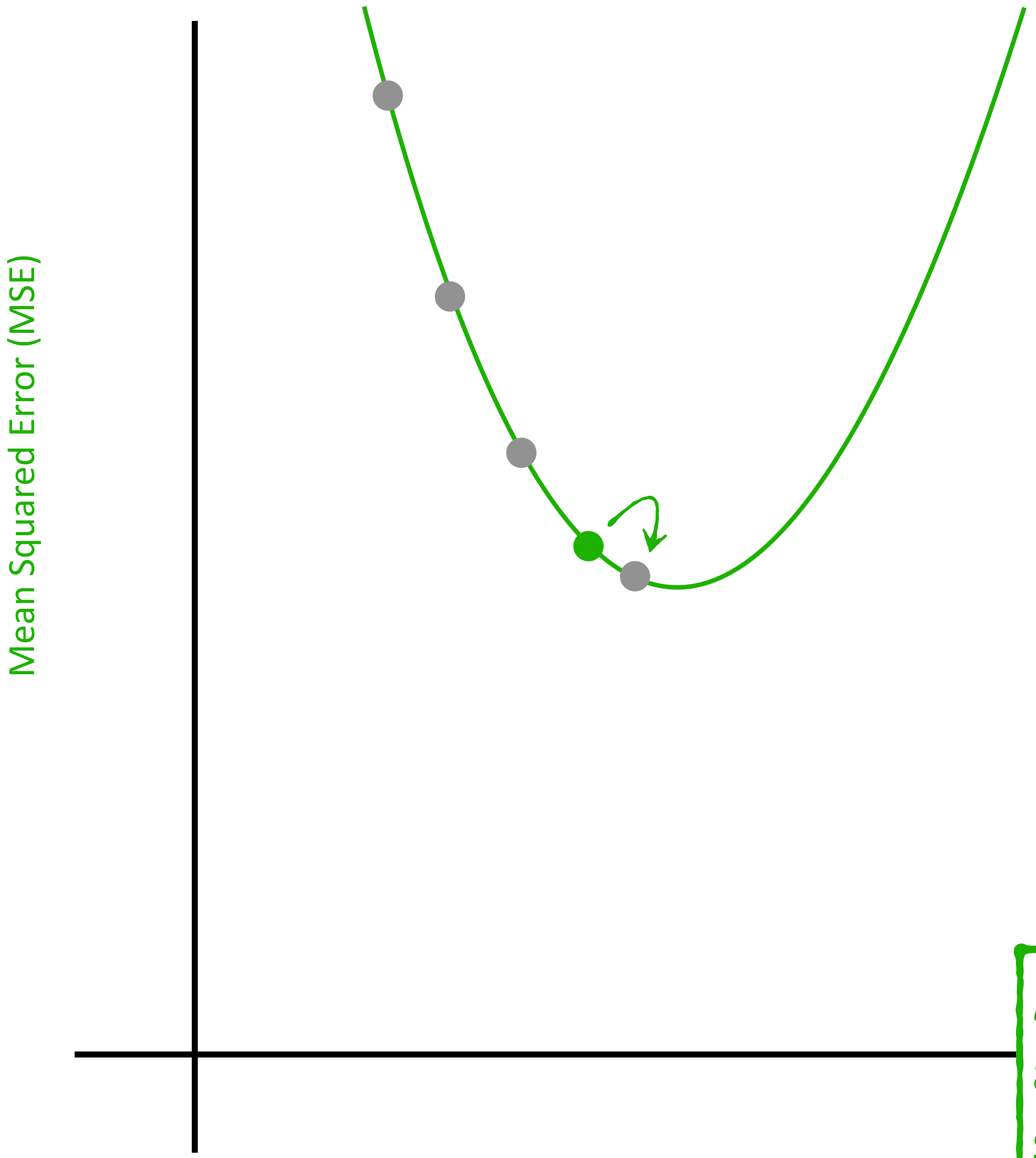
$$\frac{\partial}{\beta_1} MSE = \frac{\partial}{\partial \beta_1} \frac{1}{2n} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 = -\frac{1}{n} \sum_{i=1}^{n} (x_i y_i - \beta_0 x_i - \beta_1 x_i^2)$$

Plugin the values of x_i and y_i from the observations and the values of β_0 and β_1

i	Height (in)	Weight (lbs)
1	62	138
2	55	178
3	44	123
4	75	200
5	65	229
6	50	102

Gradient Descent

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



$$\beta_0 = 1.8$$
$$\beta_1 = 3.1$$

Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

Step 3: Calculate a step size that is proportional to the slope

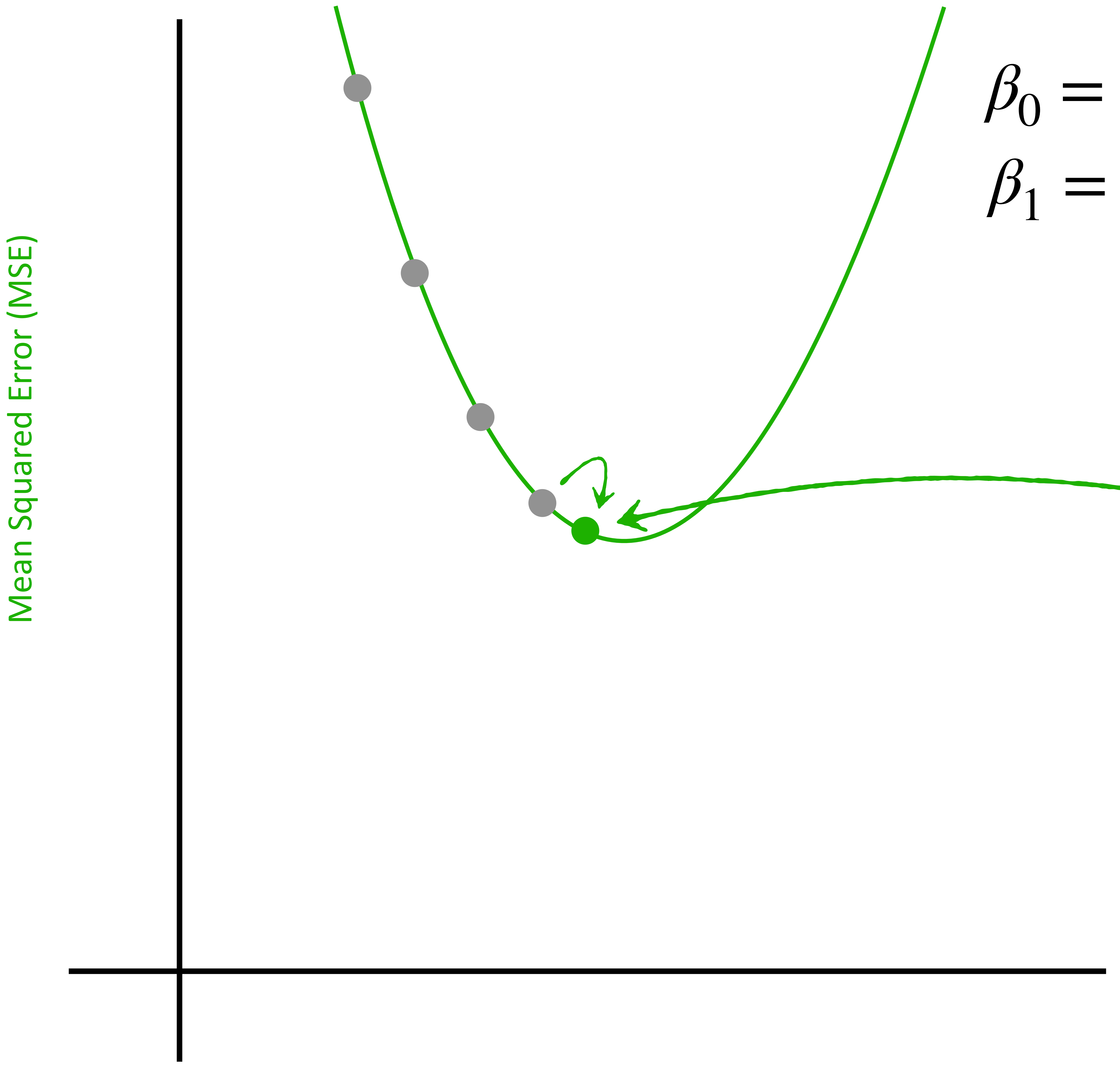
$$step\_size_{\beta_0} = \frac{\partial}{\beta_0} MSE \times learning\_rate$$

$$step\_size_{\beta_1} = \frac{\partial}{\beta_1} MSE \times learning\_rate$$

$learning\_rate$ is a small value that determines how the algorithm adjusts the parameters on each iteration. Too large and it will take big steps and fail to converge. Too small and it will take many small steps and take too long to converge.

Gradient Descent

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



$$\beta_0 = 1.9$$
$$\beta_1 = 3.2$$

Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

Step 3: Calculate a step size that is proportional to the slope

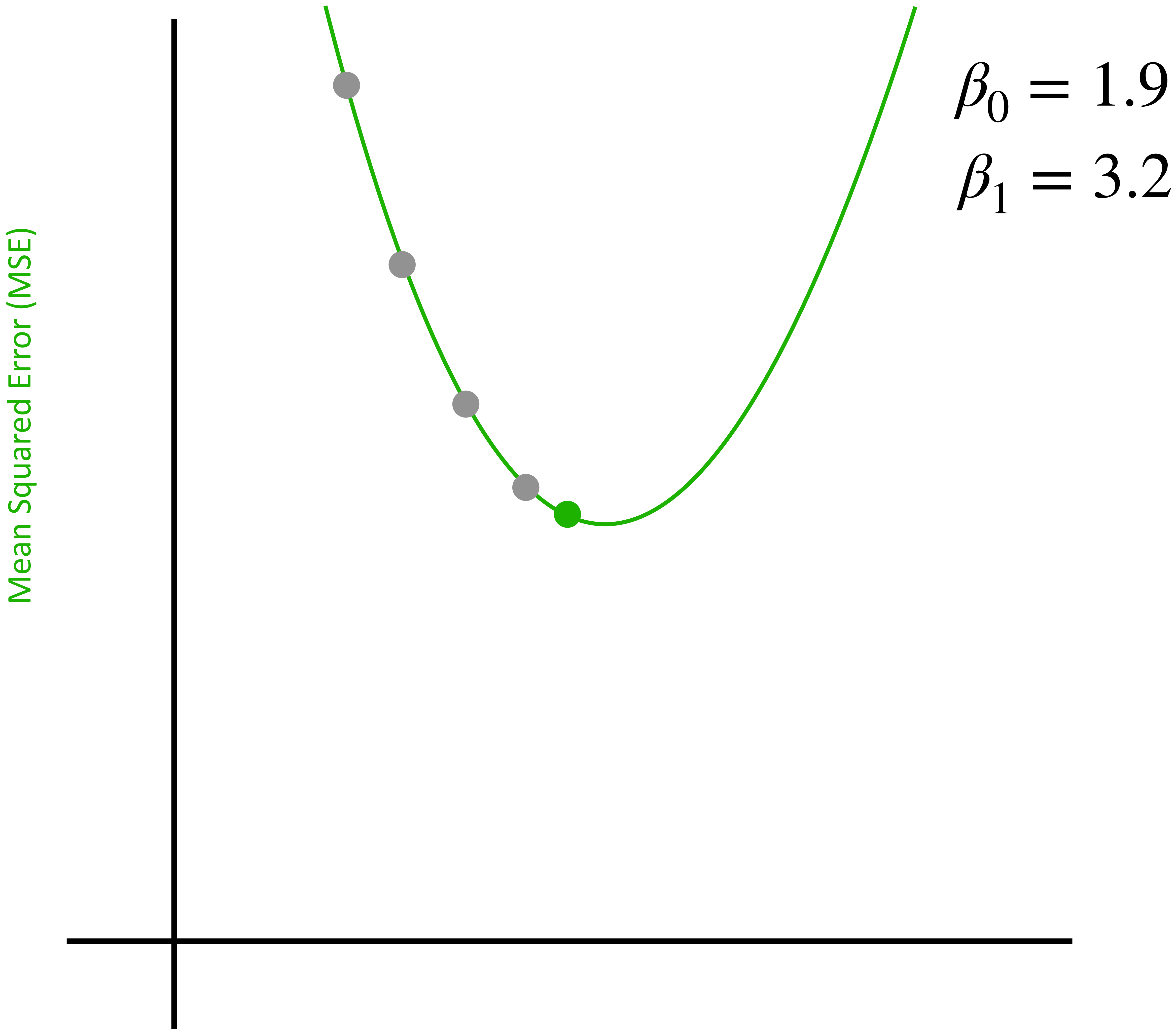
Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

$$\beta_0 = \beta_0 - step\_size_{\beta_0}$$

$$\beta_1 = \beta_1 - step\_size_{\beta_1}$$

Gradient Descent

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



Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

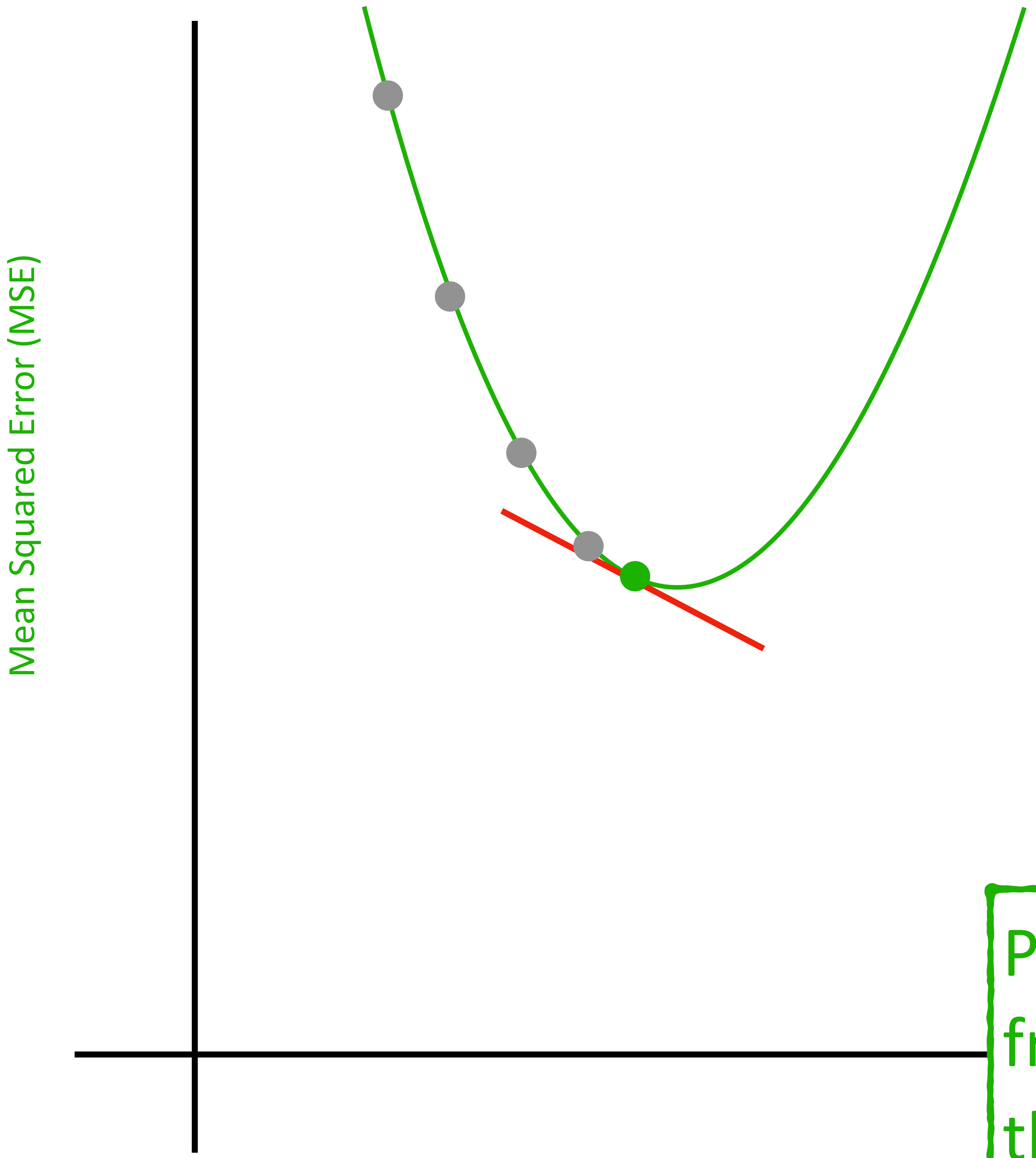
Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

Gradient Descent

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



$\beta_0 = 1.9$

$\beta_1 = 3.2$

Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

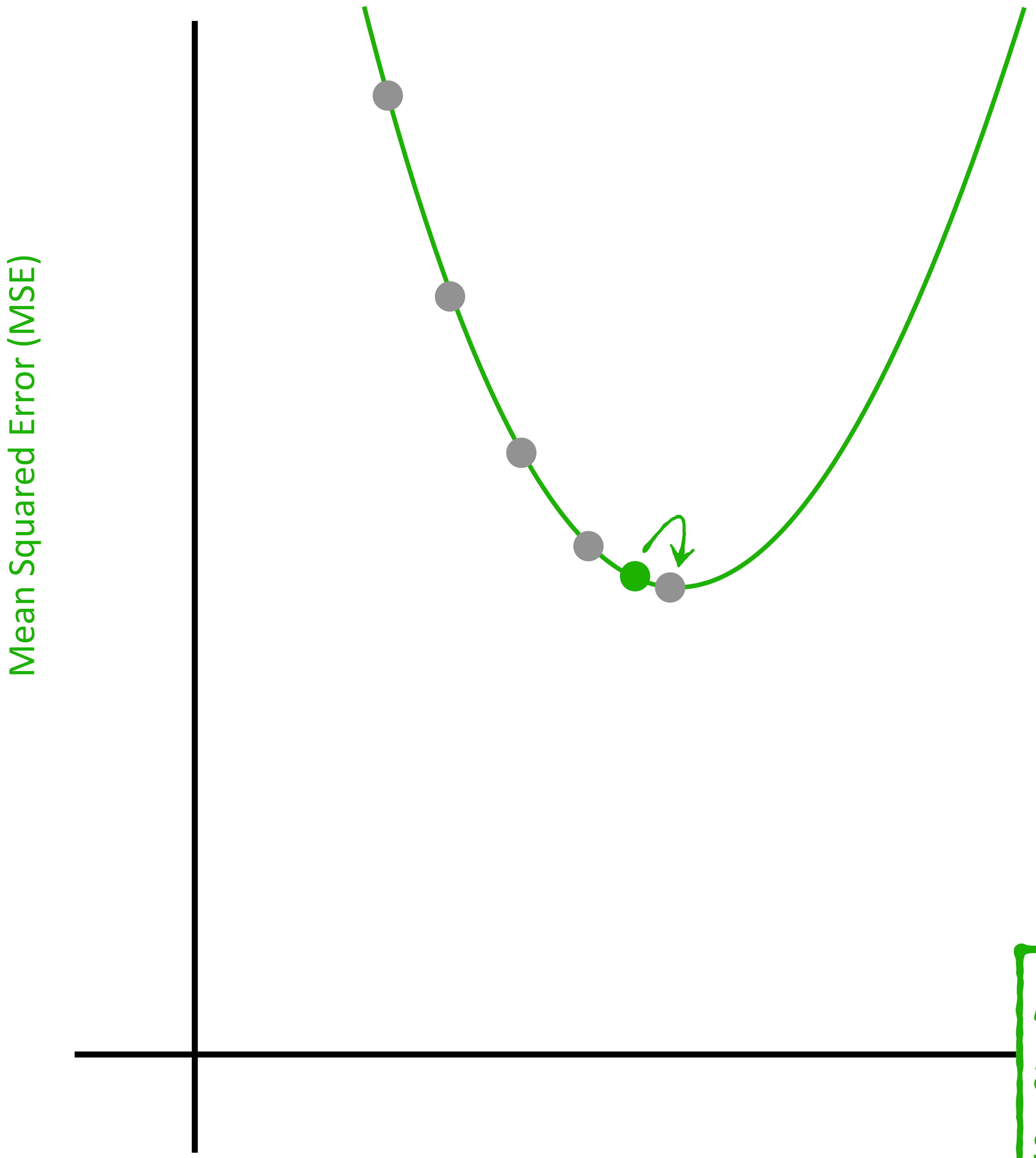
$$\frac{\partial}{\beta_0}MSE = \frac{\partial}{\partial \beta_0}\frac{1}{2n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = -\frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)$$

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

Plugin the values of x_i and y_i from the observations and the values of β_0 and β_1

i	Height (in)	Weight (lbs)
1	62	138
2	55	178
3	44	123
4	75	200
5	65	229
6	50	102

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



$$\beta_0 = 1.9$$
$$\beta_1 = 3.2$$

Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

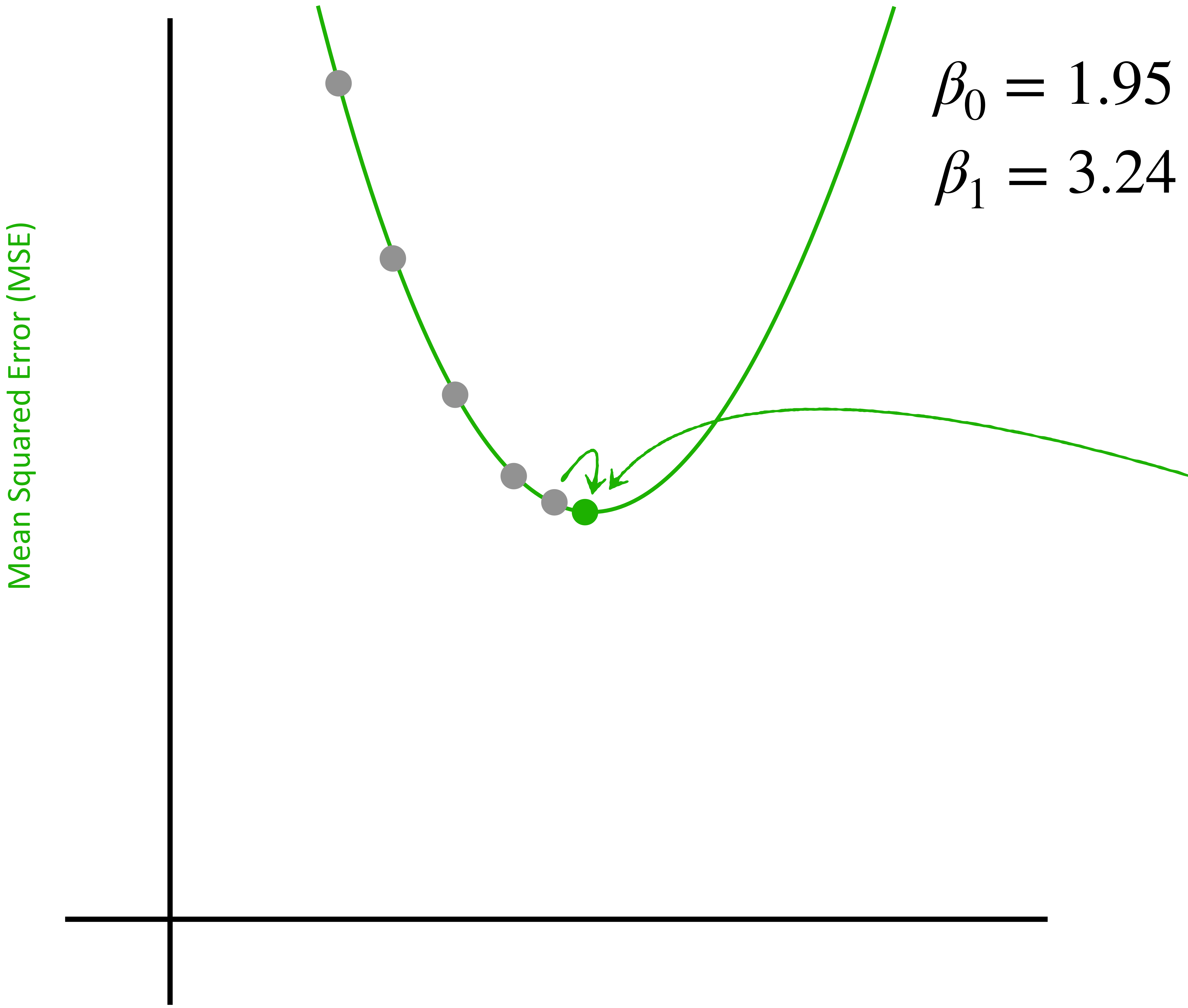
Step 3: Calculate a step size that is proportional to the slope

$$step\_size_{\beta_0} = \frac{\partial}{\beta_0} MSE \times learning\_rate$$

$$step\_size_{\beta_1} = \frac{\partial}{\beta_1} MSE \times learning\_rate$$

learning_rate is a small value that determines how the algorithm adjusts the parameters on each iteration. Too large and it will take big steps and fail to converge. Too small and it will take many small steps and take too long to converge.

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



Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

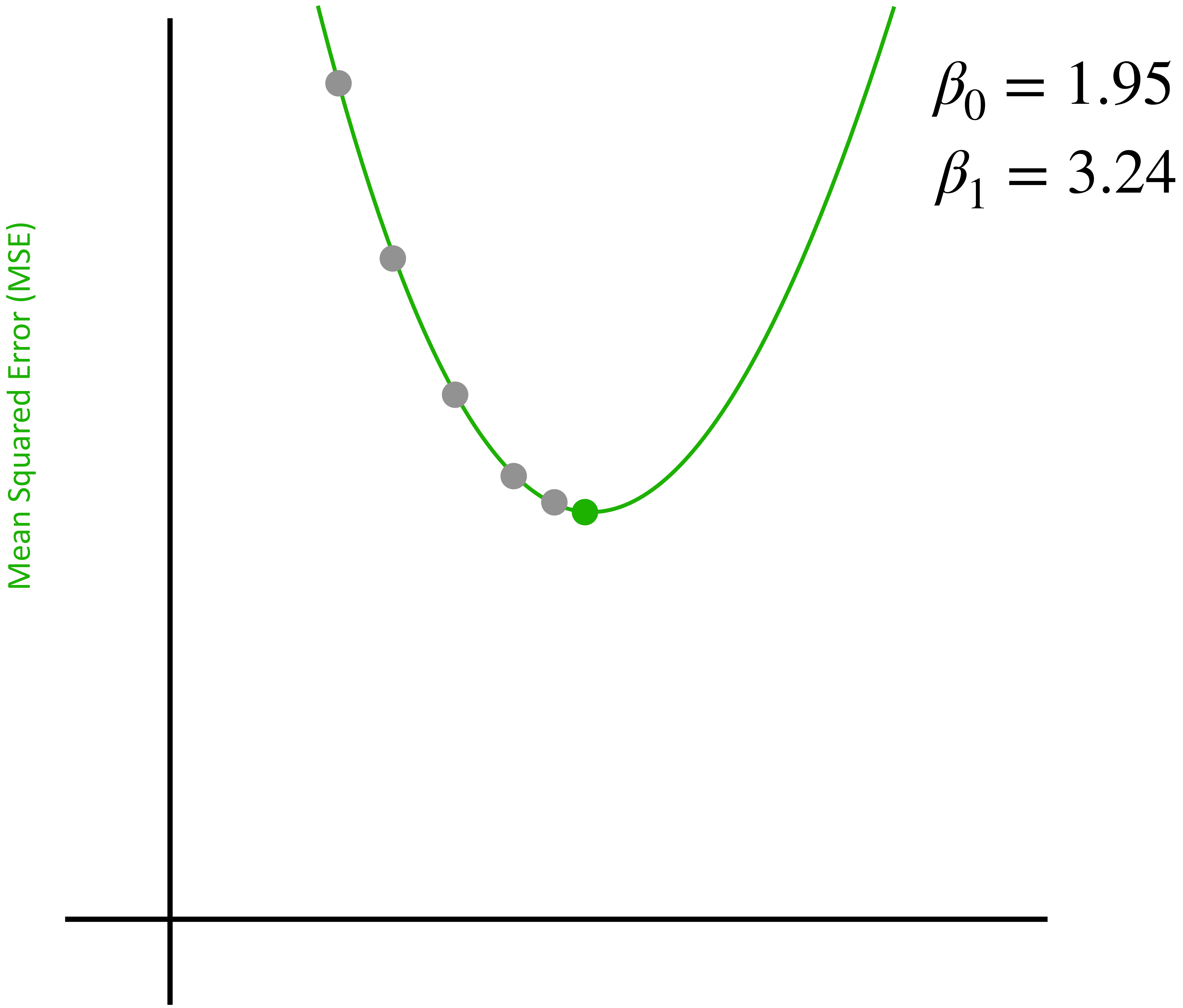
Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

$$\beta_0 = \beta_0 - step\_size_{\beta_0}$$

$$\beta_1 = \beta_1 - step\_size_{\beta_1}$$

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



Gradient Descent

Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

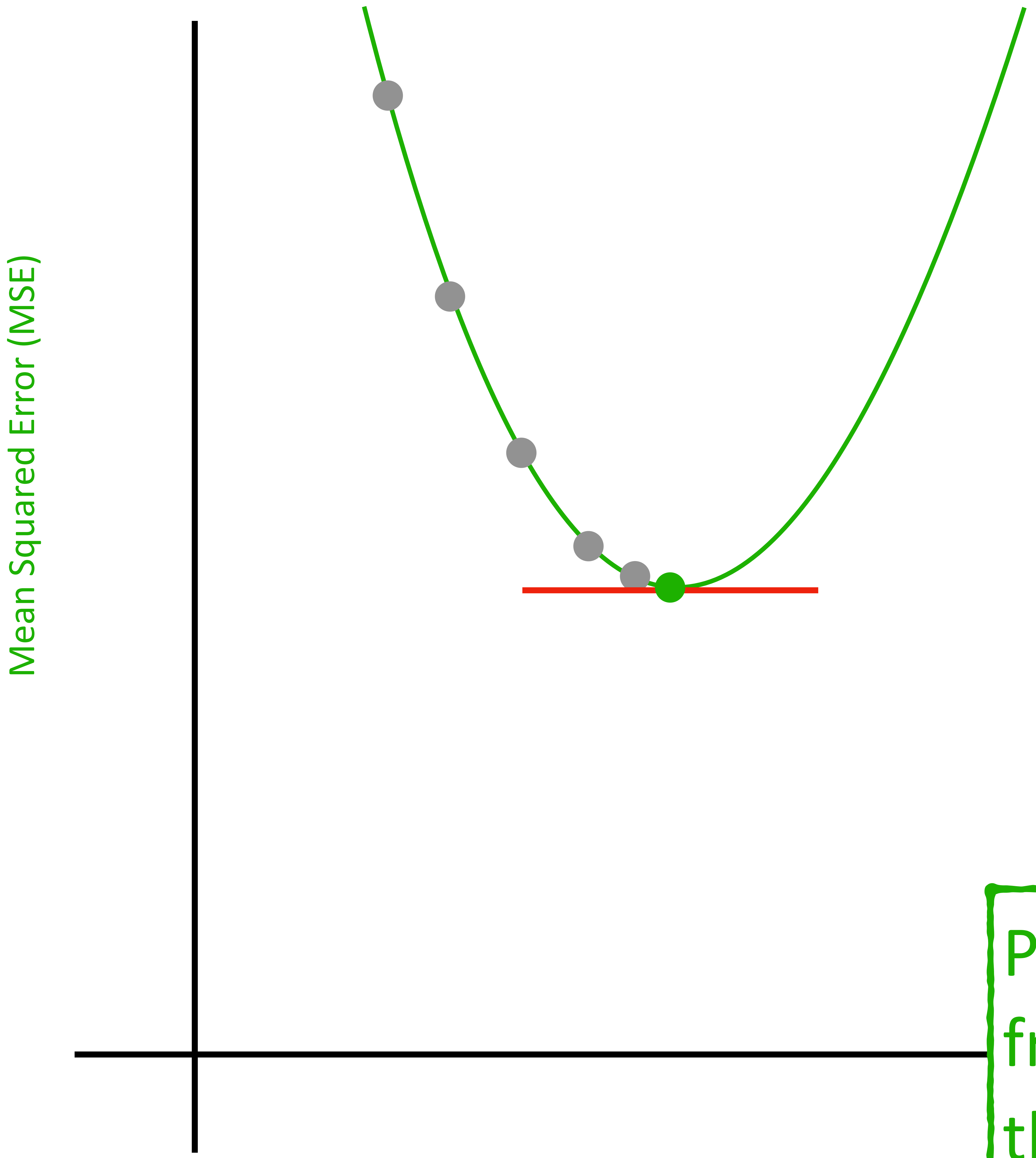
Step 3: Calculate a step size that is proportional to the slope

Step 4: Calculate new values for β_0 and β_1 by subtracting the step size

Step 5: Go to step 2 and repeat

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

Gradient Descent



$$\beta_0 = 1.95$$
$$\beta_1 = 3.24$$

Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

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

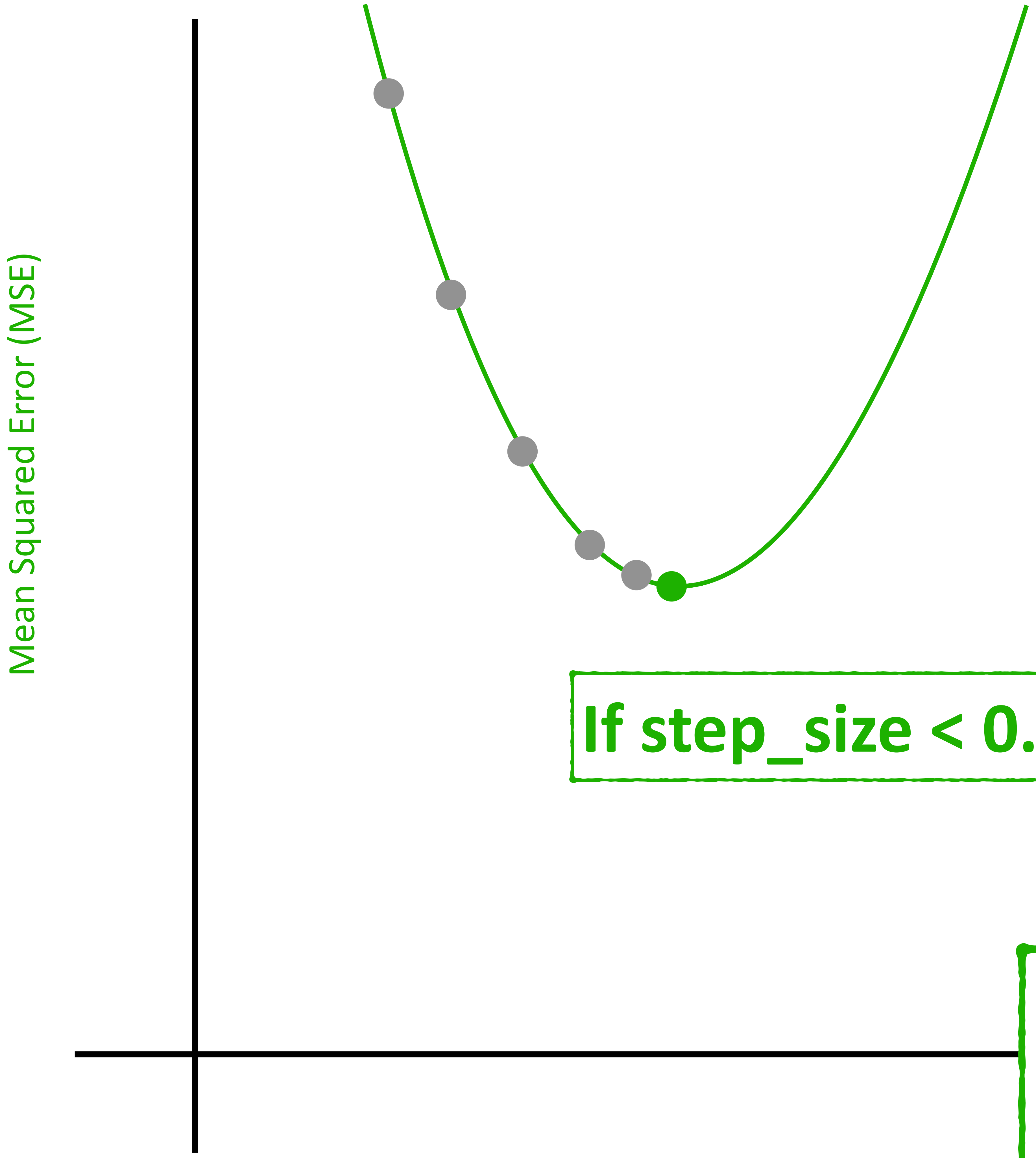
$$\frac{\partial}{\beta_1}MSE = \frac{\partial}{\partial \beta_1}\frac{1}{2n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = -\frac{1}{n}\sum_{i=1}^{n}(x_i y_i - \beta_0 x_i - \beta_1 x_i^2)$$

Plugin the values of x_i and y_i from the observations and the values of β_0 and β_1

i	Height (in)	Weight (lbs)
1	62	138
2	55	178
3	44	123
4	75	200
5	65	229
6	50	102

Gradient Descent

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



$\beta_0 = 1.9$

$\beta_1 = 3.2$

If step_size < 0.0001 then stop

Gradient Descent: Basic Concept

Step 1: Start with random values for β_0 and β_1

Step 2: Compute the partial derivative of the MSE w.r.t β_0 and β_1 - this is the slope at that point

Step 3: Calculate a step size that is proportional to the slope

$$step\_size_{\beta_0} = \frac{\partial}{\beta_0} MSE \times learning\_rate$$

$$step\_size_{\beta_1} = \frac{\partial}{\beta_1} MSE \times learning\_rate$$

$learning\_rate$ is a small value that determines how the algorithm adjusts the parameters on each iteration. Too large and it will take big steps and fail to converge. Too small and it will take many small steps and take too long to converge.

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.

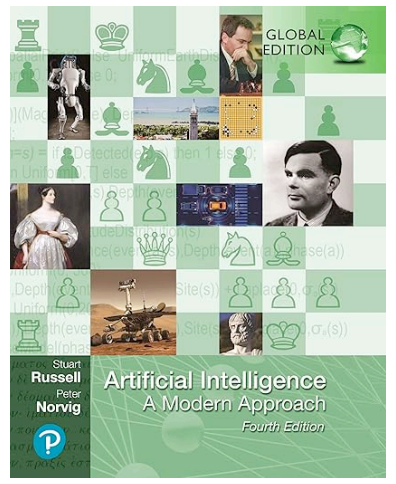
Multiple Regression

Multiple regression extends the two dimensional linear model introduced in Simple Linear Regression to k + 1 dimensions with one dependent variable, k independent variables and k+1 parameters.

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



Artificial Intelligence: A Modern Approach

by Peter Norvig, Stuart Russell

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