Pontificia Universidad JAVERIANA Bogotá

Facultad de Ingeniería

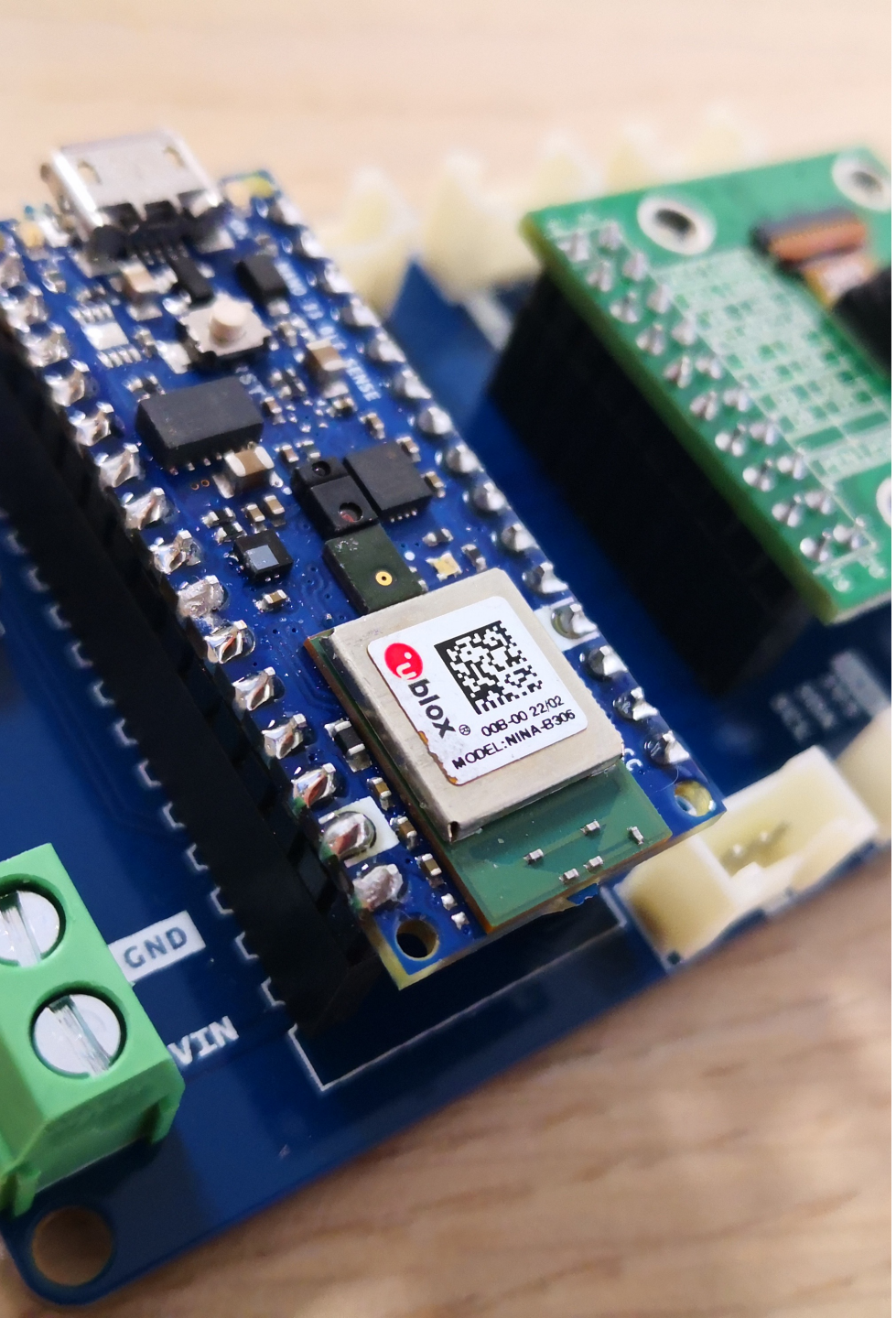
ICTP

TINY ML edu

A Brief Introduction to Deep Learning

Latin American Regional Workshop on SciTinyML:
Scientific Use of Machine Learning on Low-Power Devices

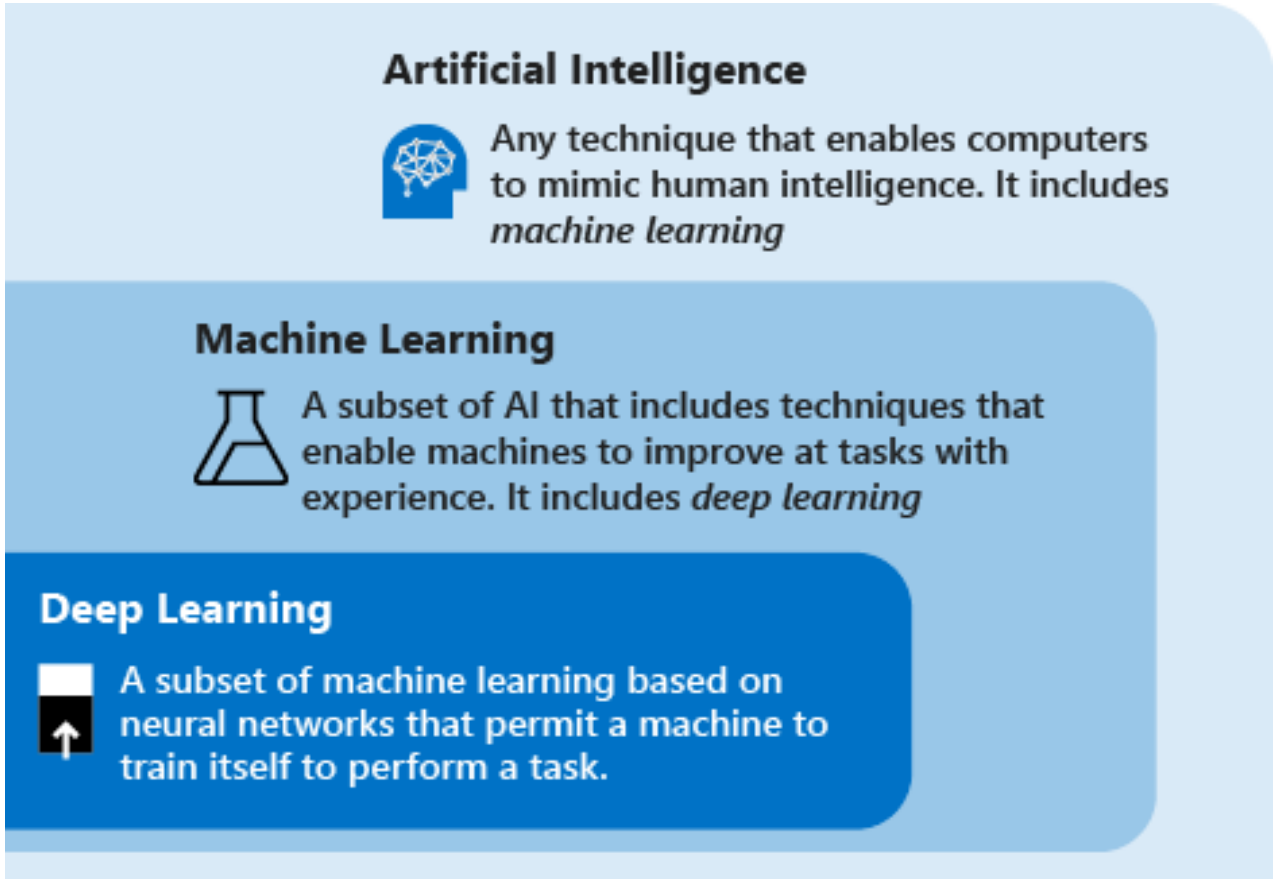
July 11th, 2022



Outline

- AI vs ML vs DL
- The Machine Learning Paradigm
- Finding the Best Solution and Fitting a Model
- Regression and Classification with NN
- ML Issues

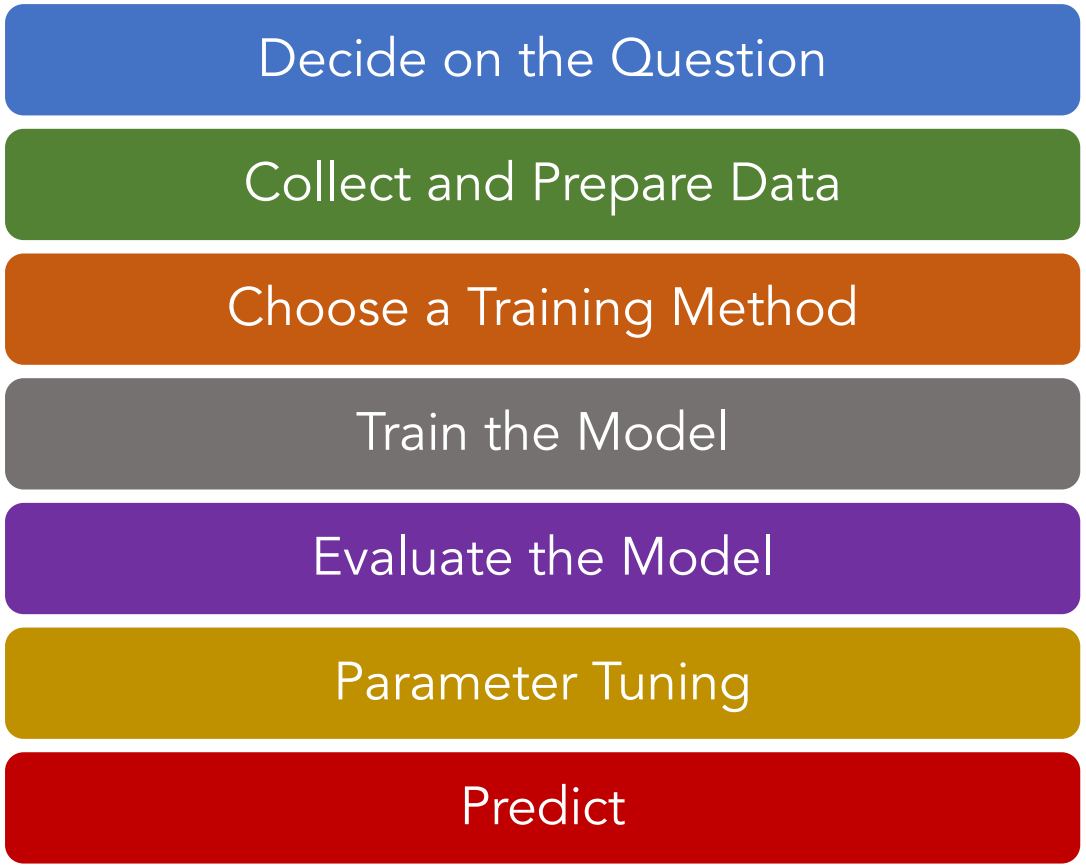
AI vs. ML vs. DL



https://docs.microsoft.com/en-us/azure/machine-learning/concept-deep-learning-vs-machine-learning

General Steps for Machine Learning

On a high level, the craft of creating machine learning (ML) processes is comprised of several steps:



https://microsoft.github.io/ML-For-Beginners/#/1-Introduction/4-techniques-of-ML/README?id=techniques-of-machine-learning

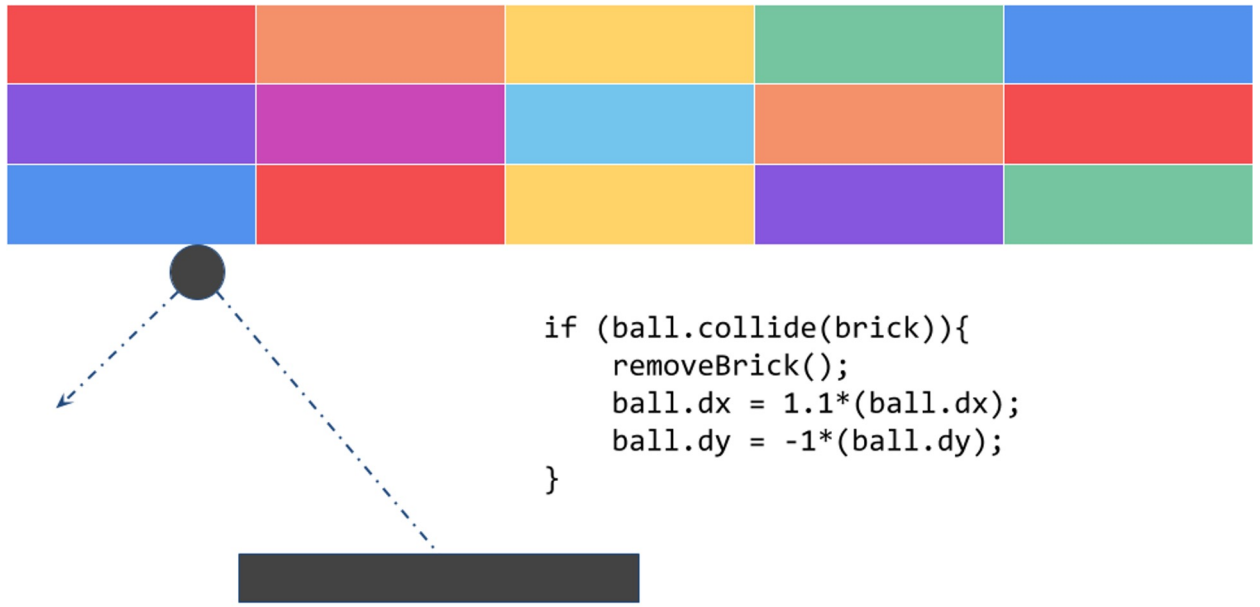
We will run through this long process

This is a first encounter with ML, but many things will be left to be experimented or developed.

The Machine Learning Paradigm

Explicit Coding

- Defining rules that determine behavior of a program
- Everything is pre-calculated and pre-determined by the programmer
- Scenarios are limited by program complexity



```
if (ball.collide(brick)){
    removeBrick();
    ball.dx = 1.1*(ball.dx);
    ball.dy = -1*(ball.dy);
}
```

The Traditional Programming Paradigm

Consider Activity Detection

```
if(speed<4){
    status=WALKING;
}
```

```
if(speed<4){
    status=WALKING;
} else {
    status=RUNNING;
}
```

```
if(speed<4){
    status=WALKING;
} else if(speed<12){
    status=RUNNING;
} else {
    status=BIKING;
}
```

```
// ???
```

Way too complex to code!

The Traditional Programming Paradigm

The Machine Learning Paradigm

Activity Detection with Machine Learning

```
0101001010100101010
1001010101001011101
0100101010010101001
0101001010100101010
```

Label = WALKING

```
1010100101001010101
0101010010010010001
0010011111010101111
1010100100111101011
```

Label = RUNNING

```
1001010011111010101
1101010111010101110
1010101111010101011
1111110001111010101
```

Label = BIKING

```
1111111111010011101
0011111010111110101
0101110101010101110
1010101010100111110
```

Label = GOLFING

The Machine Learning Paradigm

```
0101001010100101010
1001010101001011101
0100101010010101001
0101001010100101010
```

Label = WALKING

```
1010100101001010101
0101010010010010001
0010011111010101111
1010100100111101011
```

Label = RUNNING

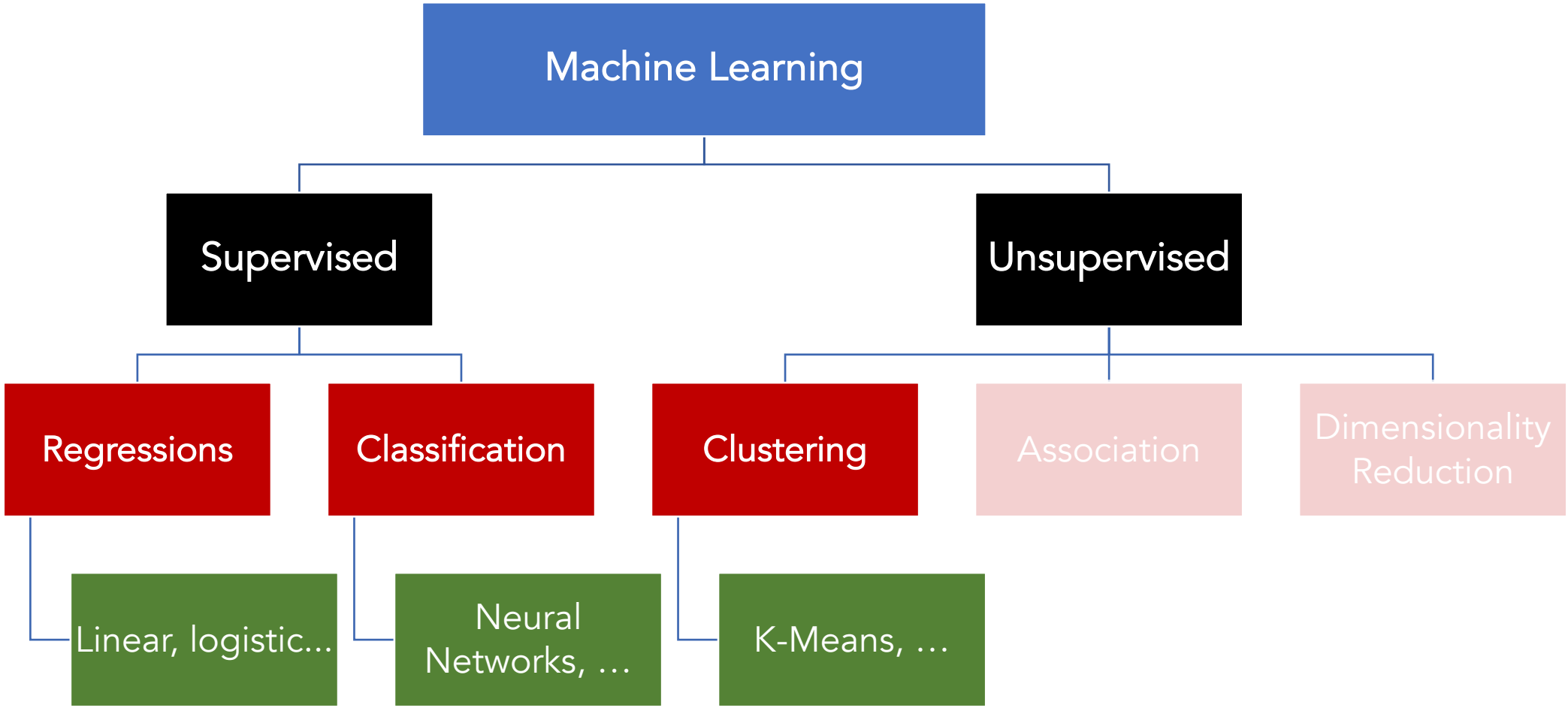
```
1001010011111010101
1101010111010101110
1010101111010101011
1111110001111010101
```

Label = BIKING

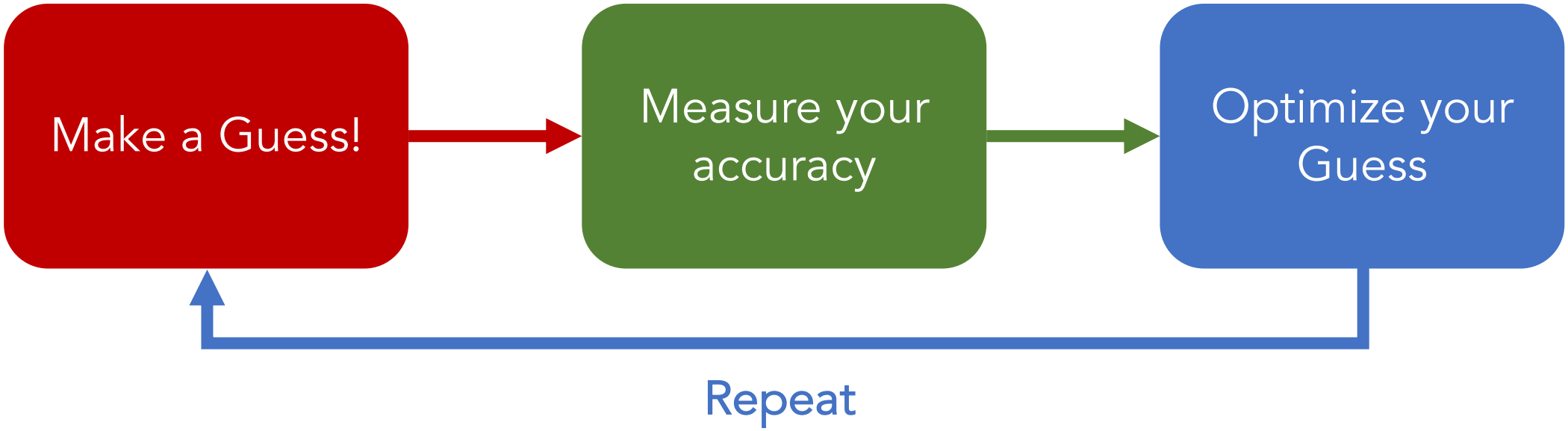
```
1111111111010011101
0011111010111110101
0101110101010101110
1010101010100111110
```

Label = GOLFING

Two Approaches

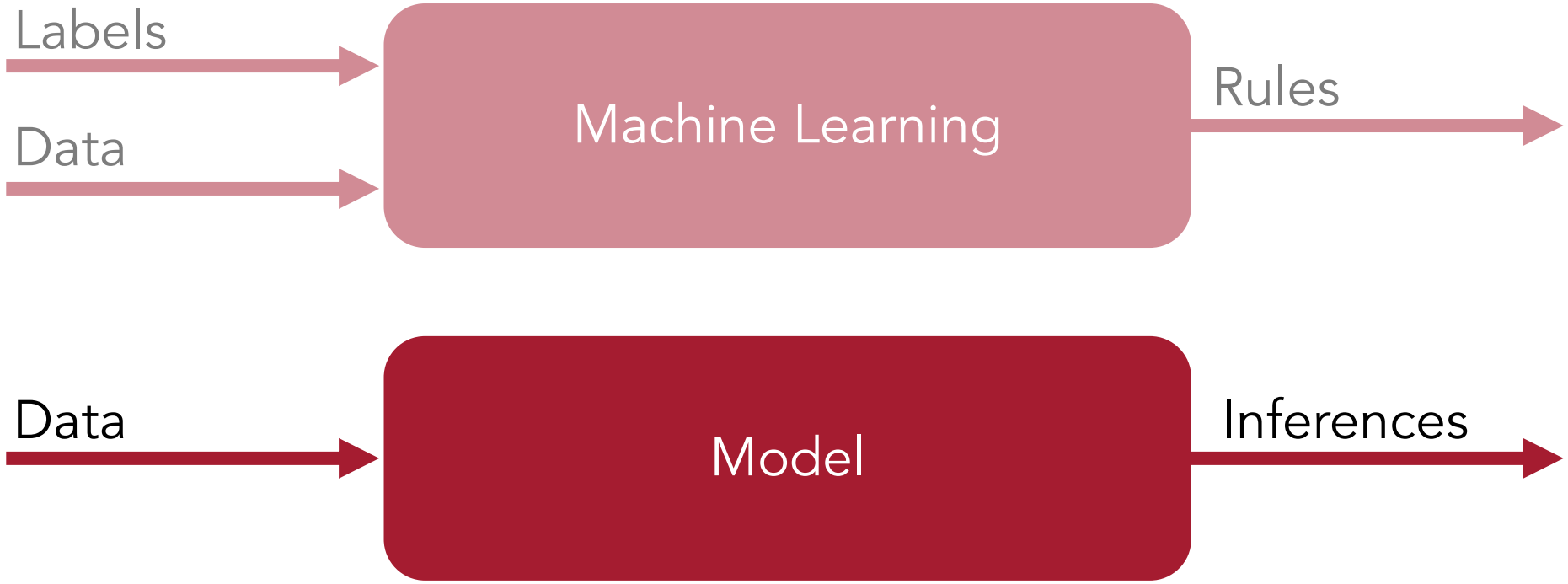


The Machine Learning Paradigm



The Machine Learning Paradigm

The Machine Learning Paradigm



How good is your model?

a way to measure your accuracy

Matching X to Y

X = { -1, 0, 1, 2, 3, 4 }

Y = { -3, -1, 1, 3, 5, 7 }

Make a guess!

$Y = 3X - 1$

$X = \{ -1, 0, 1, 2, 3, 4 \}$

My $Y = \{ -4, -1, 2, 5, 8, 11 \}$

How good is the guess?

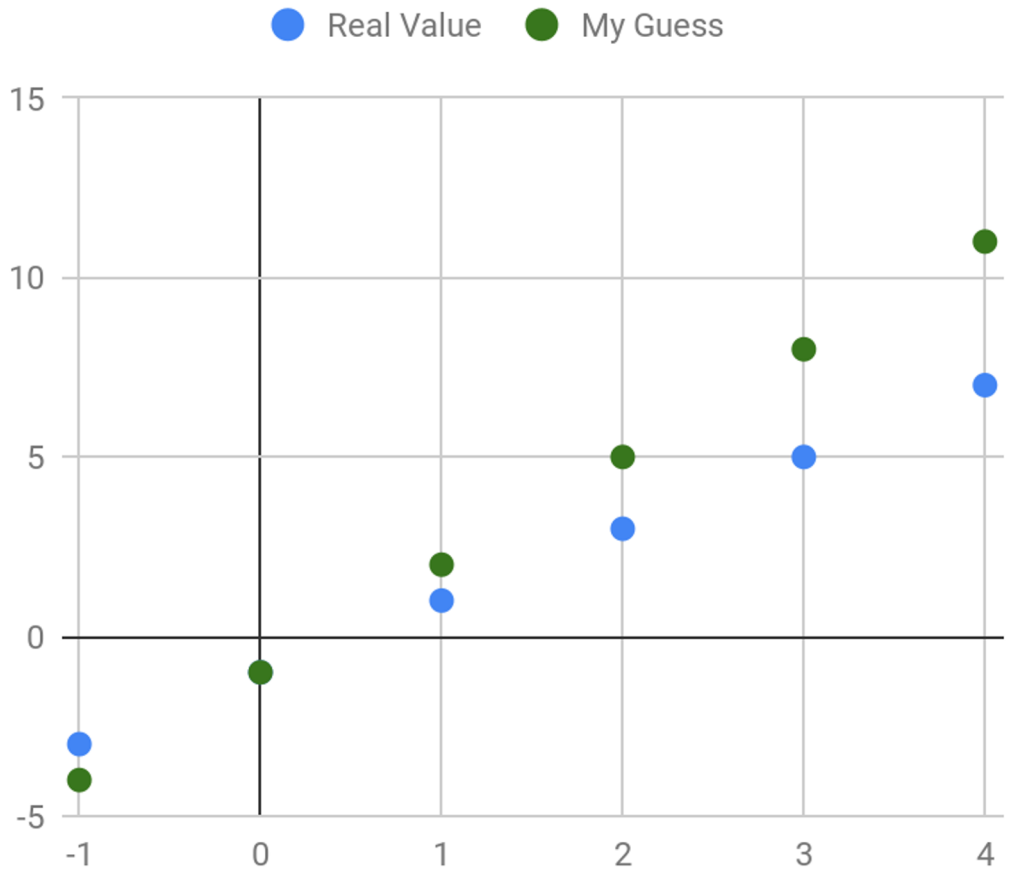
$Y = 3X - 1$

$X = \{ -1, 0, 1, 2, 3, 4 \}$

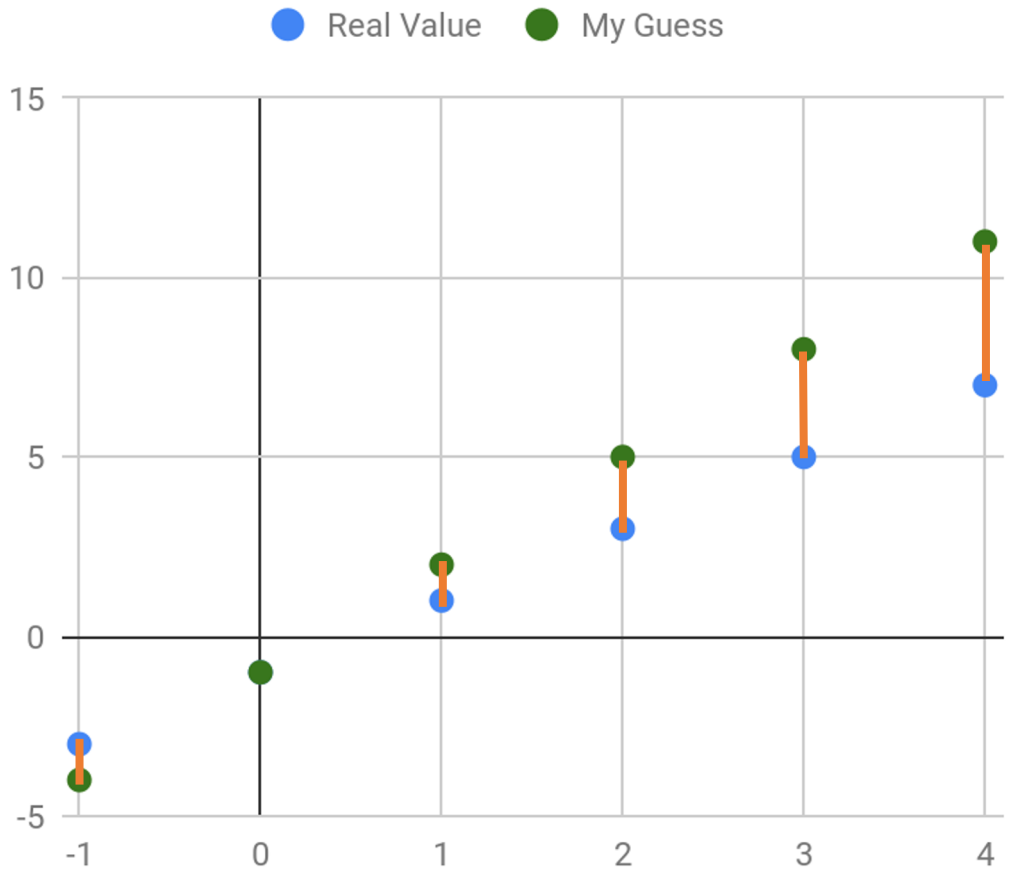
My $Y = \{ -4, -1, 2, 5, 8, 11 \}$

Real $Y = \{ -3, -1, 1, 3, 5, 7 \}$

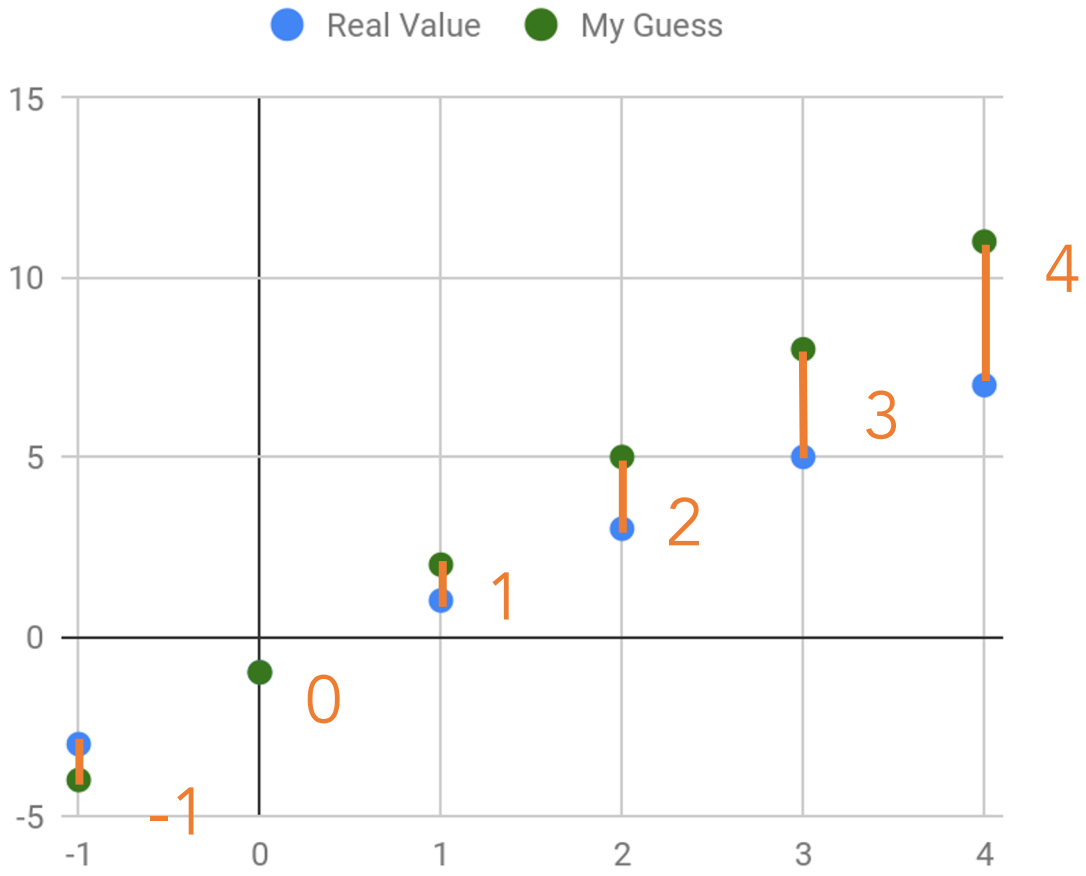
Let’s measure it!



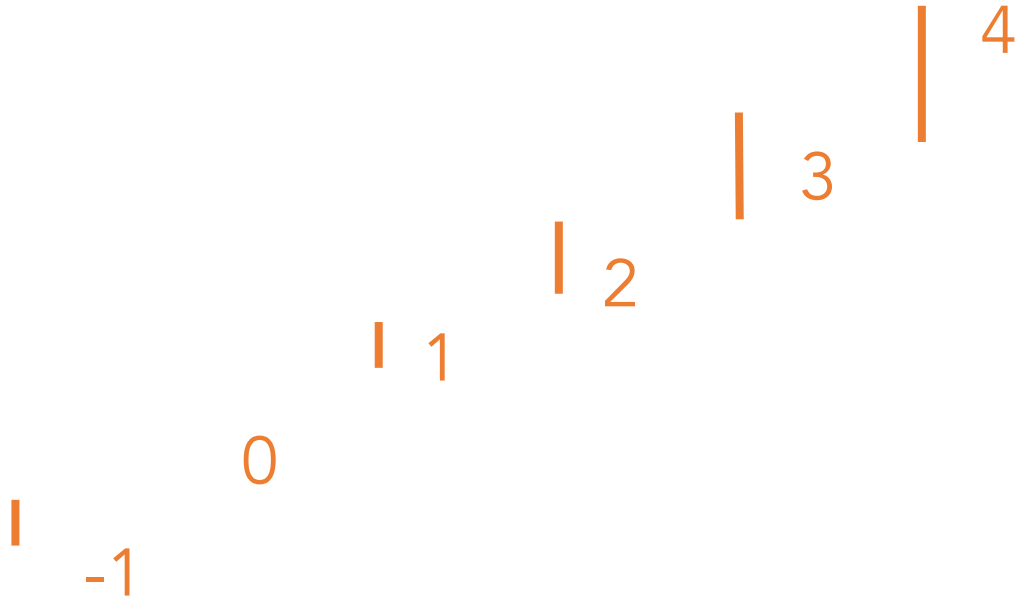
Let’s measure it!



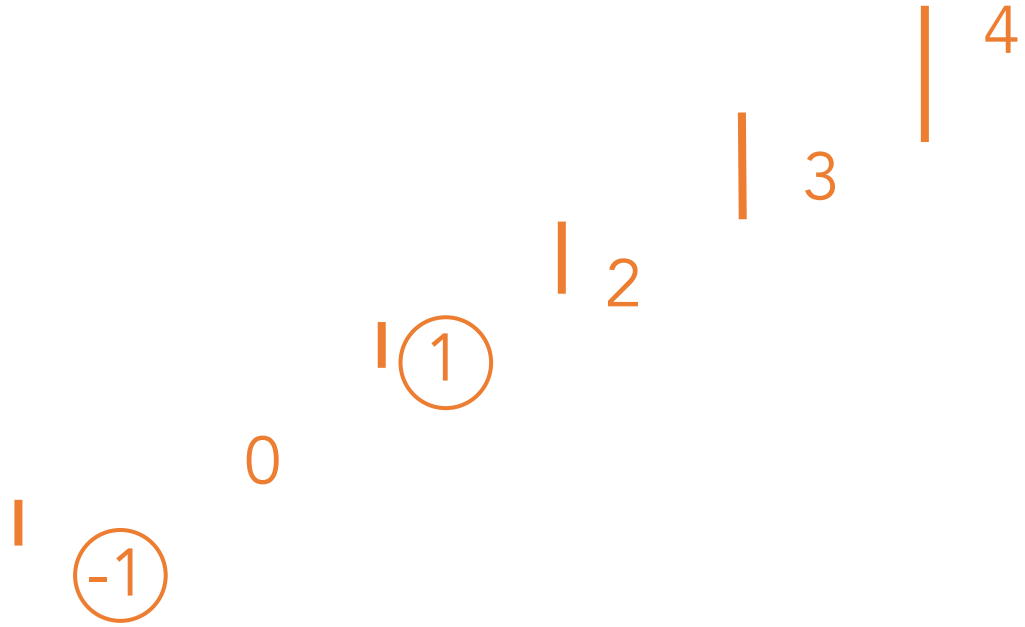
Let’s measure it!



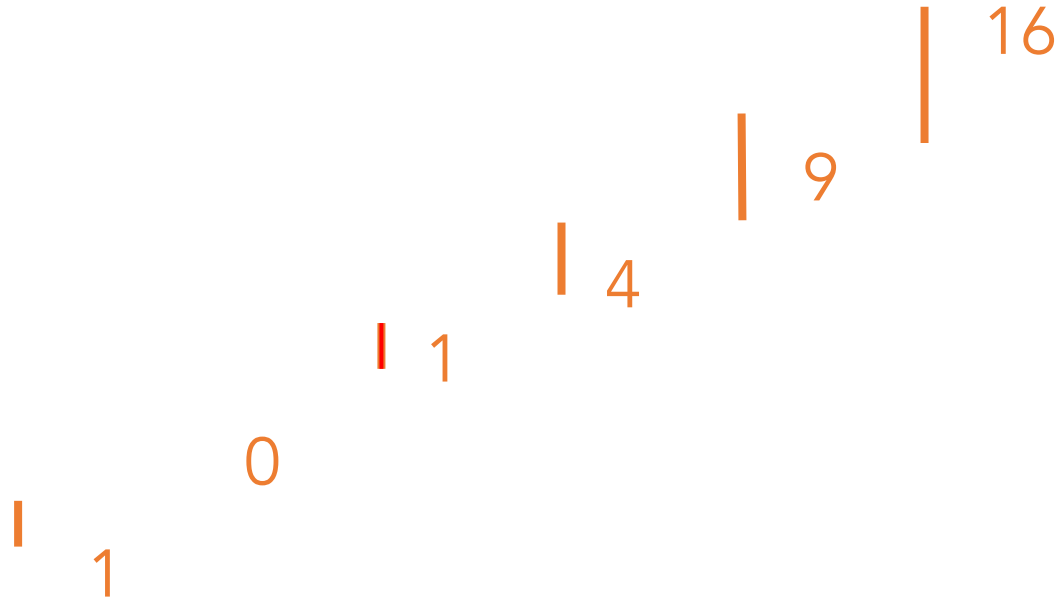
Let’s measure it!



Houston, we have a problem!



What if we **square**^{2} them?



Total that (∑) and take the square root √

sqrt(1 + 1 + 4 + 9 + 16)

= sqrt(31)
= 5.57

Make another guess!

$Y = 2X - 2$

$X = \{ -1, 0, 1, 2, 3, 4 \}$

My $Y = \{ -4, -2, 0, 2, 4, 6 \}$

Real $Y = \{-3, -1, 1, 3, 5, 7\}$

$Diff^2 = \{1, 1, 1, 1, 1\}$

Get the same difference, repeat the same process.

sqrt(1 + 1 + 1 + 1 + 1)

= sqrt(5)
= 2.23

Make another guess!

$Y = 2X - 1$

$X = \{ -1, 0, 1, 2, 3, 4 \}$

My $Y = \{ -3, -1, 1, 3, 5, 7 \}$

Real $Y = \{-3, -1, 1, 3, 5, 7\}$

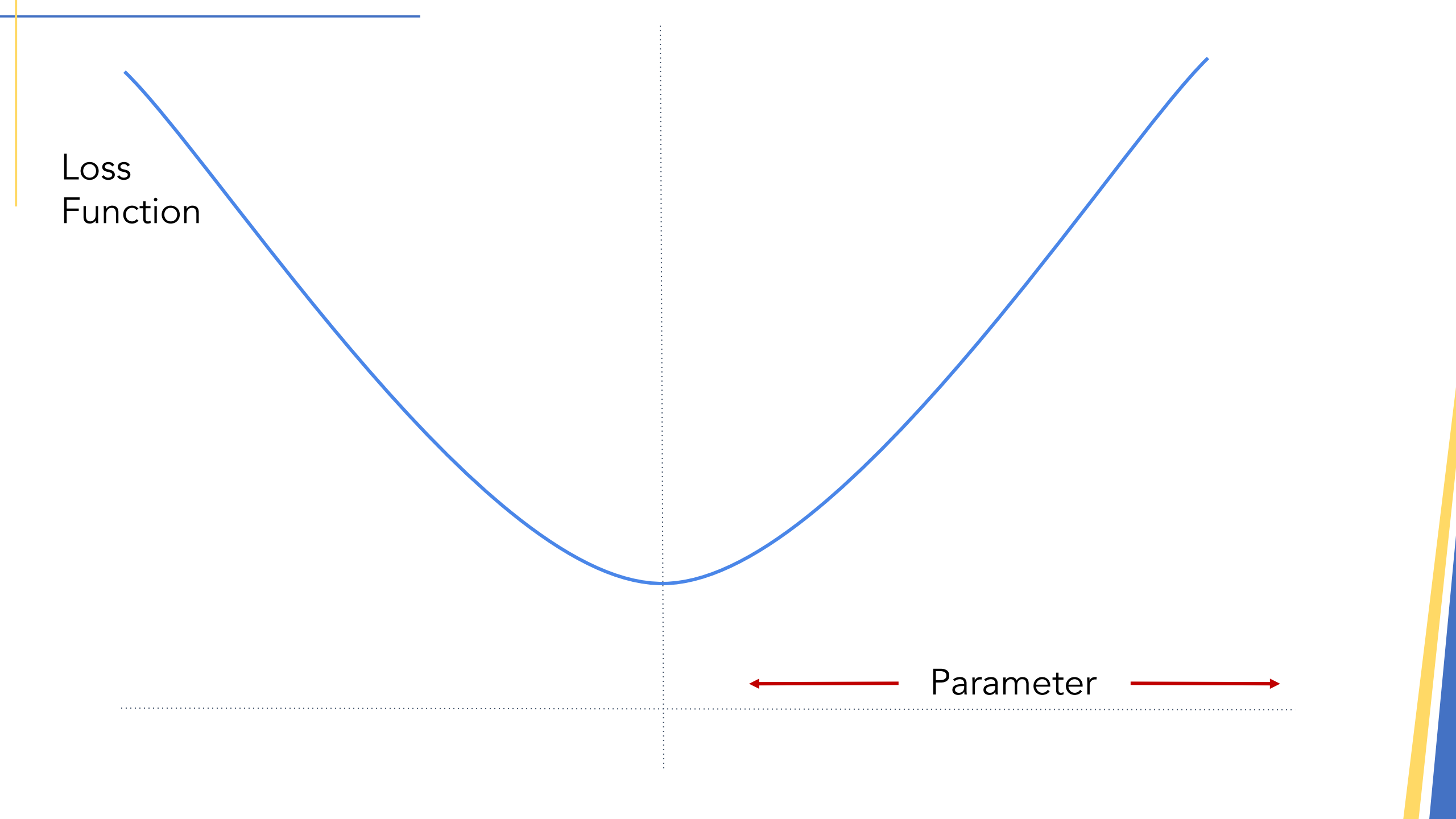
$Diff^2 = \{0, 0, 0, 0, 0\}$

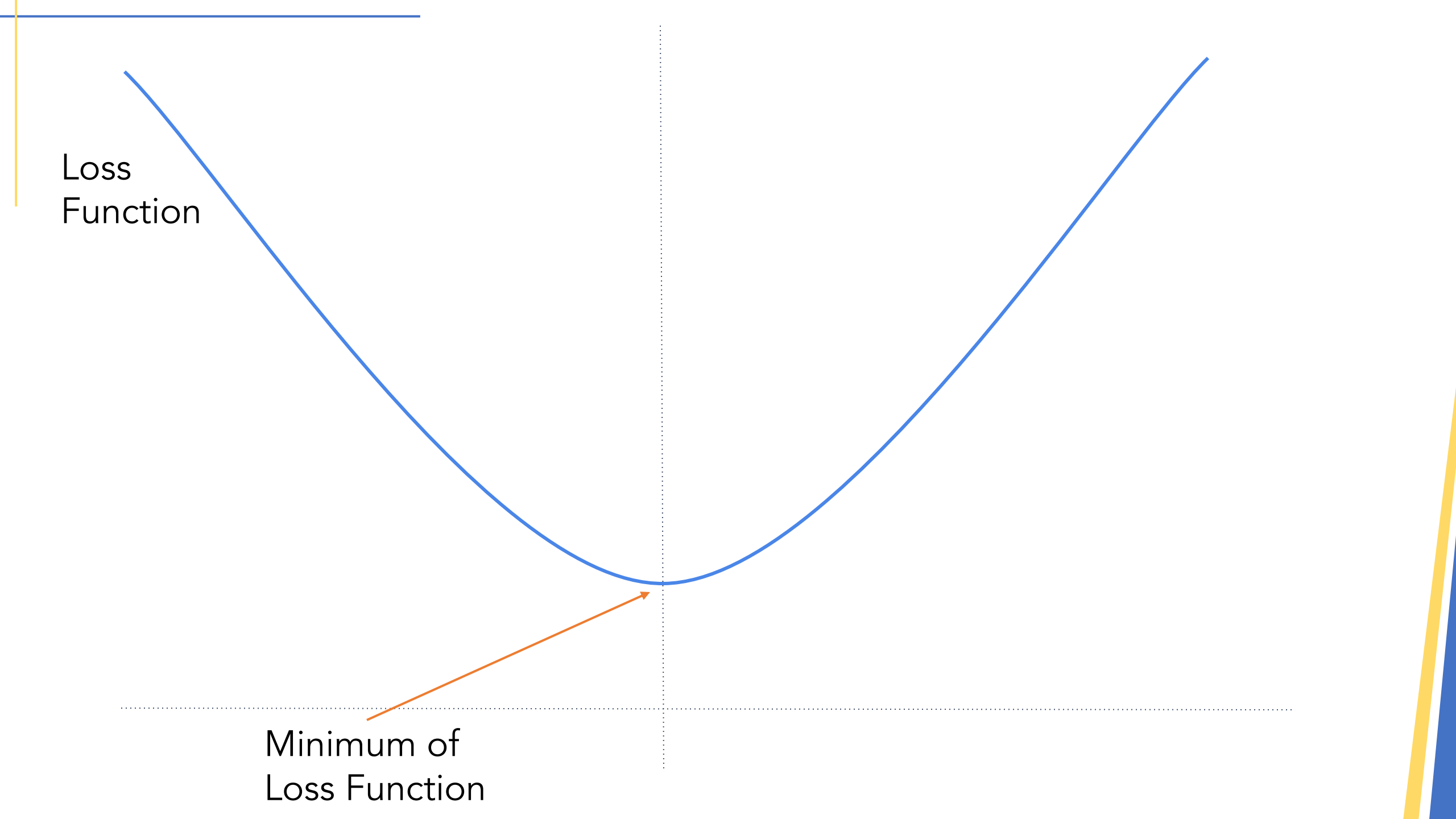
Root-mean-square deviation

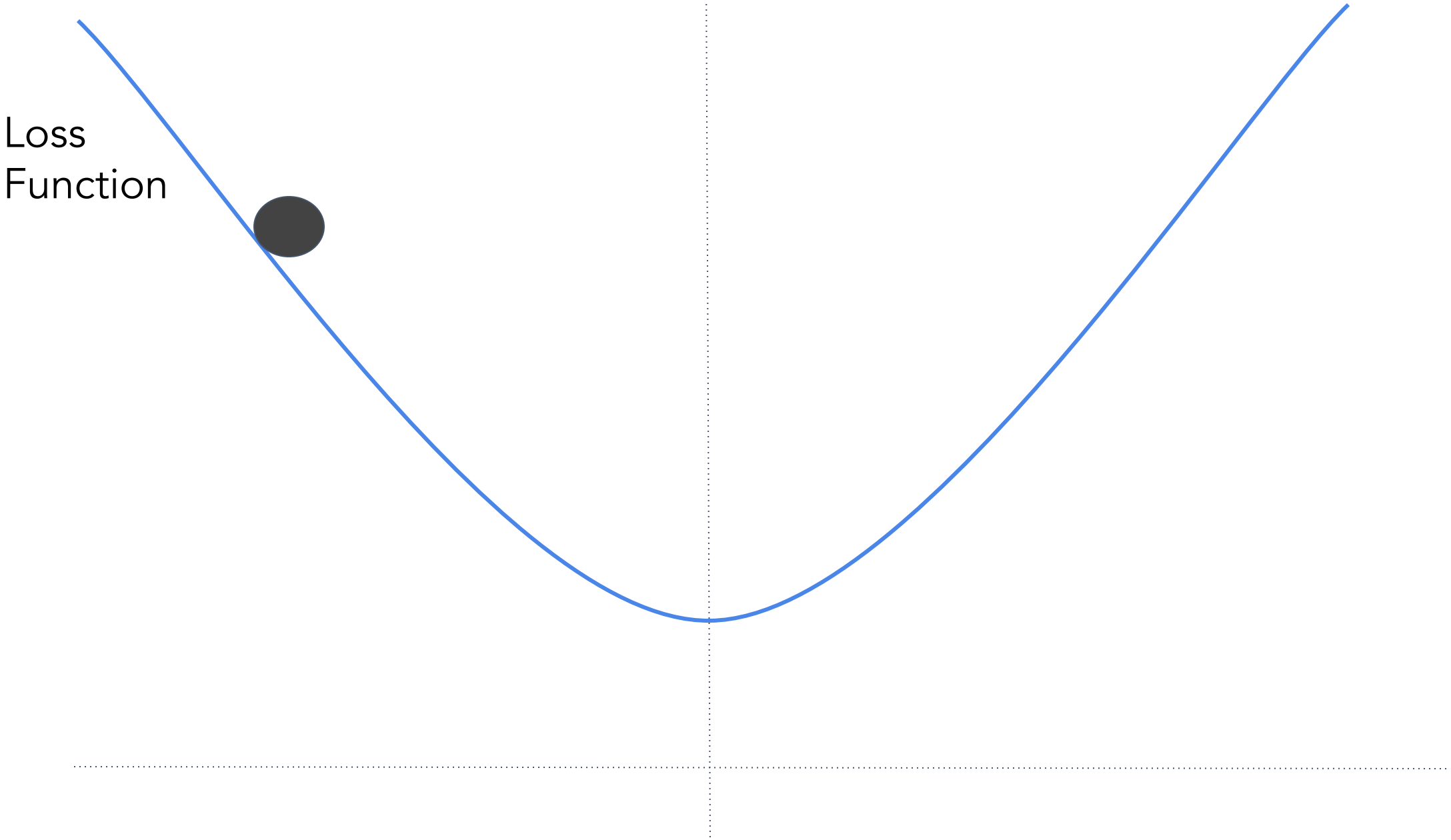
$$\mathrm{RMSD} = \sqrt{\frac{\sum_{t=1}^{T} (\hat{y}_t - y_t)^2}{T}}.$$

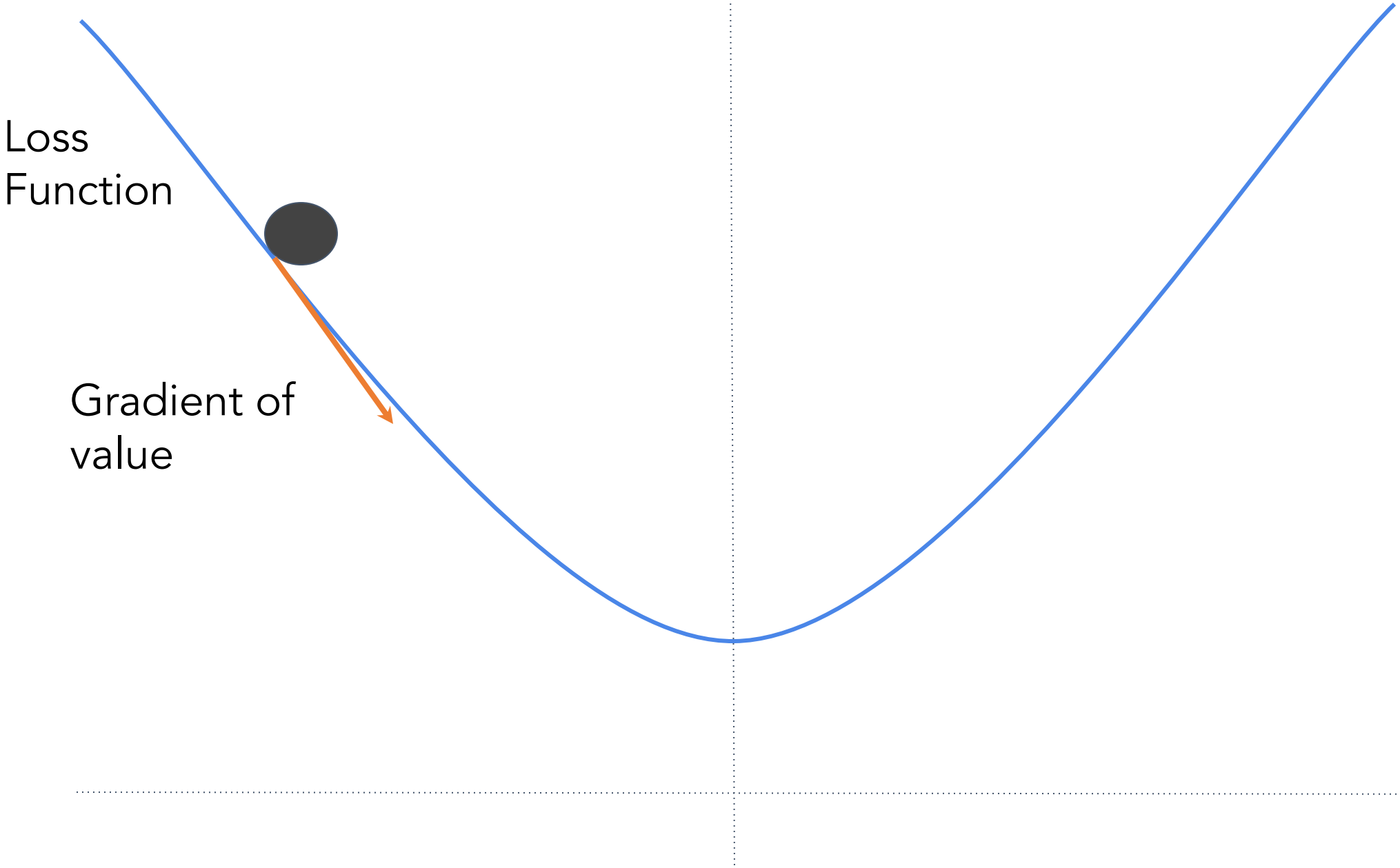
Finding out the best solution

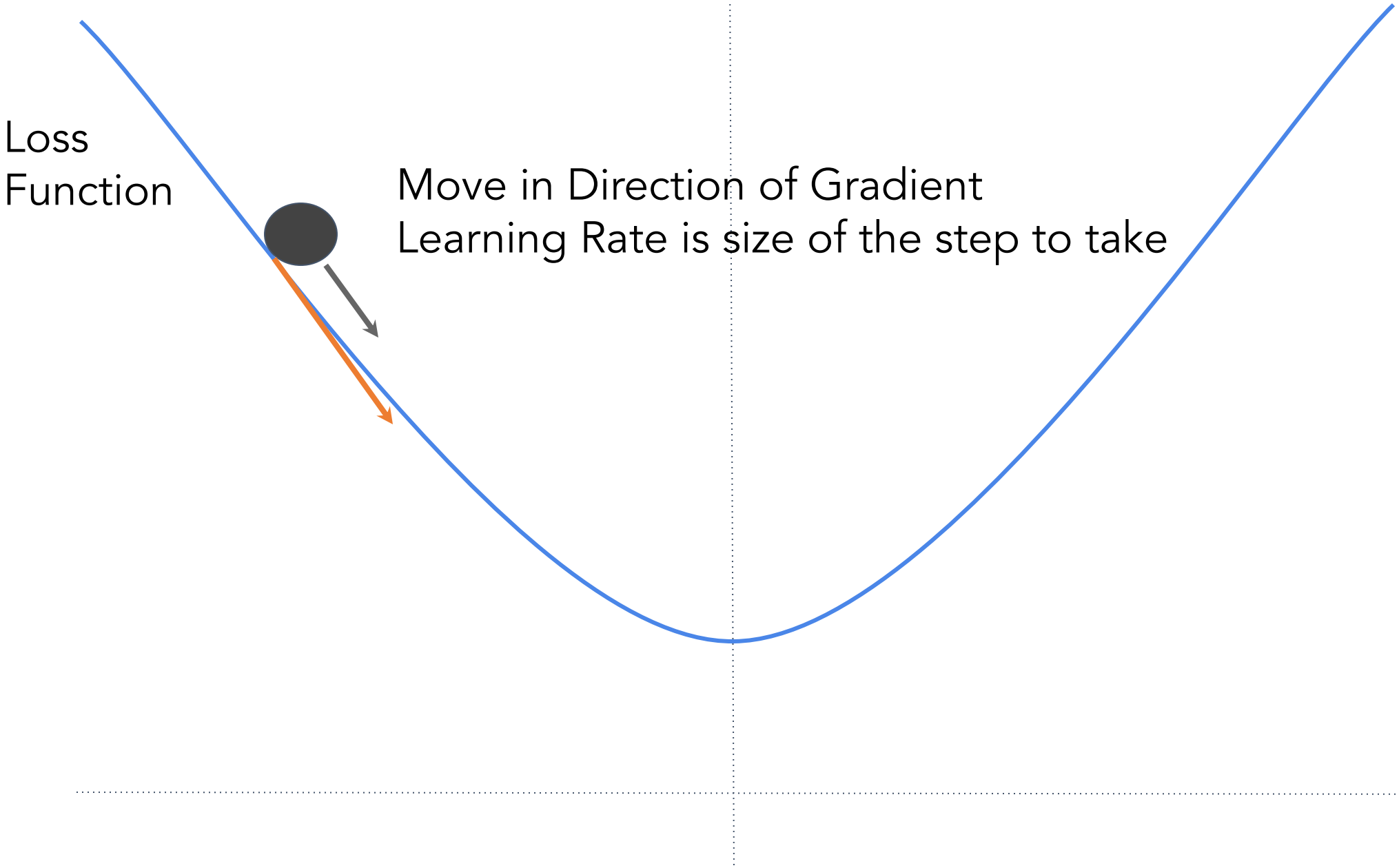
Trial and error approach

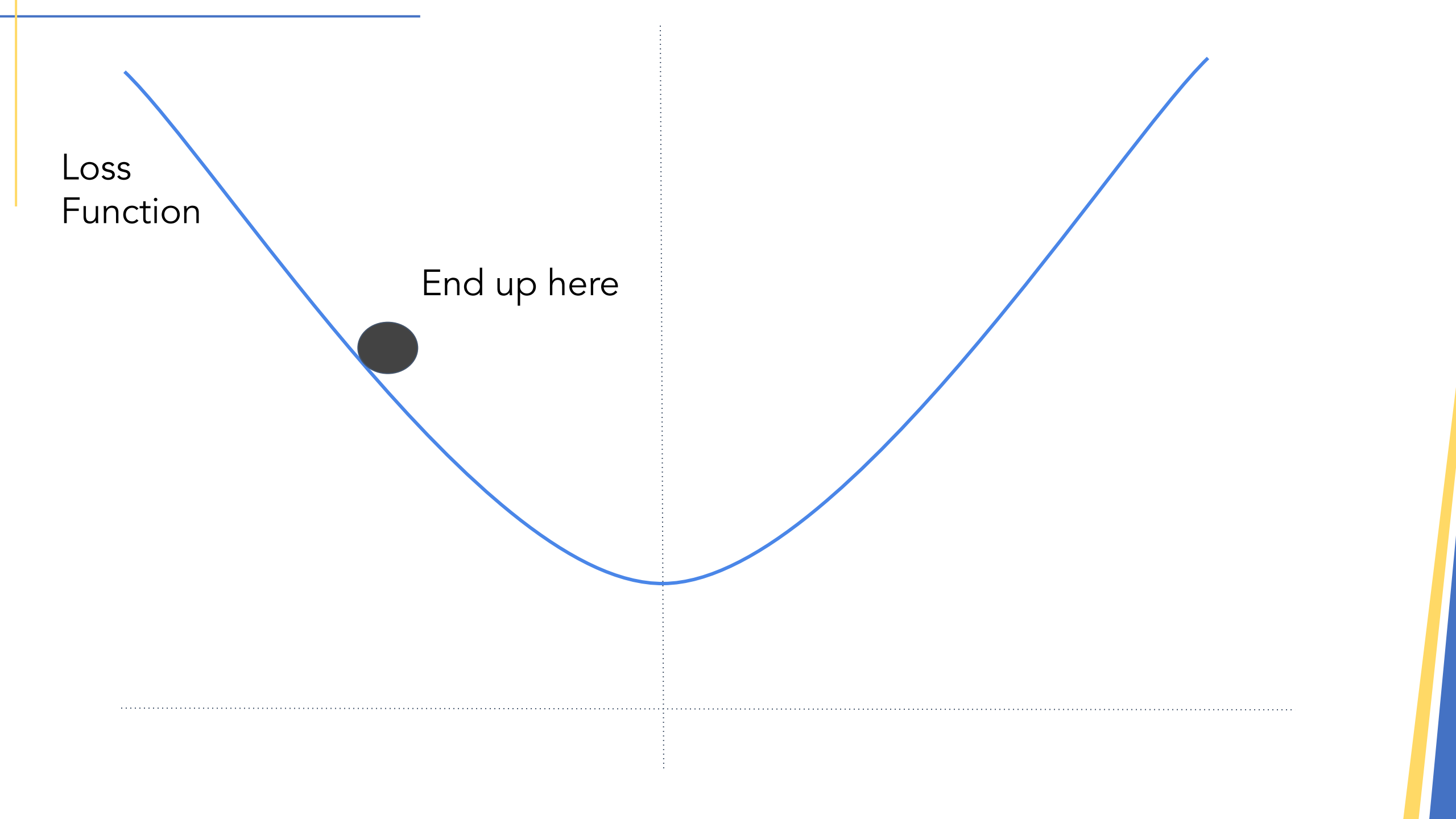

Loss
Function
Parameter

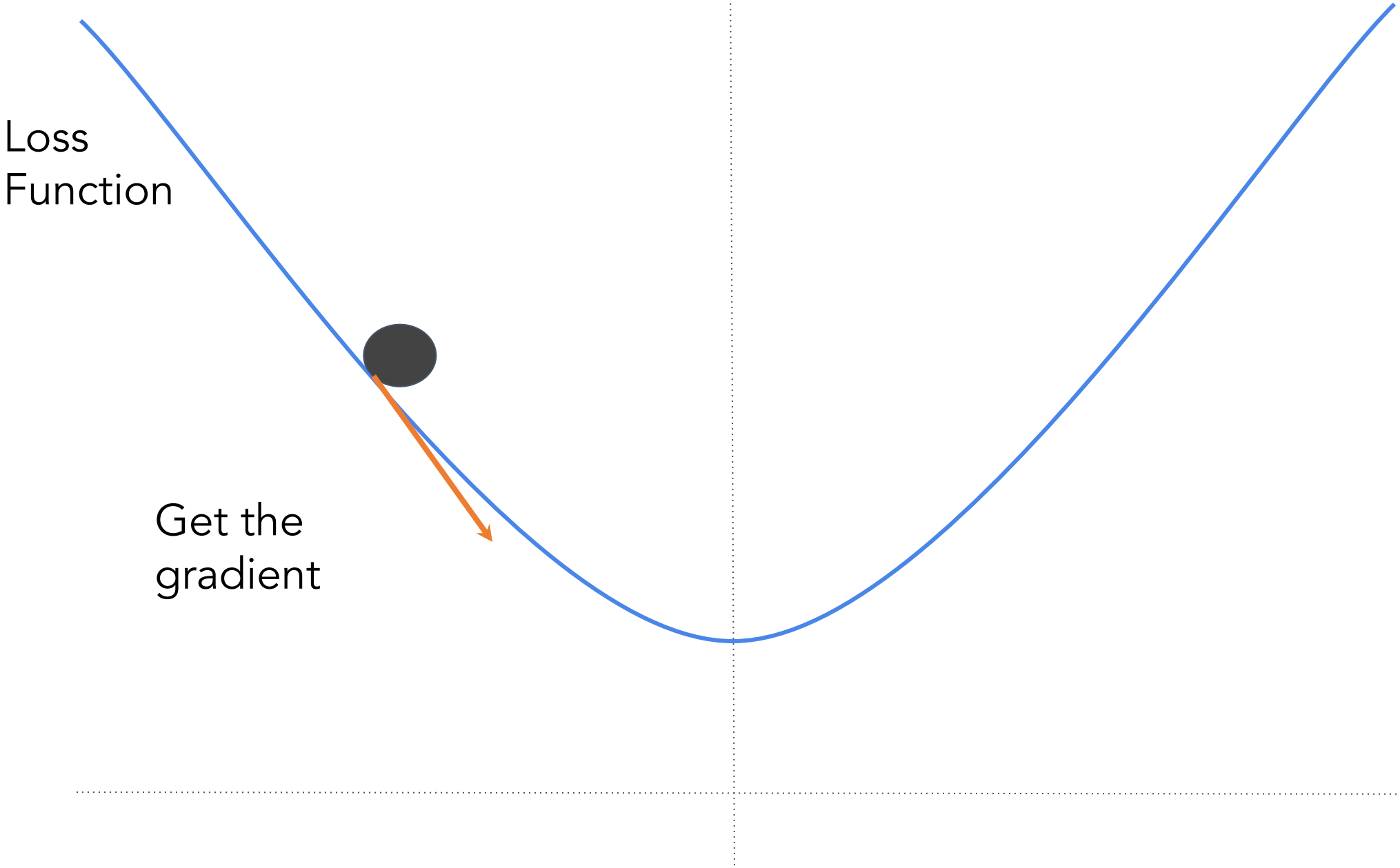

Loss
Function
Minimum of
Loss Function

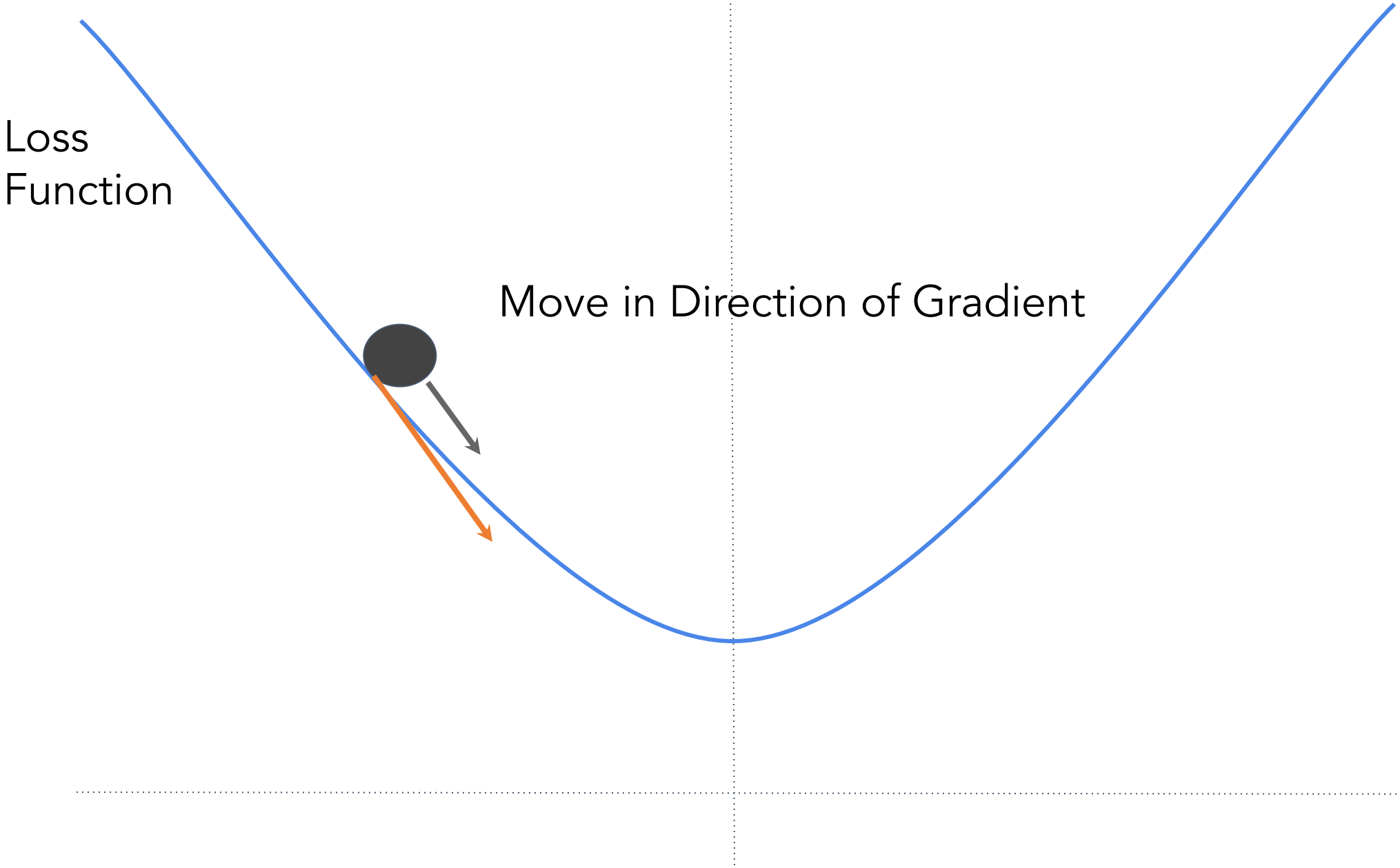

Loss
Function


Loss
Function
Gradient of
value


Loss
Function
Move in Direction of Gradient
Learning Rate is size of the step to take

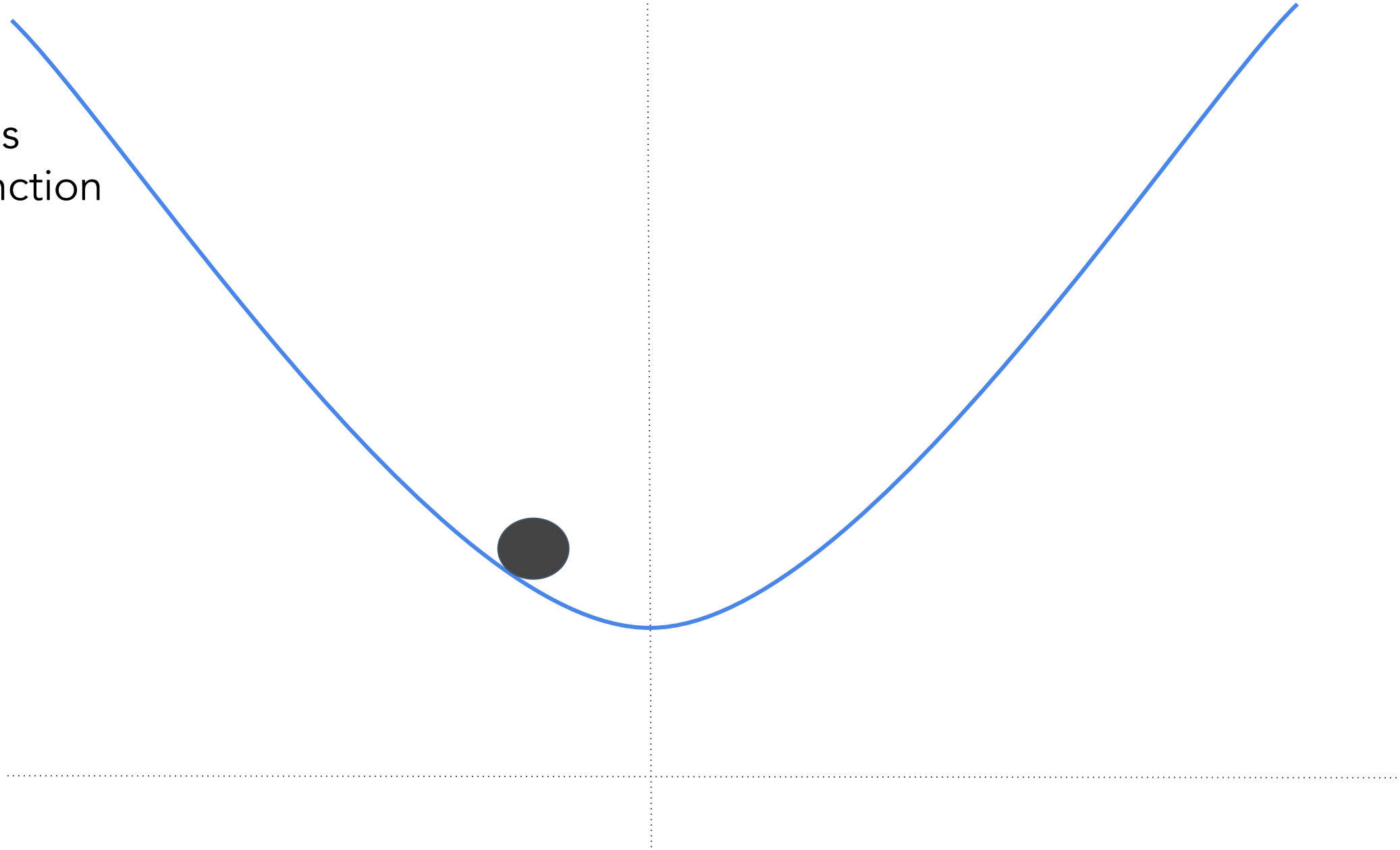

Loss
Function
End up here


Loss
Function
Get the
gradient

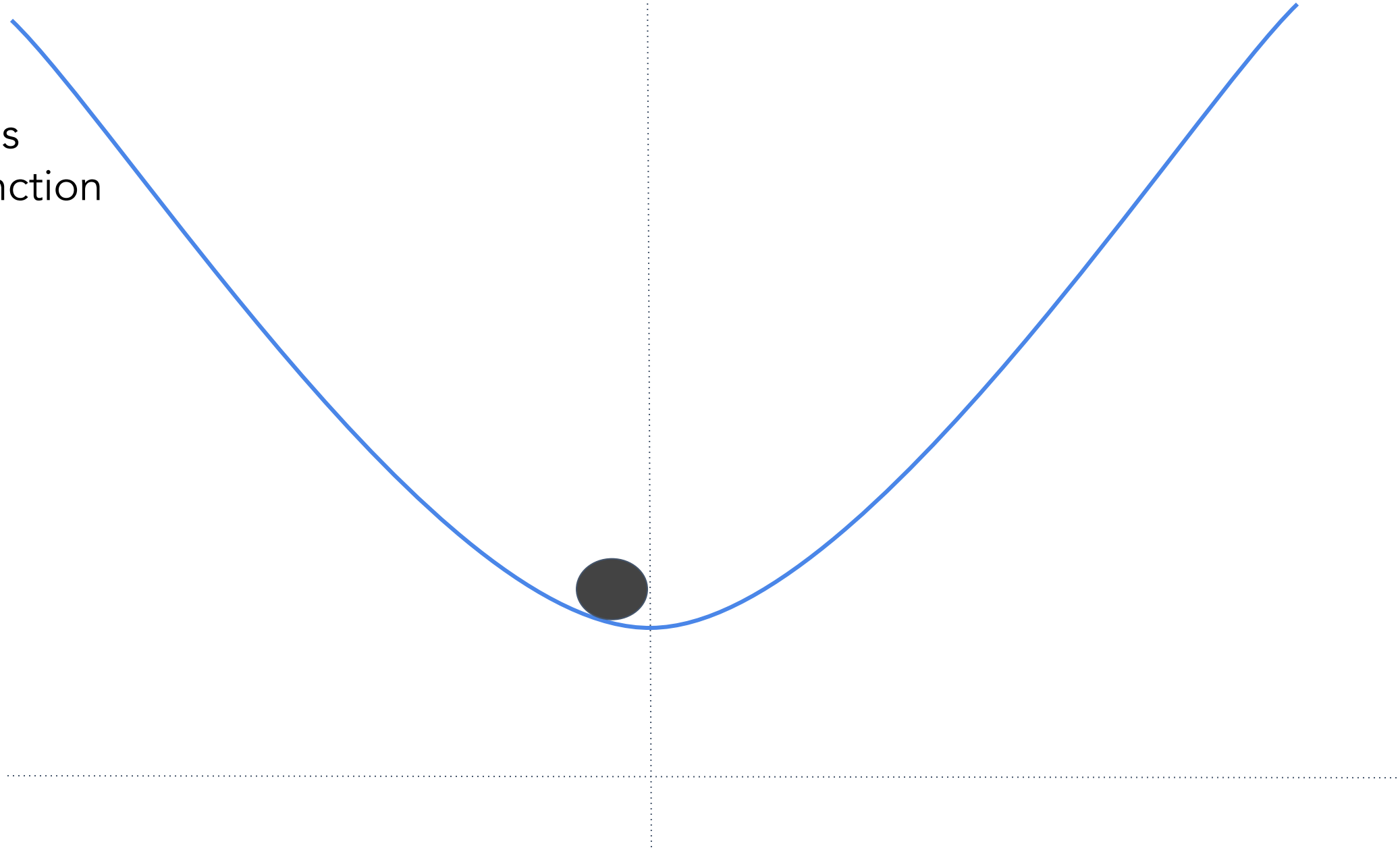

Loss
Function
Move in Direction of Gradient


Loss
Function
End Up here

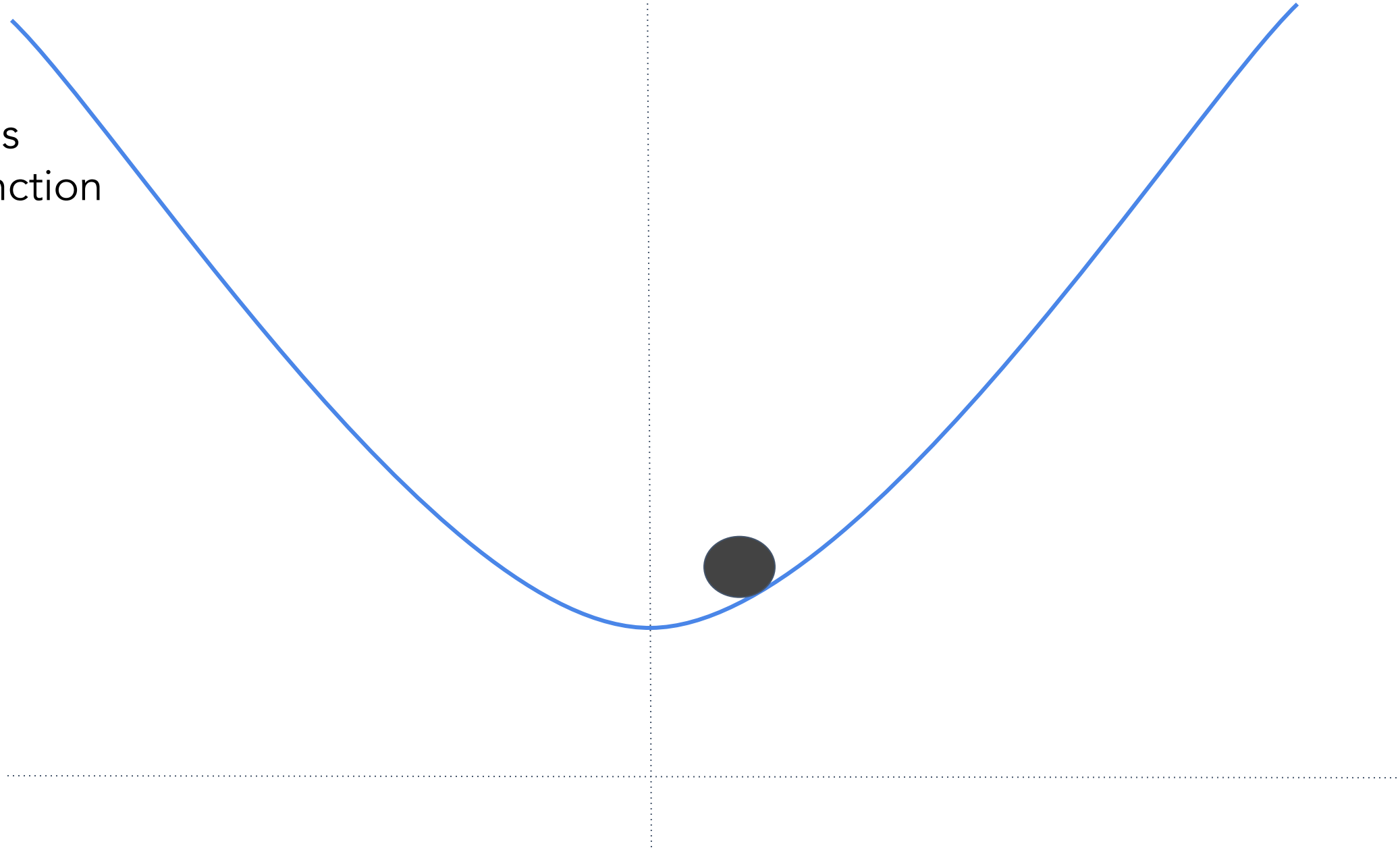
Loss
Function



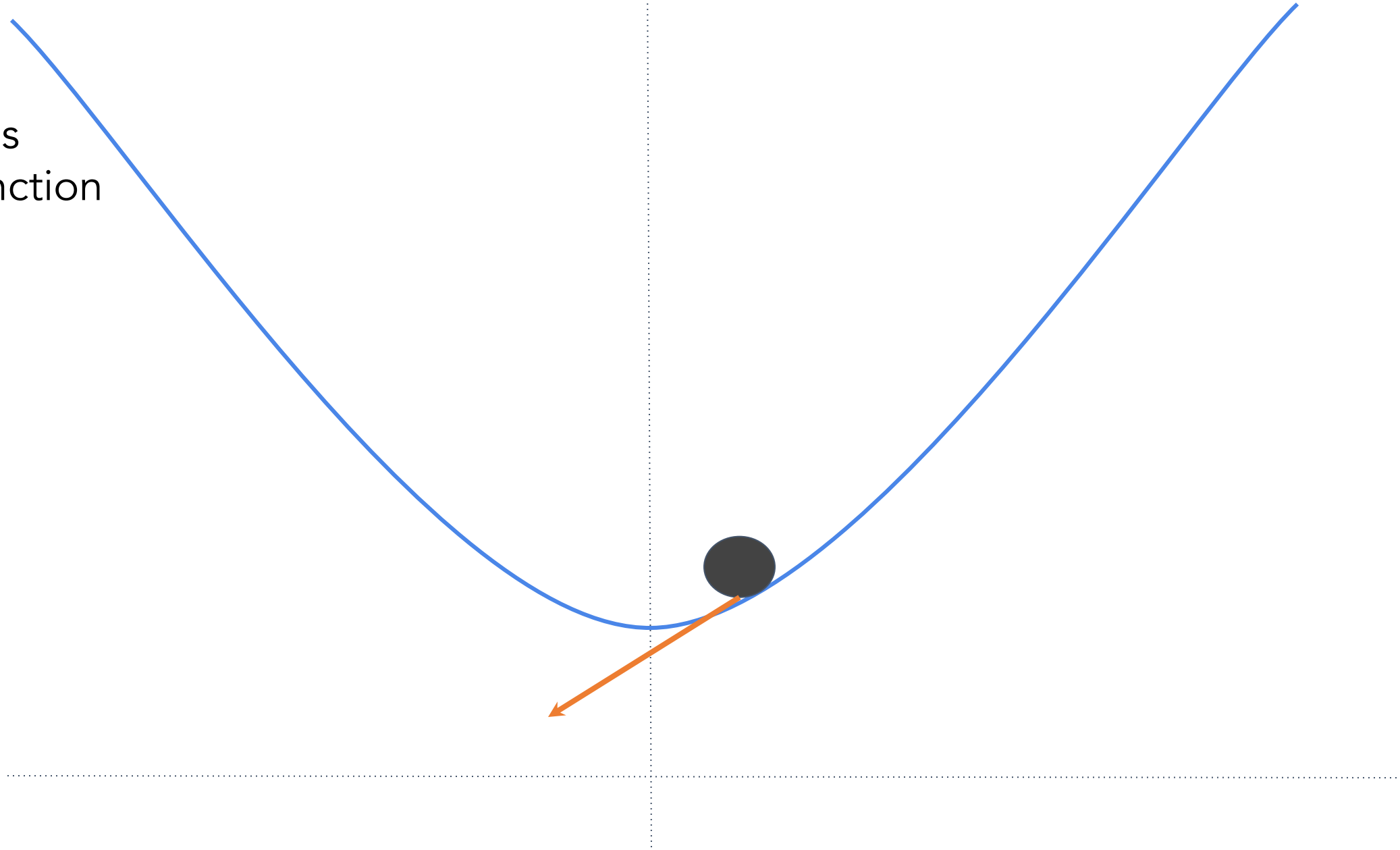
Loss
Function



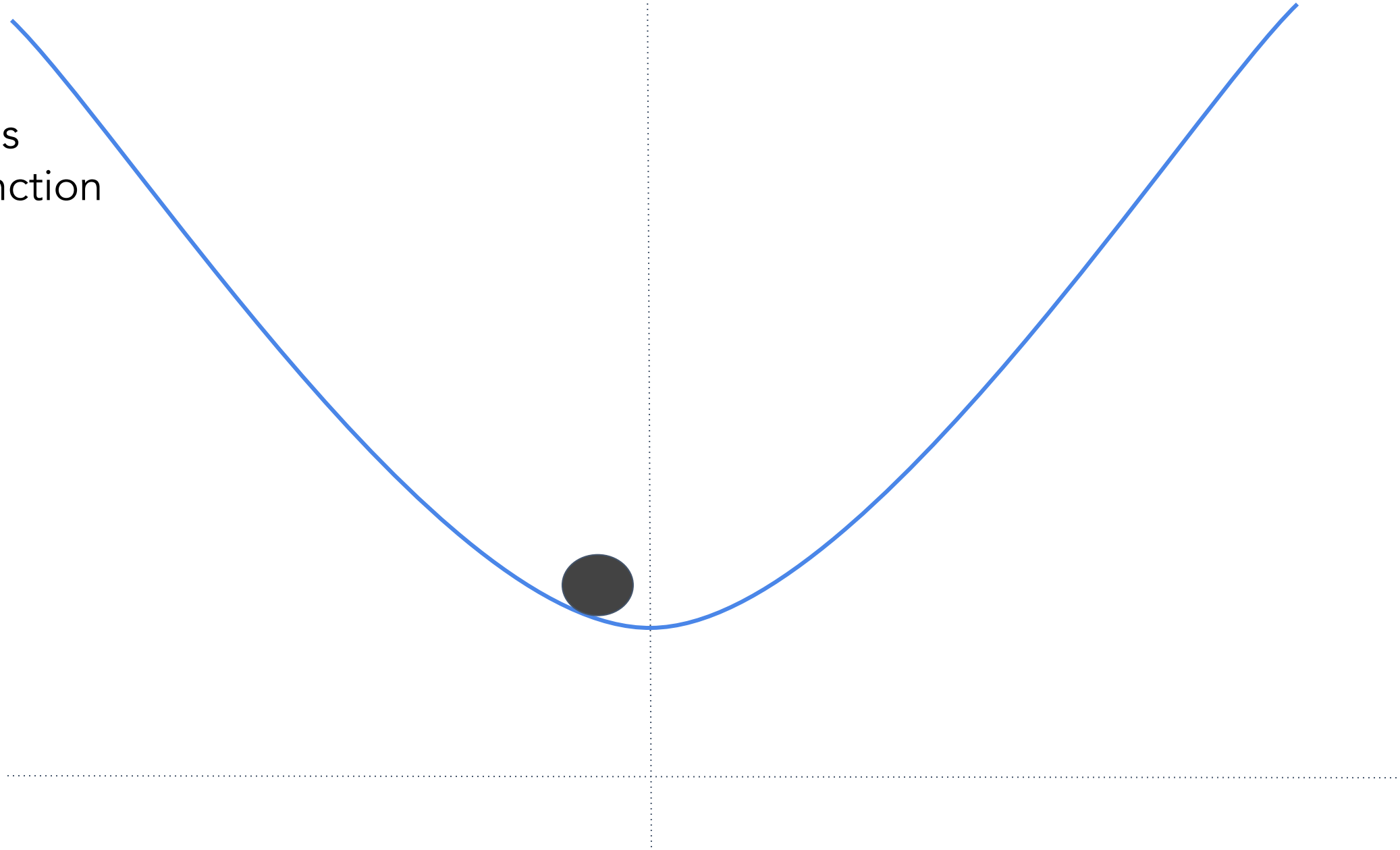
Loss
Function



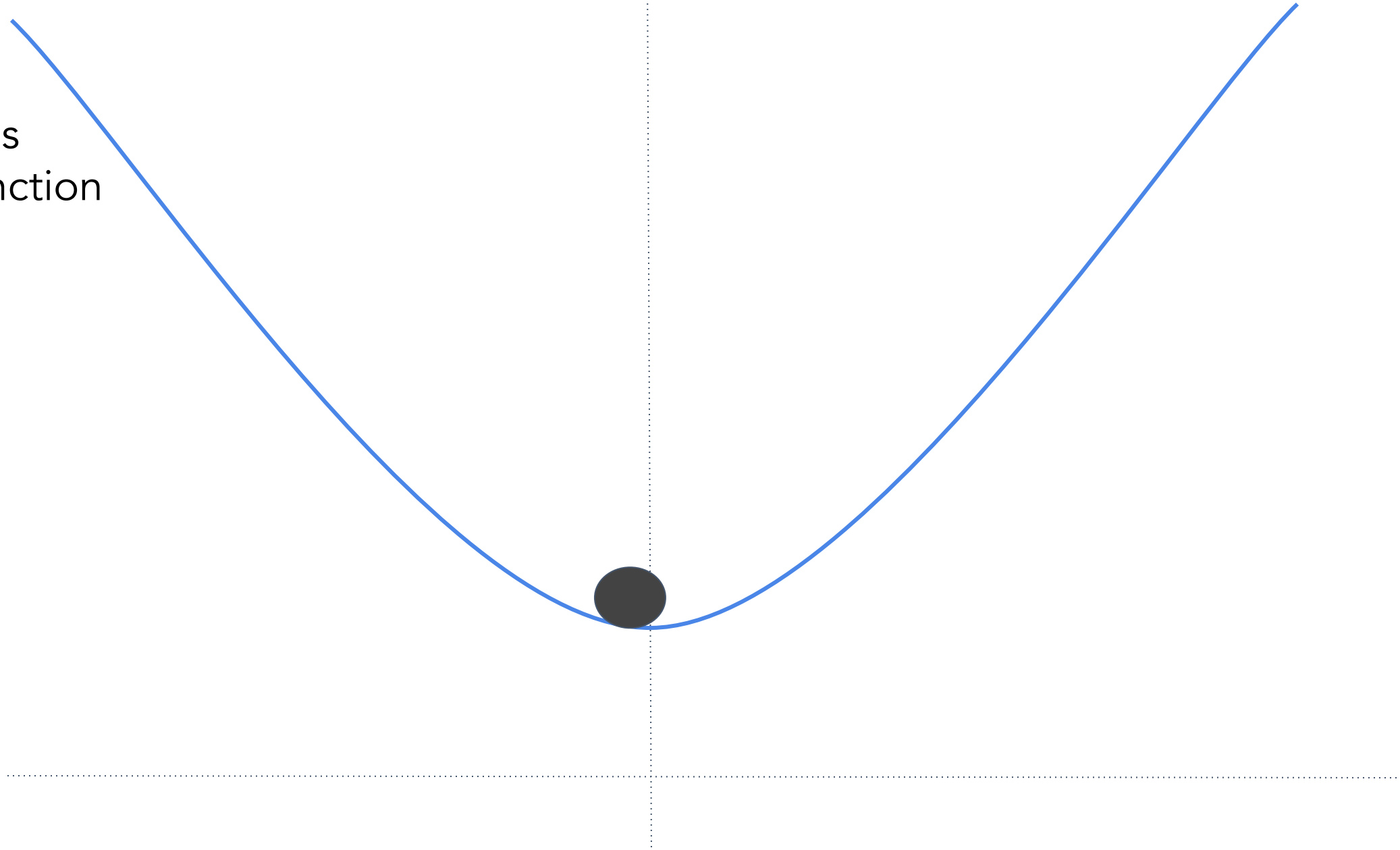
Loss
Function



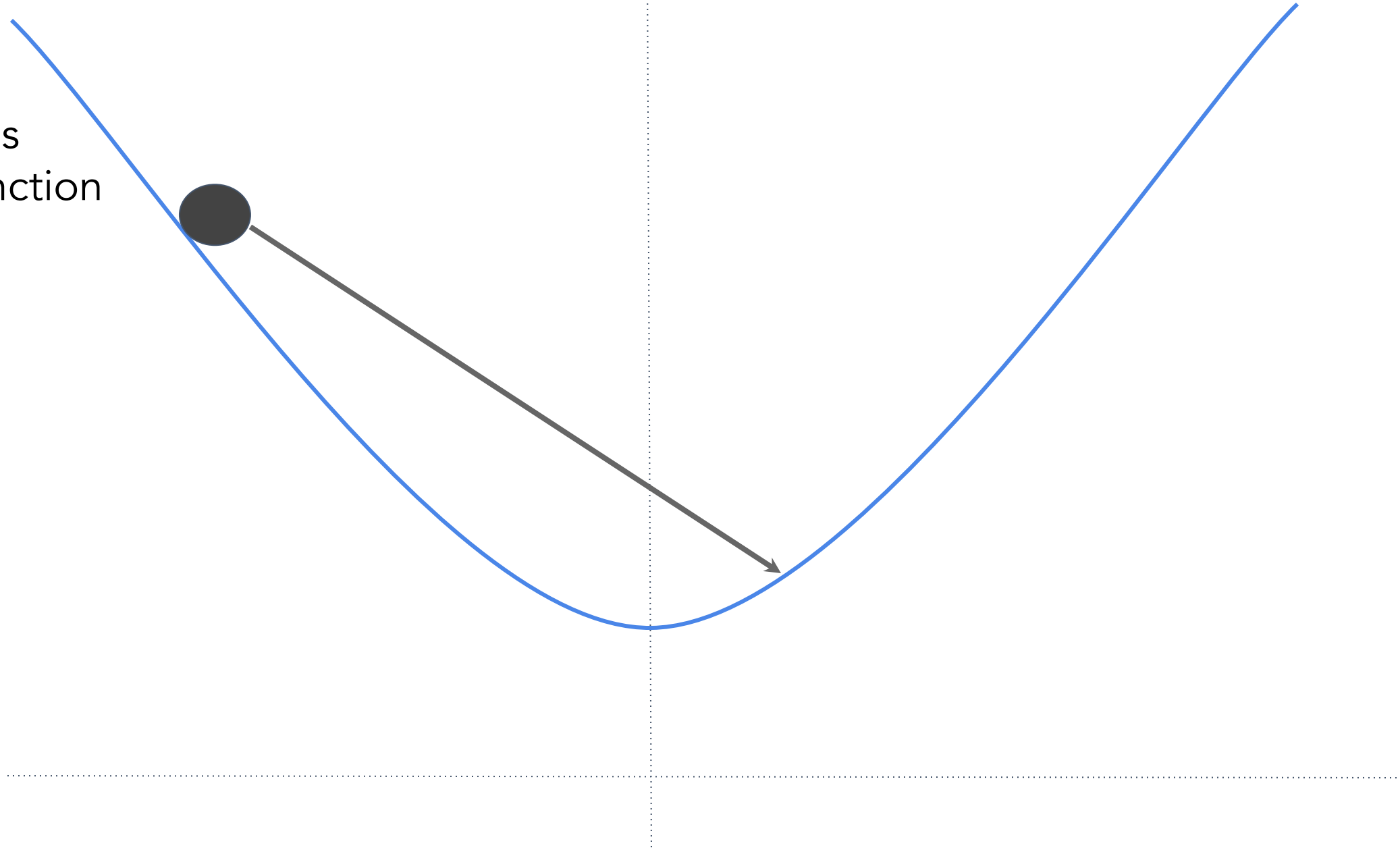
Loss
Function



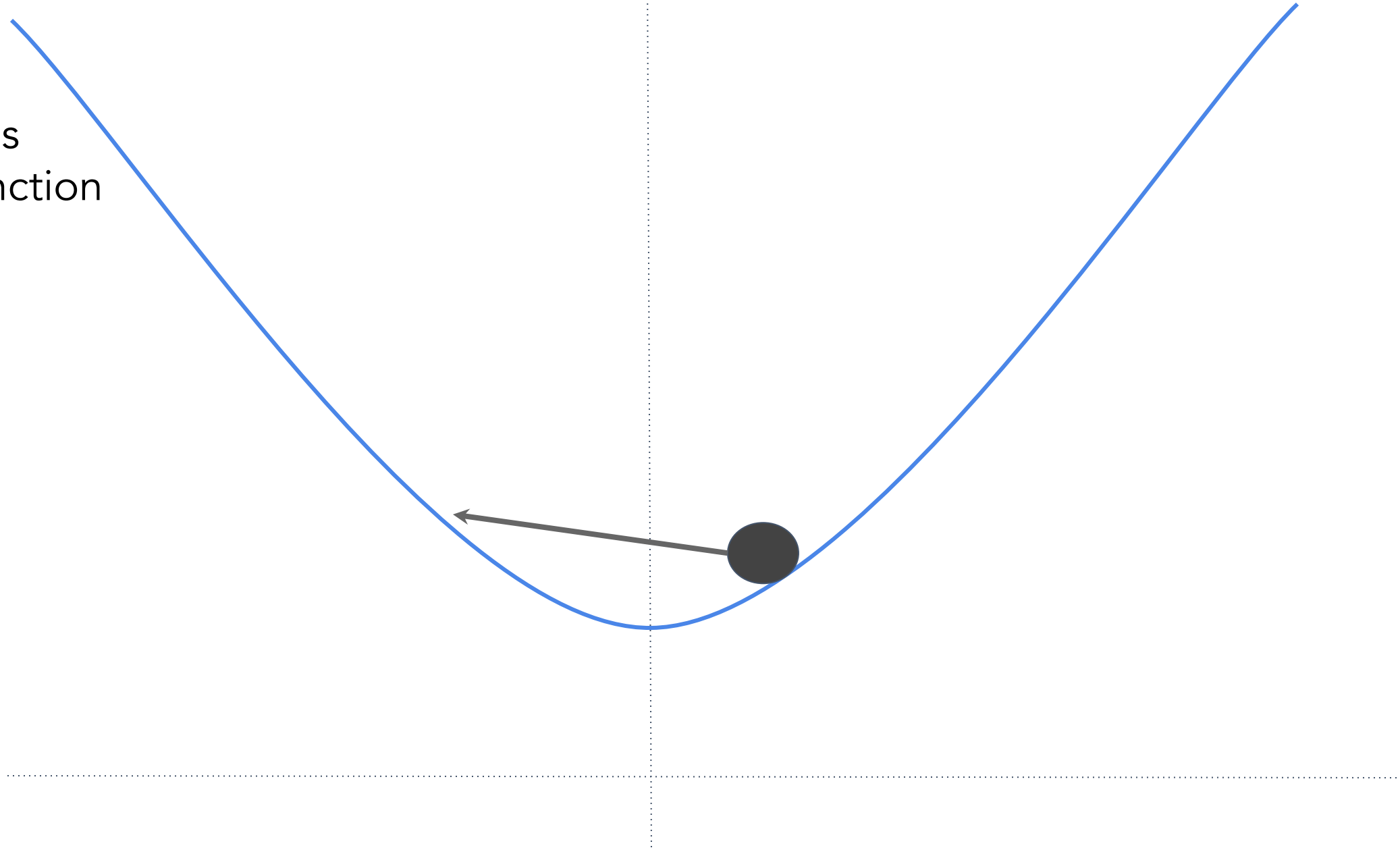
Loss
Function



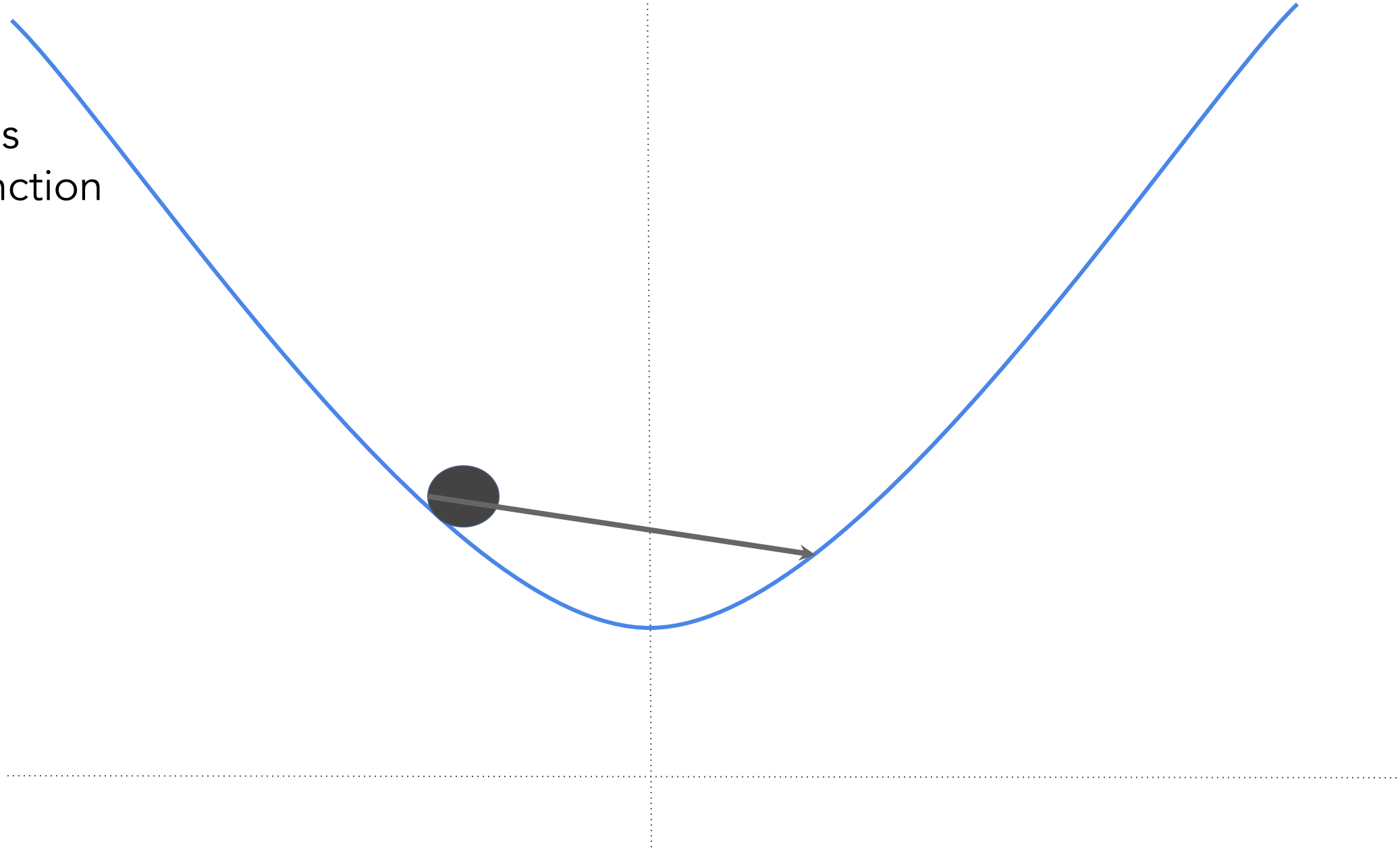
Loss Function

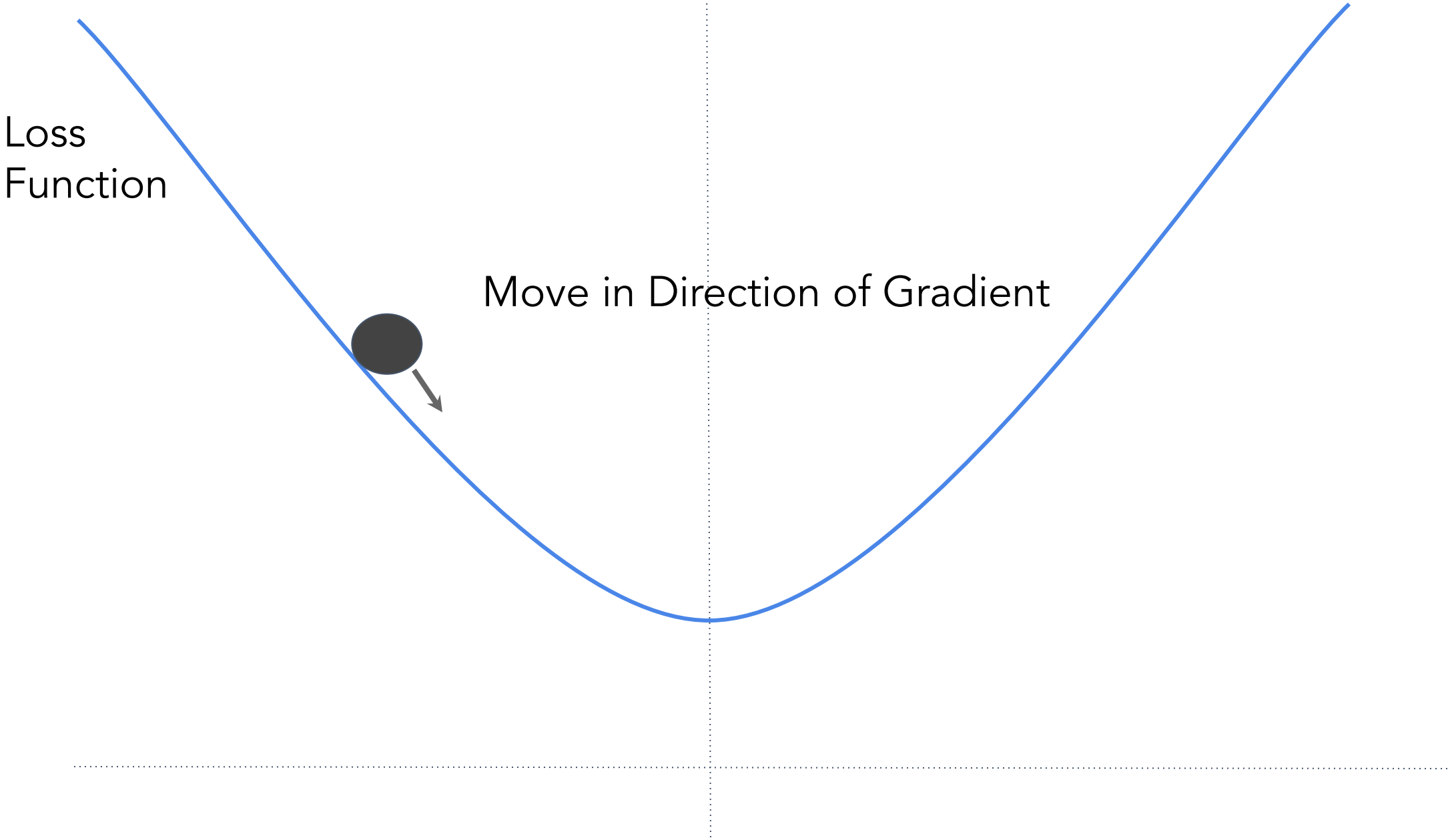


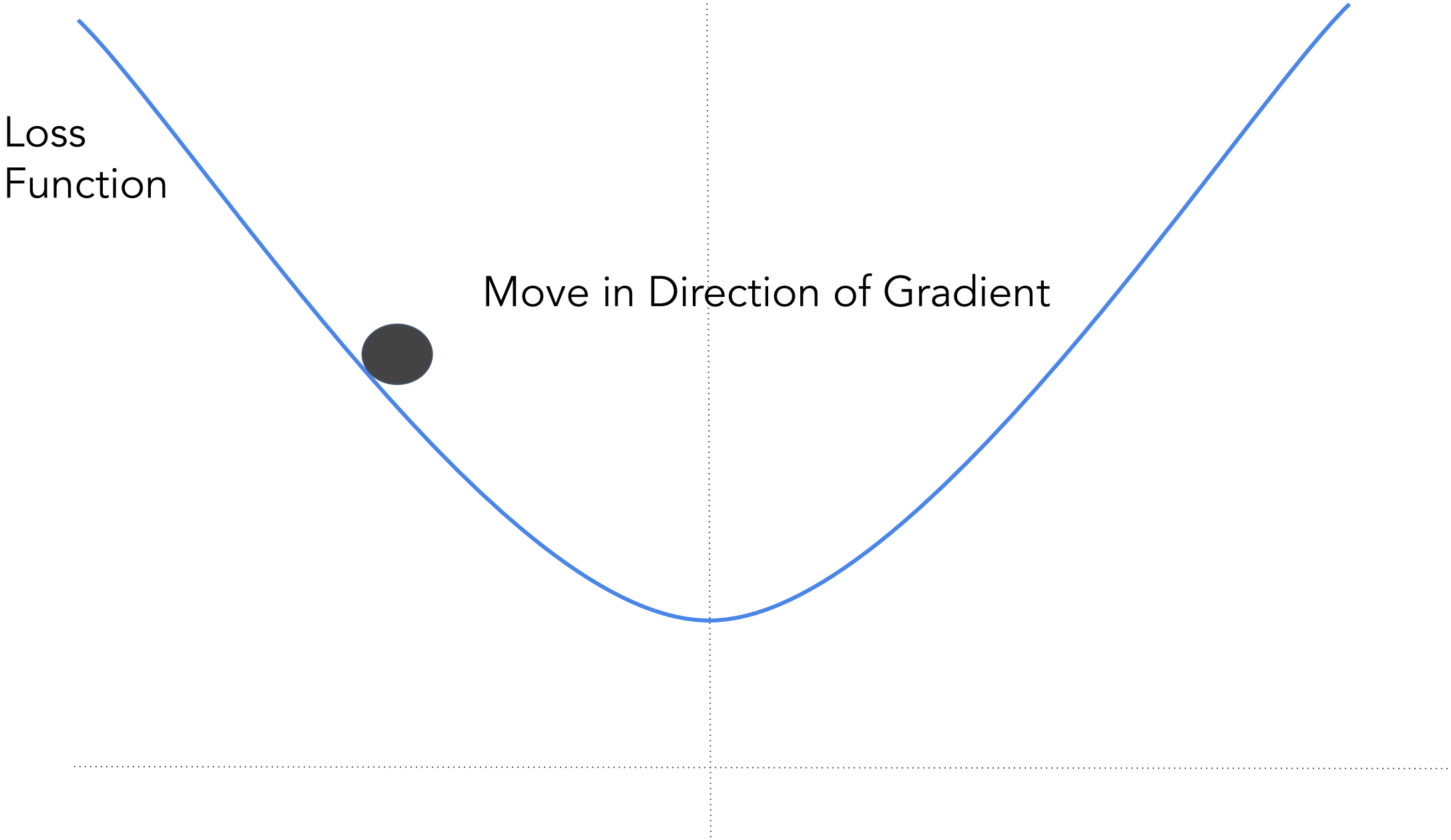
Loss
Function

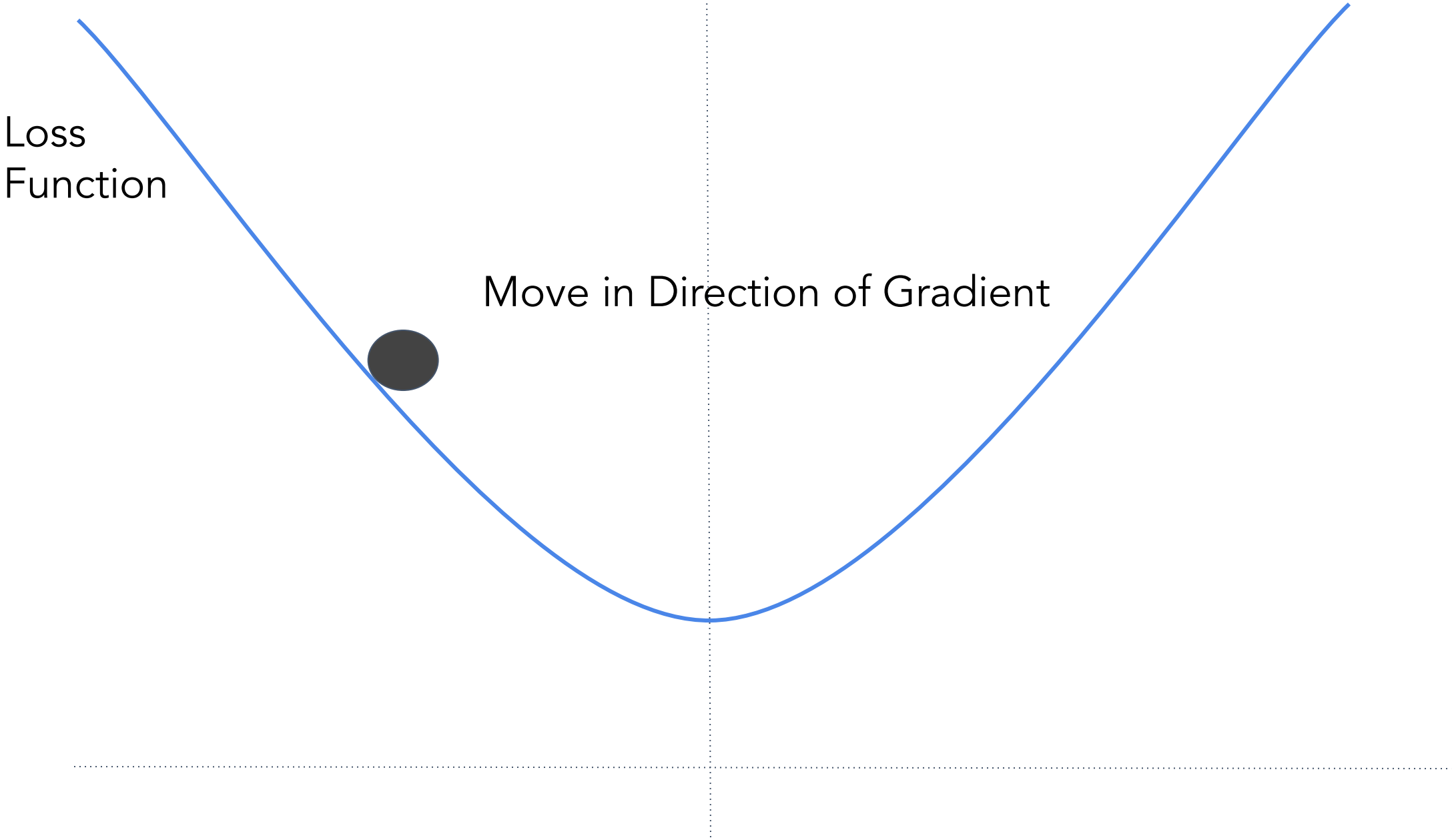


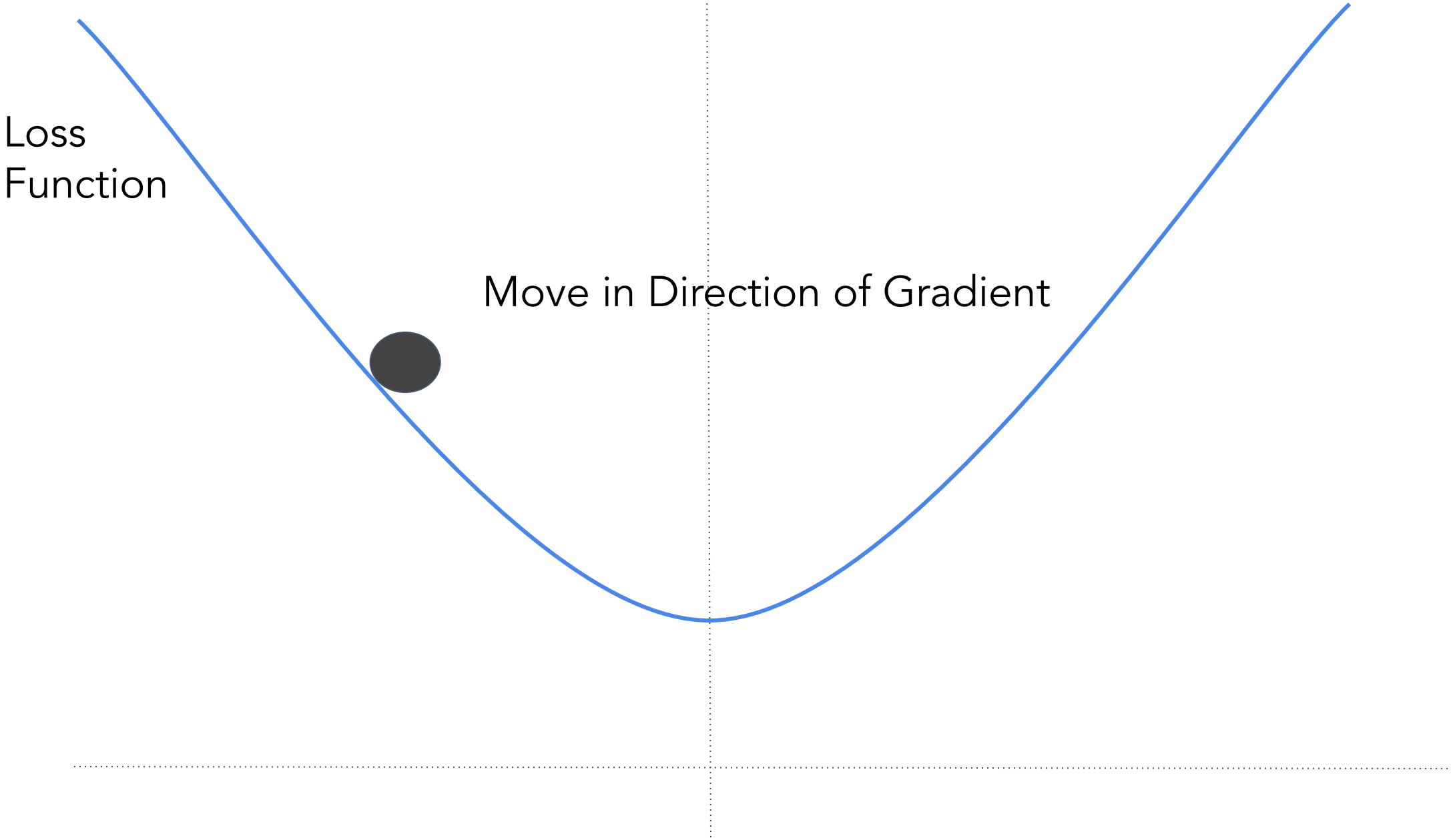
Loss
Function

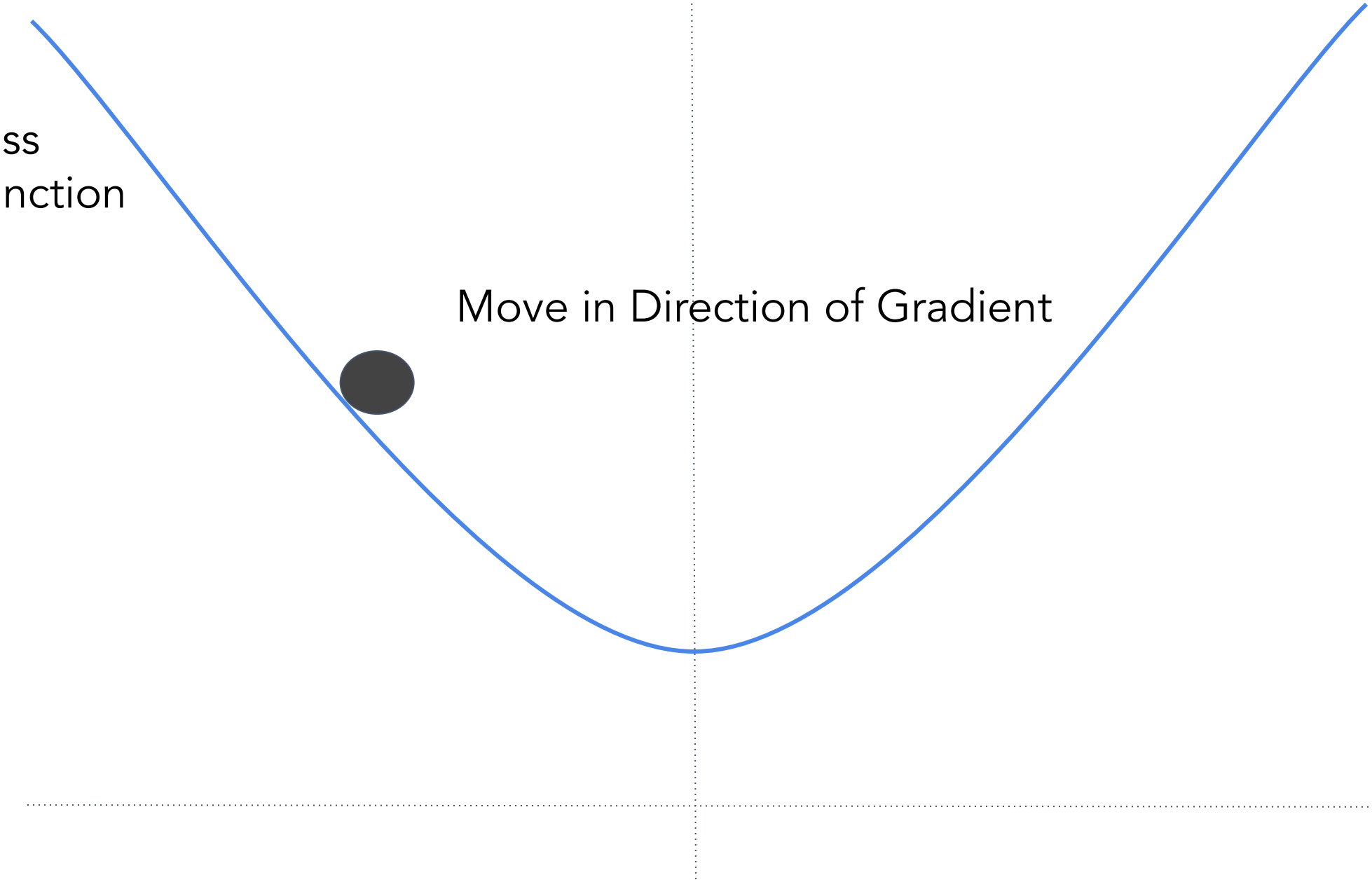



Loss
Function
Move in Direction of Gradient

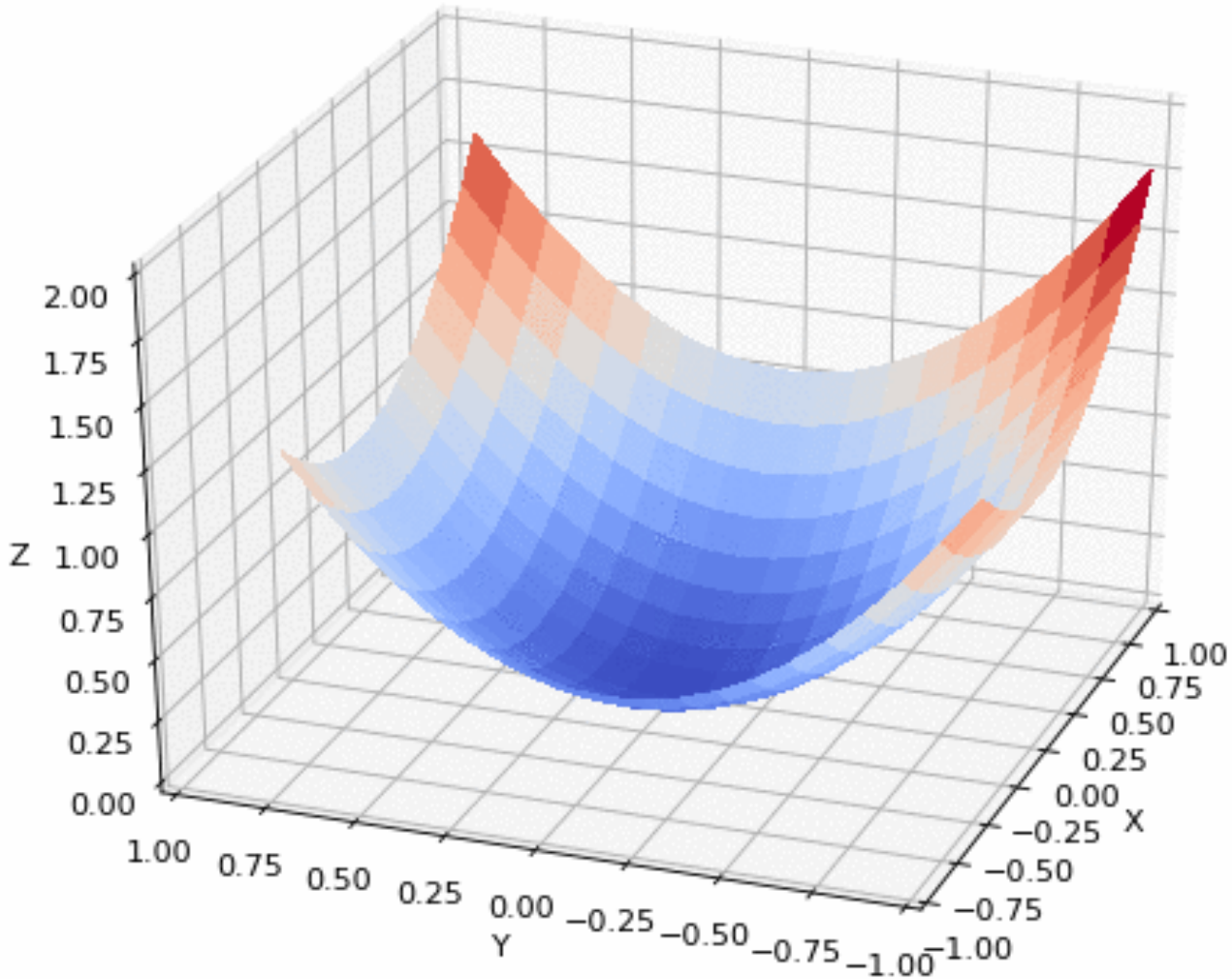

Loss
Function
Move in Direction of Gradient


Loss
Function
Move in Direction of Gradient


Loss
Function
Move in Direction of Gradient


Loss
Function
Move in Direction of Gradient

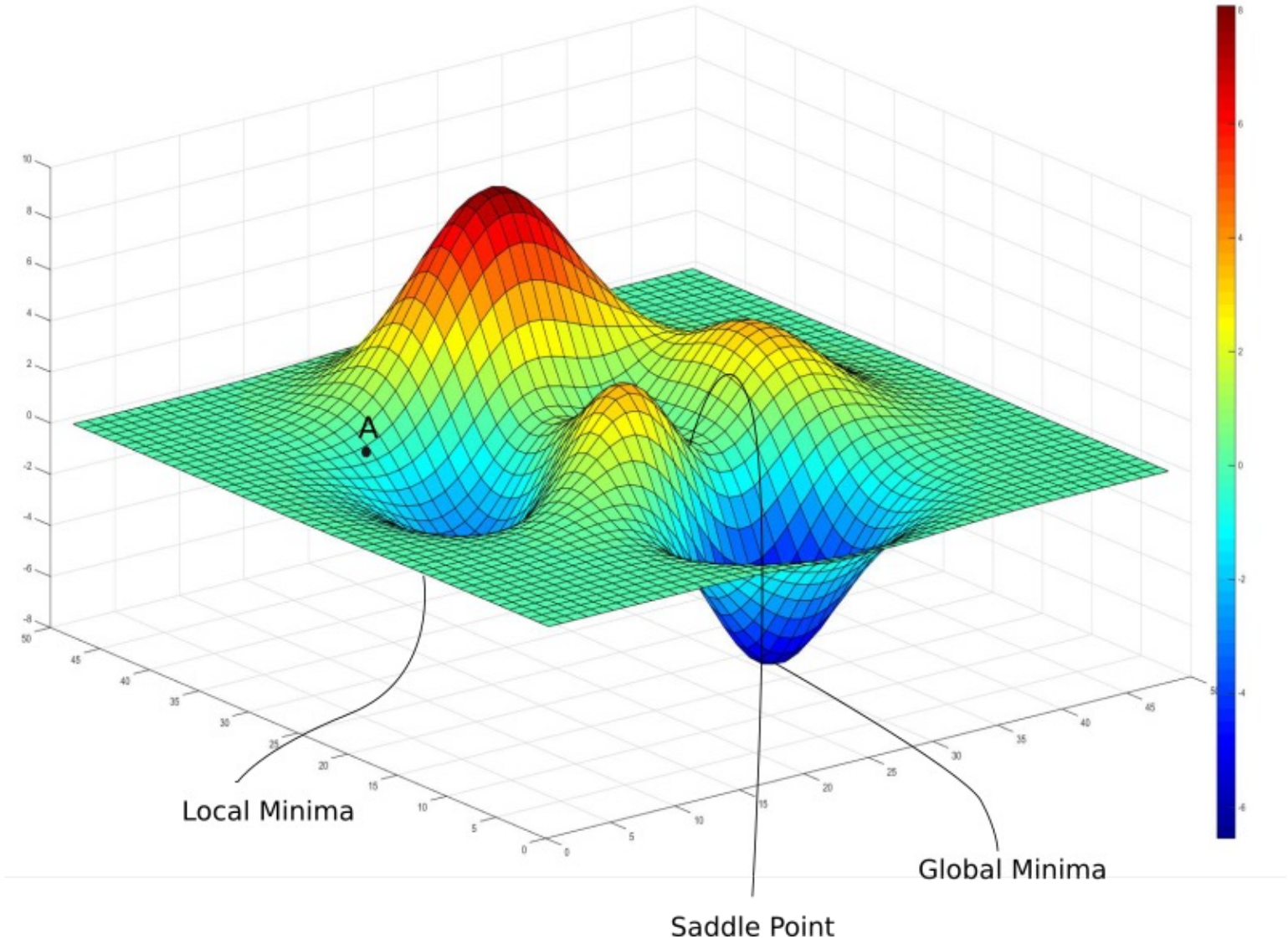
Gradient Descent for Two Parameters



A single minima

Global minima

Gradient Descent for Two Parameters



Thank you!

Prof. Diego Méndez Chaves, Ph.D

Associate Professor - Electronics Engineering Department
Director of the Master Program in Internet of Things
Director of the Master Program in Electronics Engineering
email: diego-mendez@javeriana.edu.co
Website: www.javeriana.edu.co/blogs/diego-mendez/