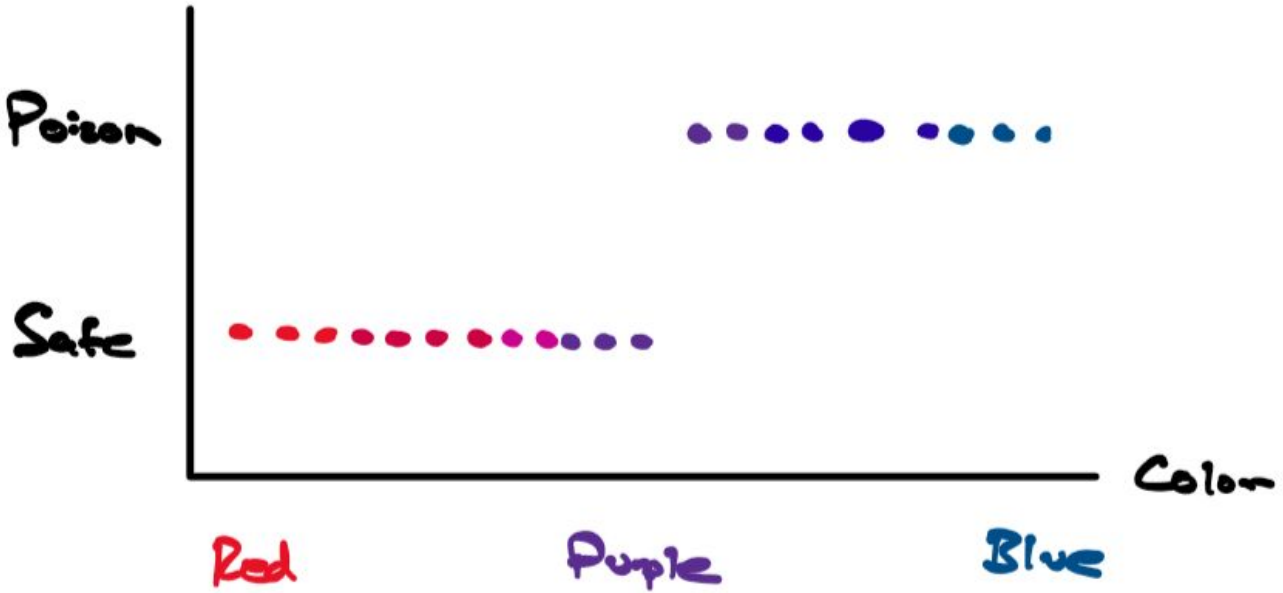
Gradient Descent + PyTorch

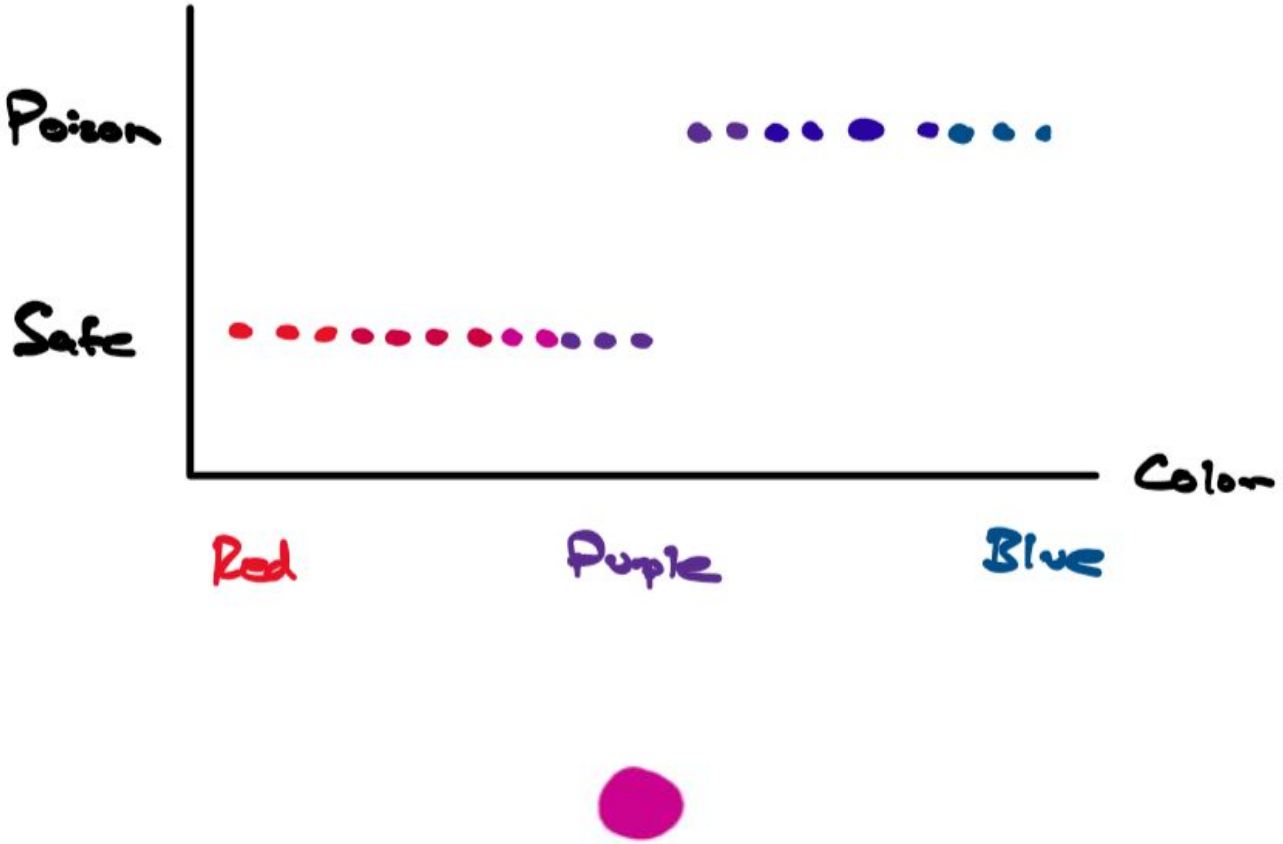
The Jubilee Institute, January 2023

Oh no! We have some poisonous berries!

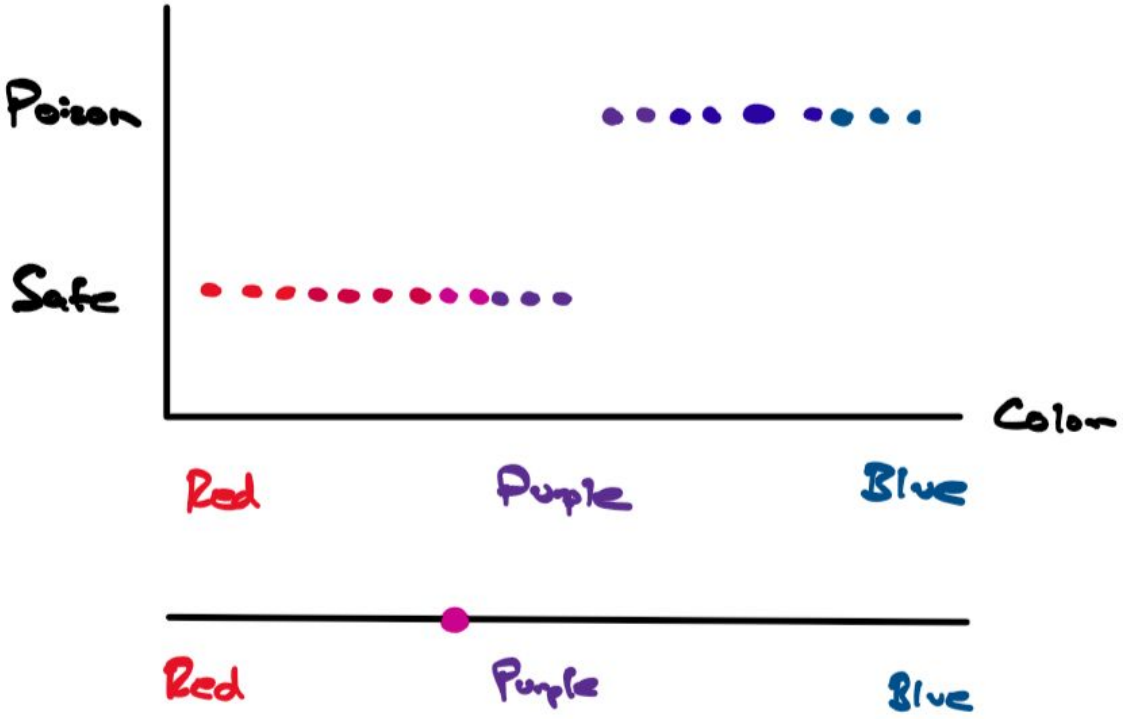
After some sacrifices...



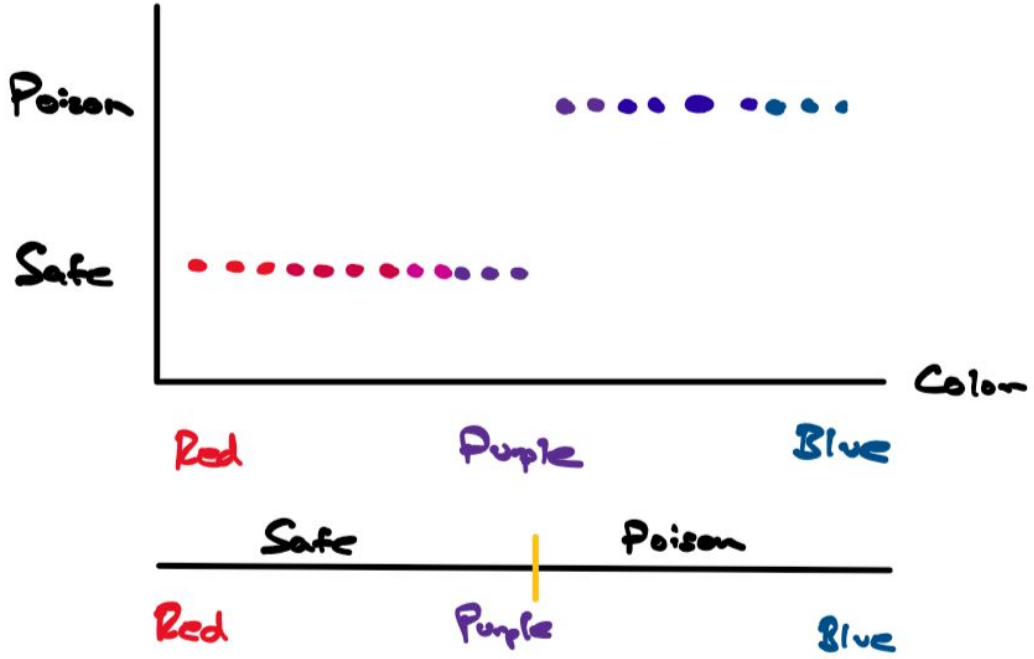
Just looking at color, is this berry poisonous?


Poison
Safe
Color
Red
Purple
Blue

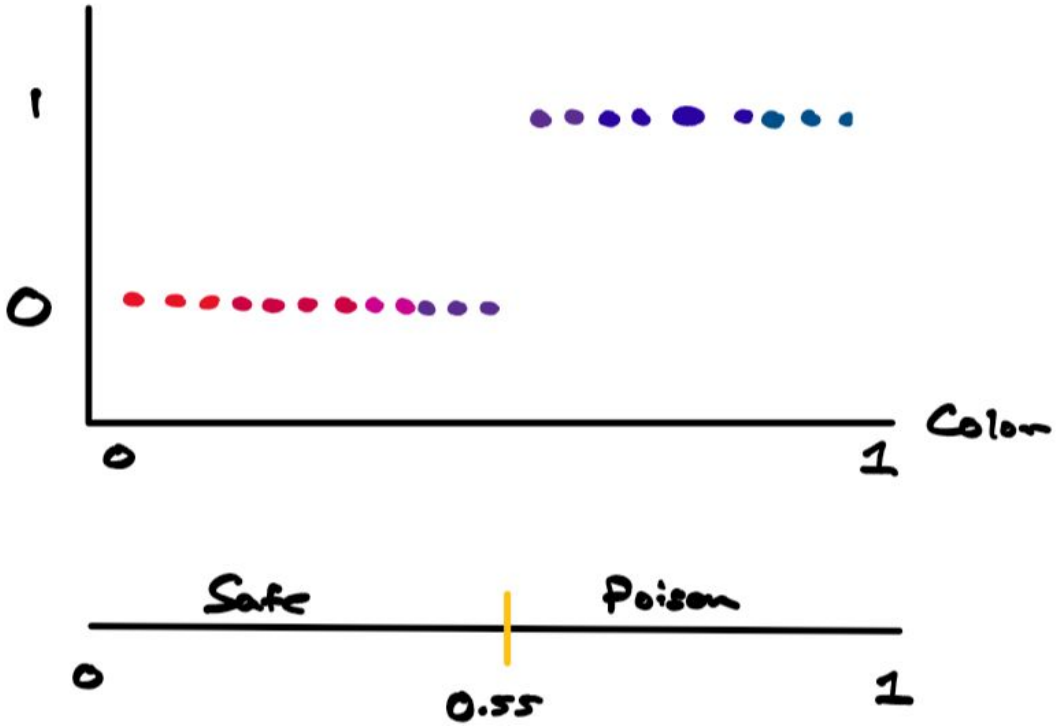
No, it is probably safe...



We can easily find a separator.



Let's convert our graph to numbers.



We are safe now!

- We know that if the color rating x is less than 0.55, the berry is probably safe to eat. If it is greater than 0.55, it is probably going to be poisonous.
- Color is our **parameter**, and x = 0.55 is our separator.
- This was an easy problem. We only had 1 parameter, and it was obvious where the separator should be by just looking at the data.
- What if we had 2 parameters? 10 parameters? 784 parameters?
- We have to find a better way to do this...

What if we started with a random guess?

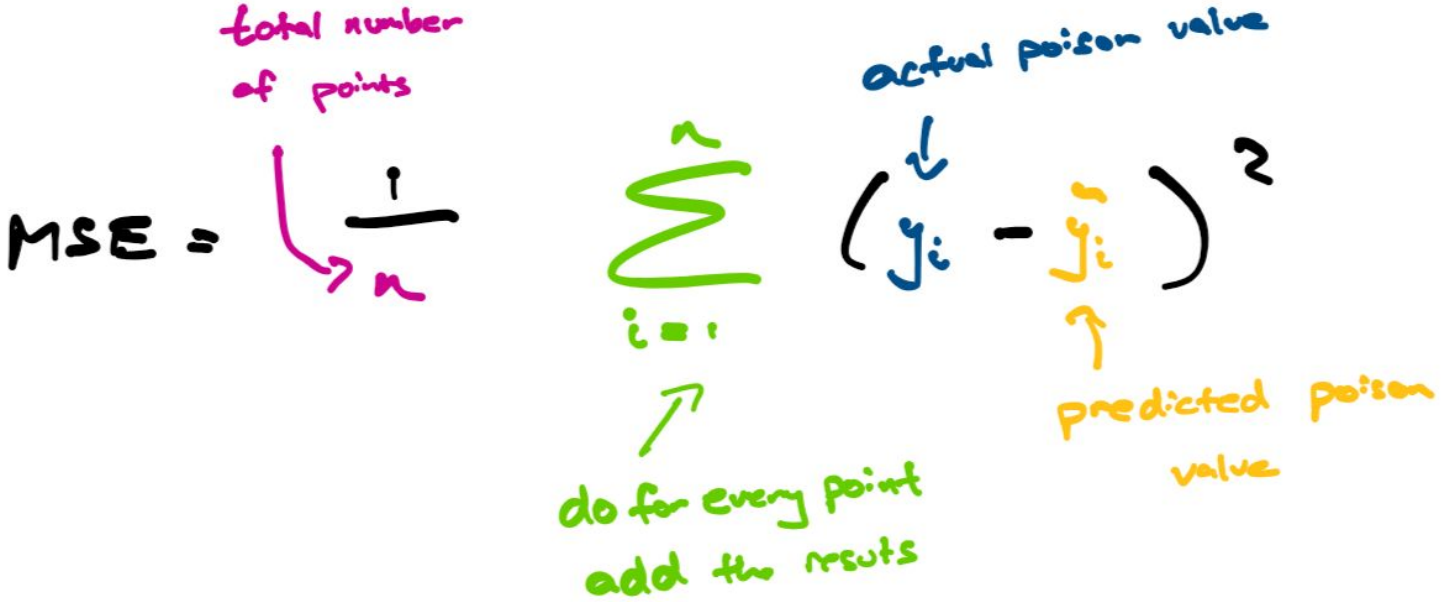
How can we determine how good or bad that guess is?

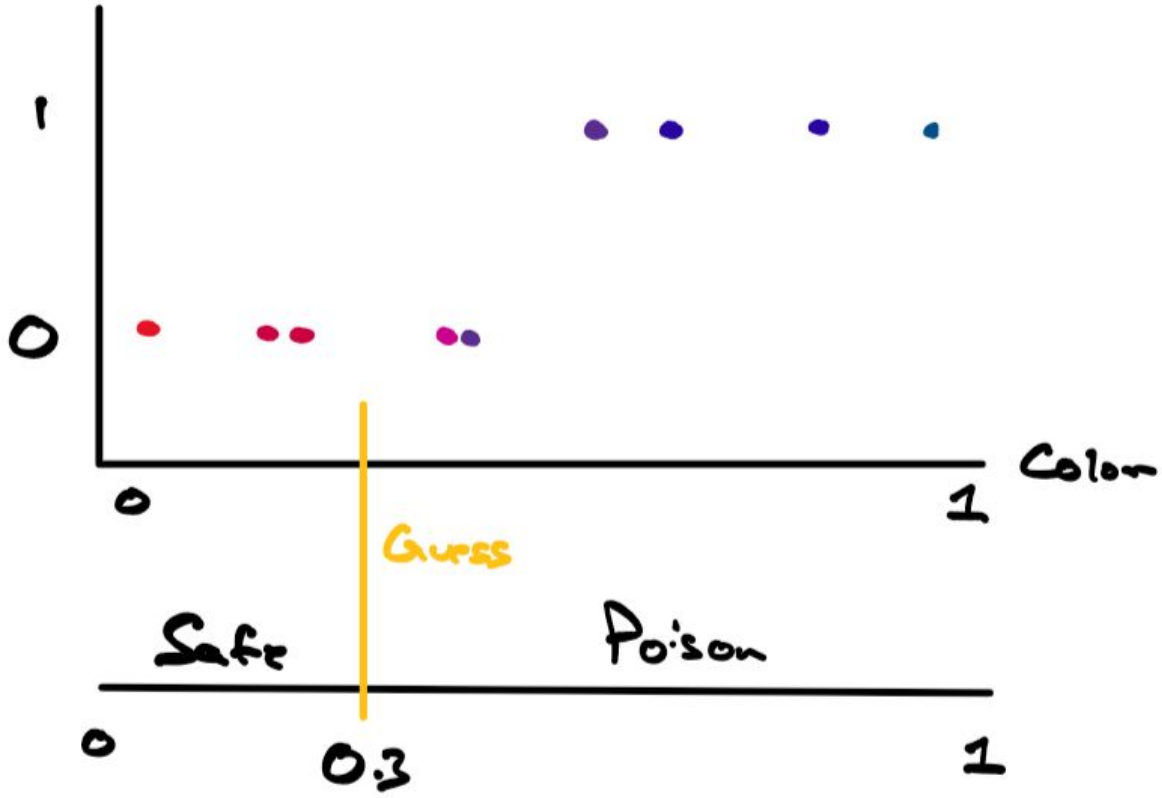
Cost Function C(x)

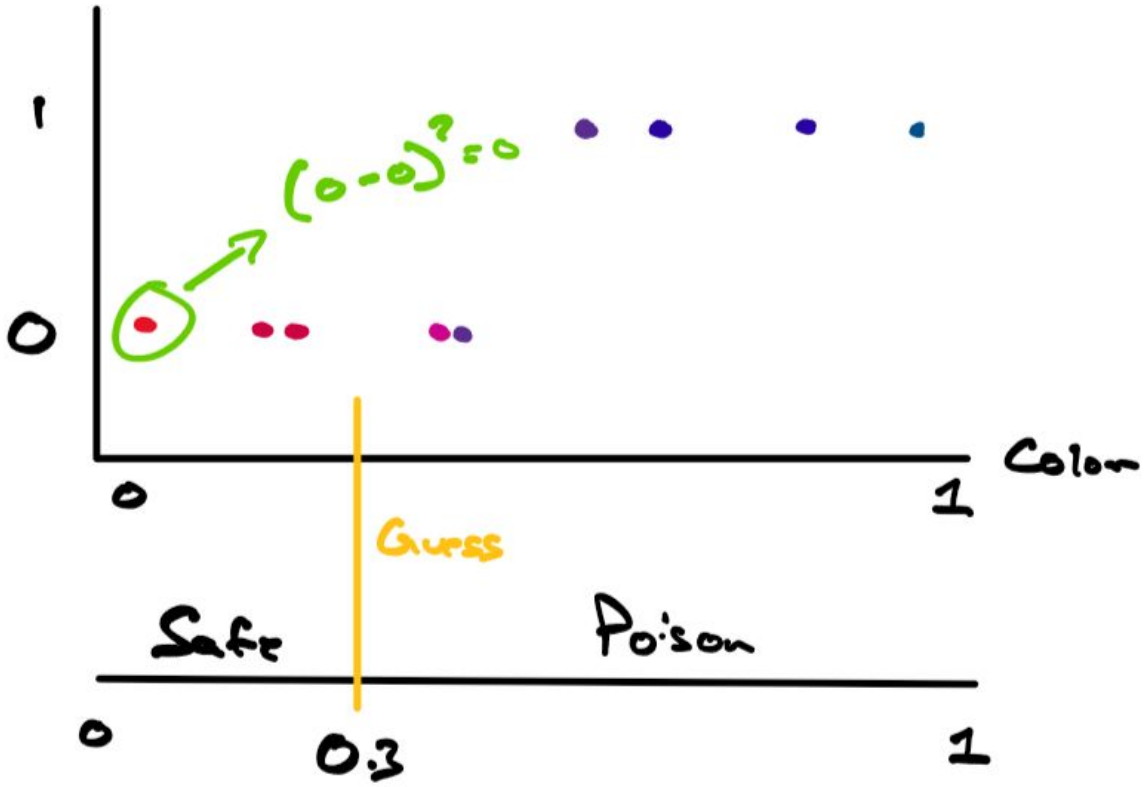
A function C(x) that tells us how bad our random guess is.

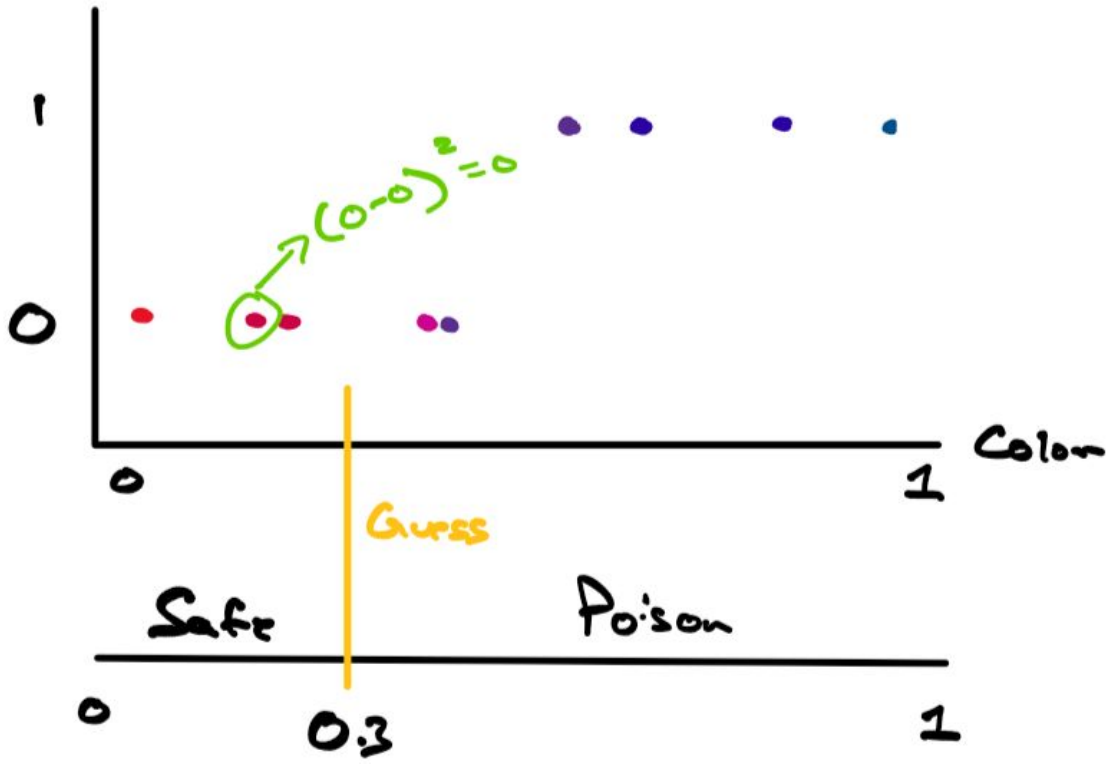
Mean-Squared Error

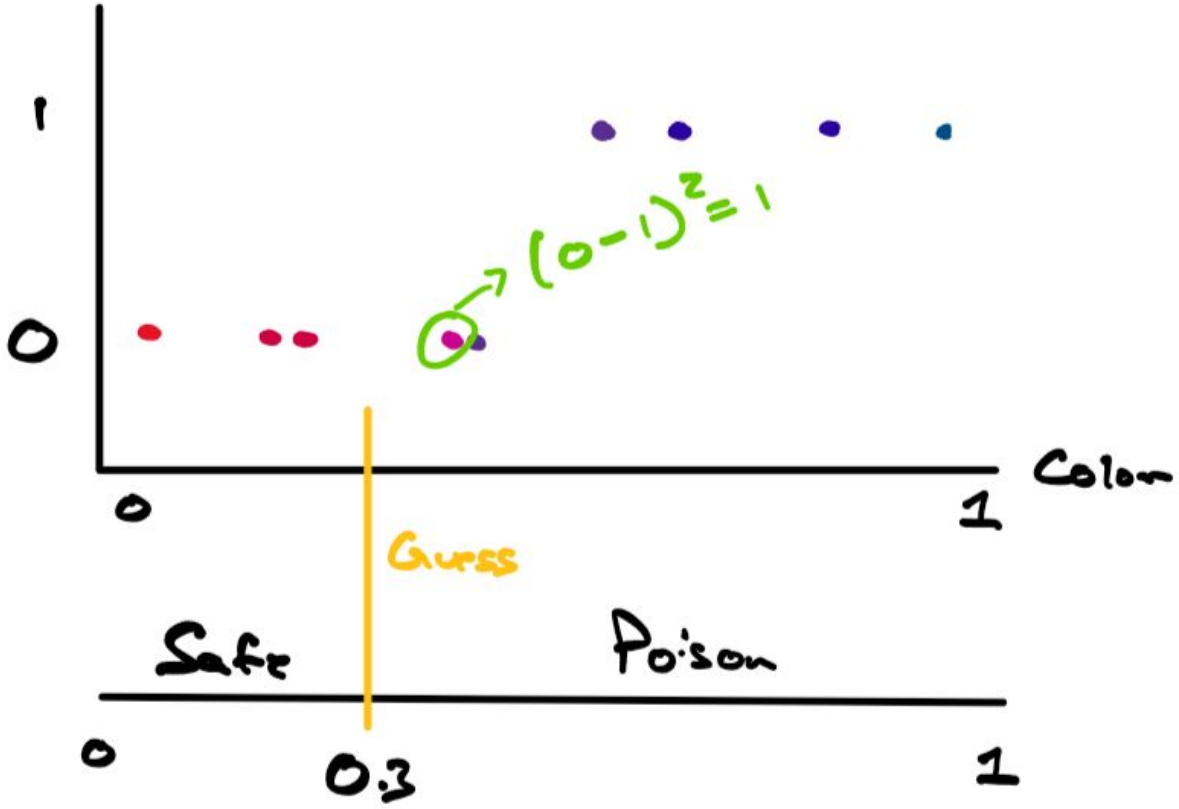
$$\mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2$$


total number of points
actual poison value
MSE =
1/n
$\sum_{i=1}^{n}$
$(y_i - \hat{y}_i)^2$
do for every point add the results
predicted poison value


1
0
Color
0
1
Guess
Safe
Poison
0
0.3
1


1
0
(0-0)²=0
0
1
Color
Guess
Safe
Poison
0
0.3
1

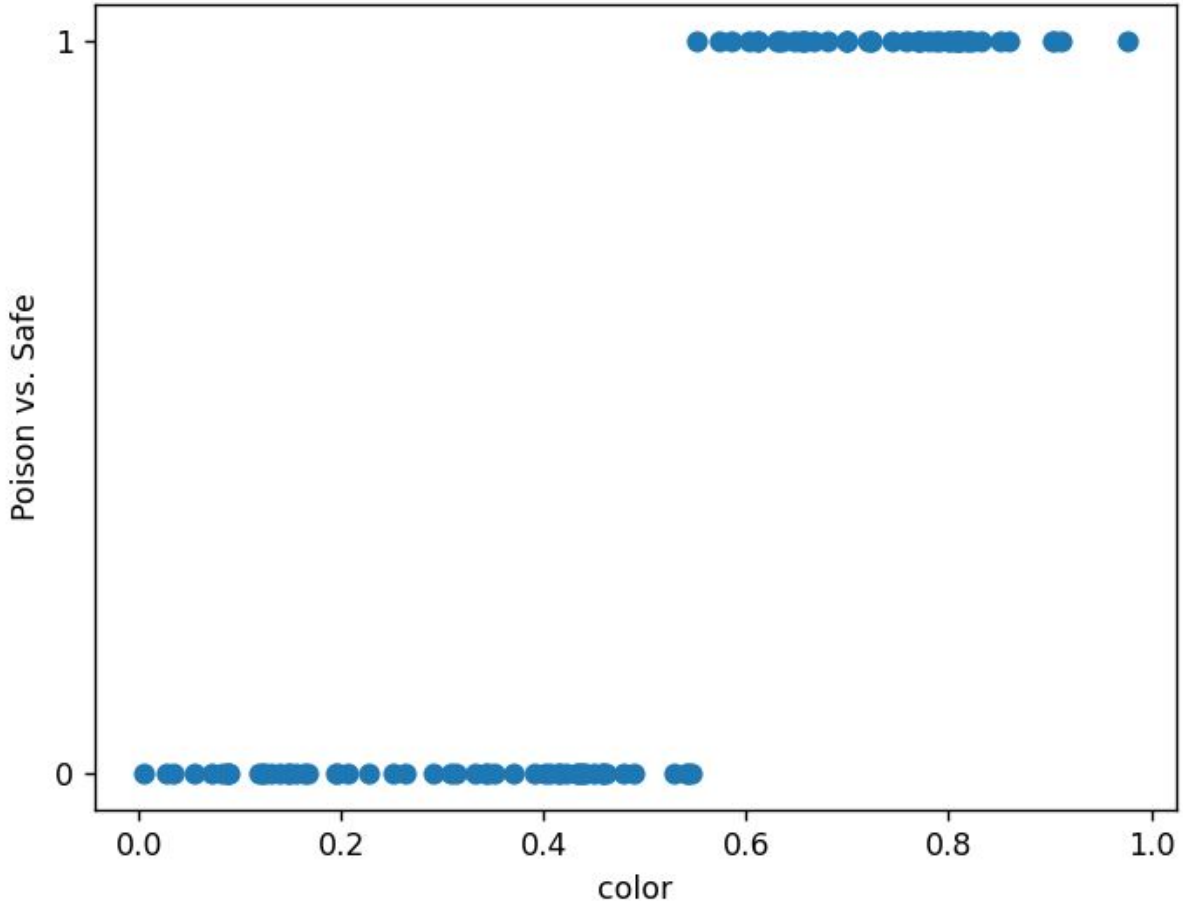

1
0
$(0-0)^2=0$
0
1
Color
Guess
Safe
Poison
0
0.3
1

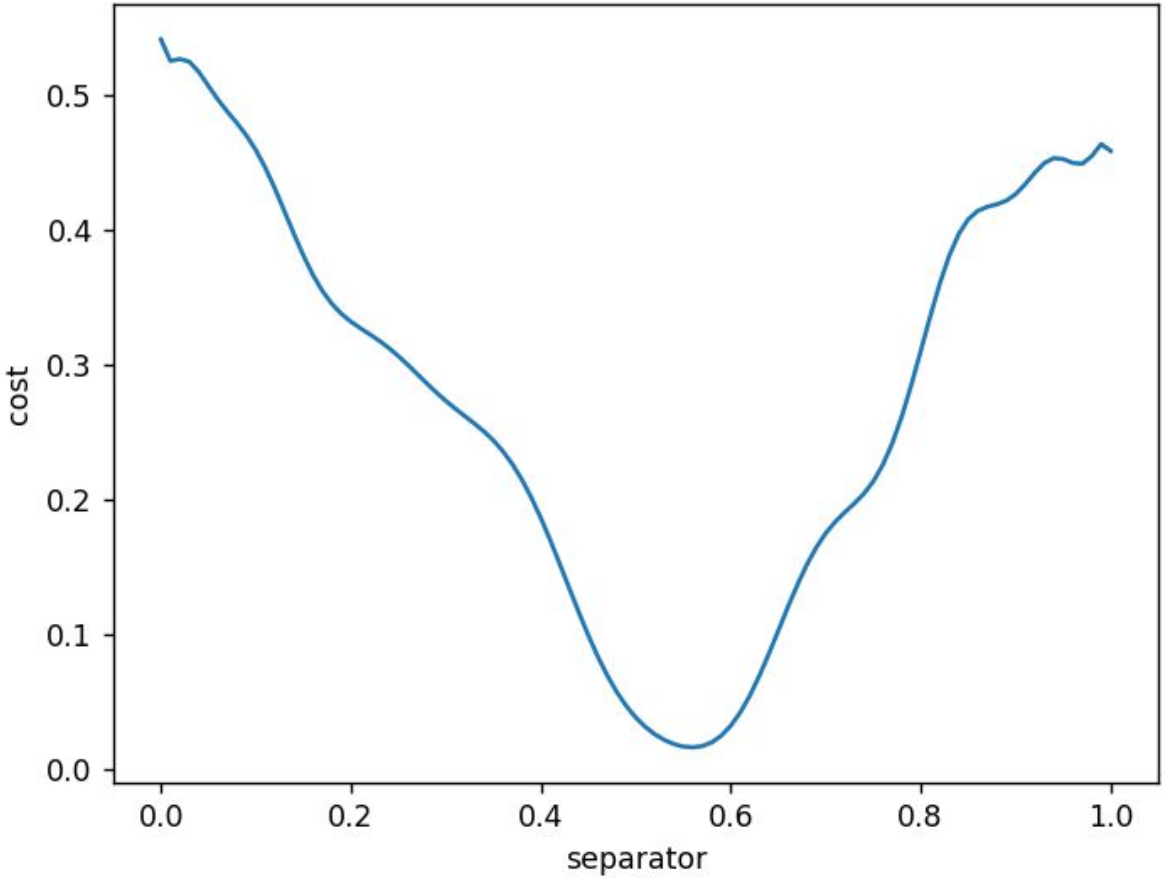

1
0
$(0-1)^2=1$
Color
0
1
Guess
Safe
Poison
0
0.3
1

$$MSE = \frac{1}{9} \cdot 2 = 0.22$$

Cost of Separator $x = 0.3$

$$C(x) = 0.22$$


1
0
Poison vs. Safe
0.0
0.2
0.4
0.6
0.8
1.0
color


0.5
0.4
0.3
0.2
0.1
0.0
cost
0.0
0.2
0.4
0.6
0.8
1.0
separator

How do we find the separator with the minimum cost?

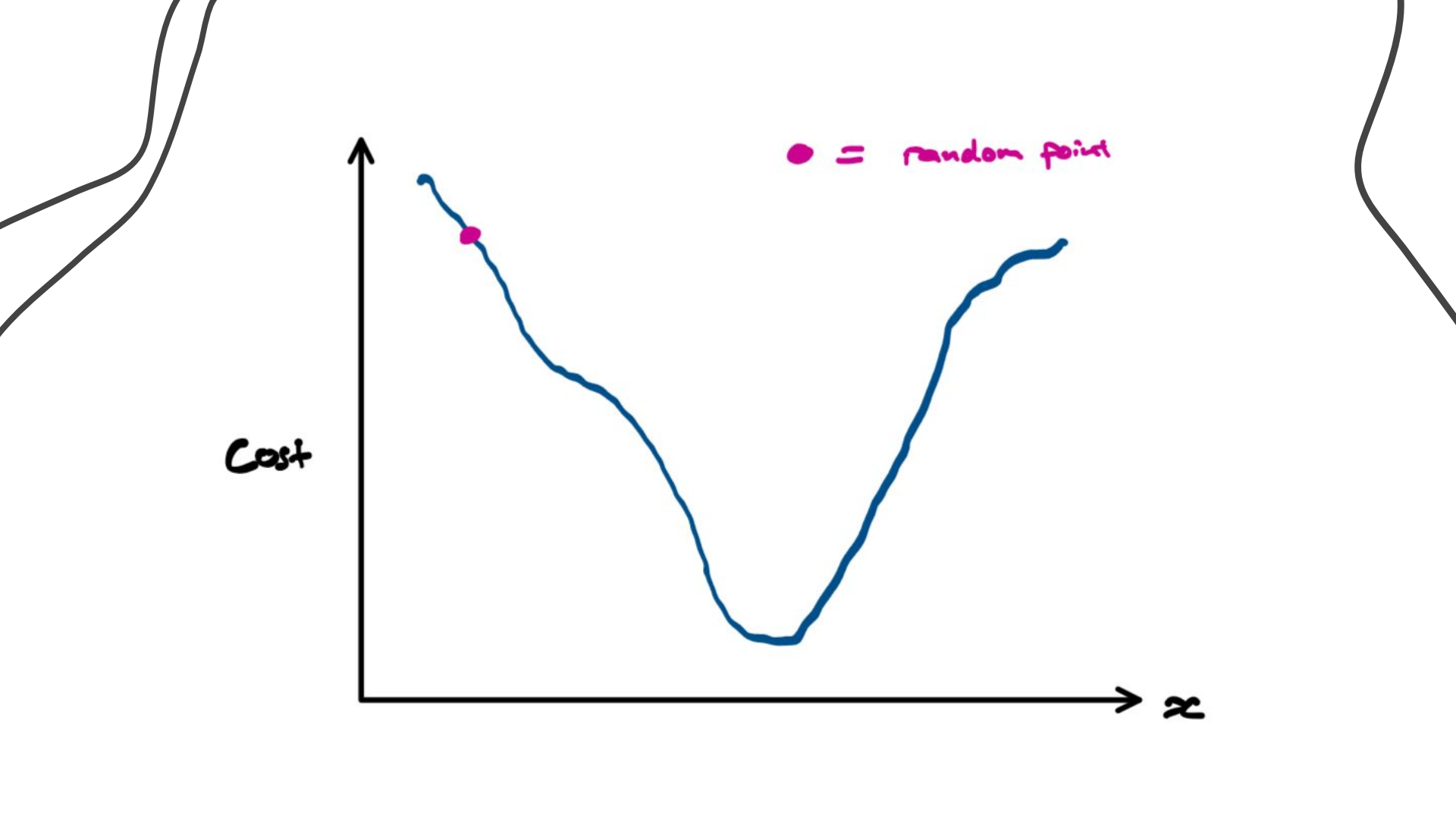
Gradient Descent

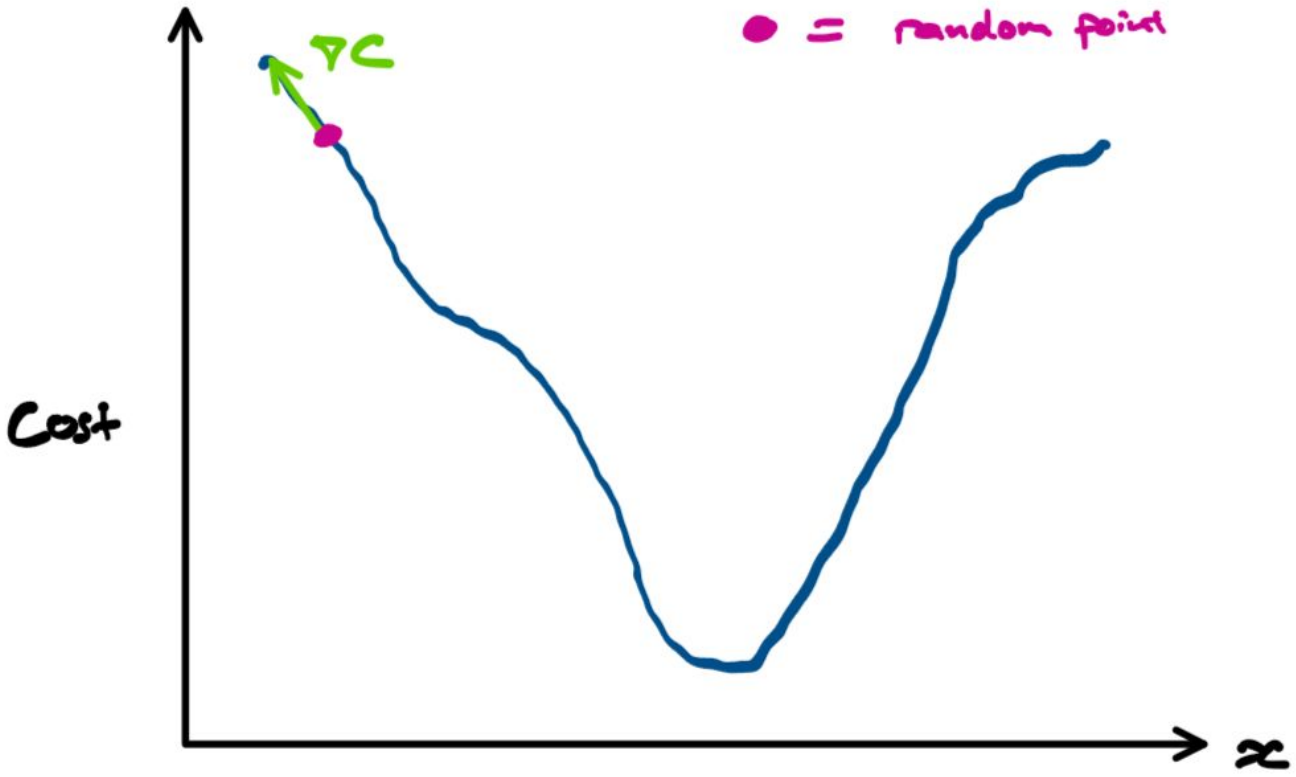
using the **gradient** (direction of fastest increase) of the cost function to find a **minimum**

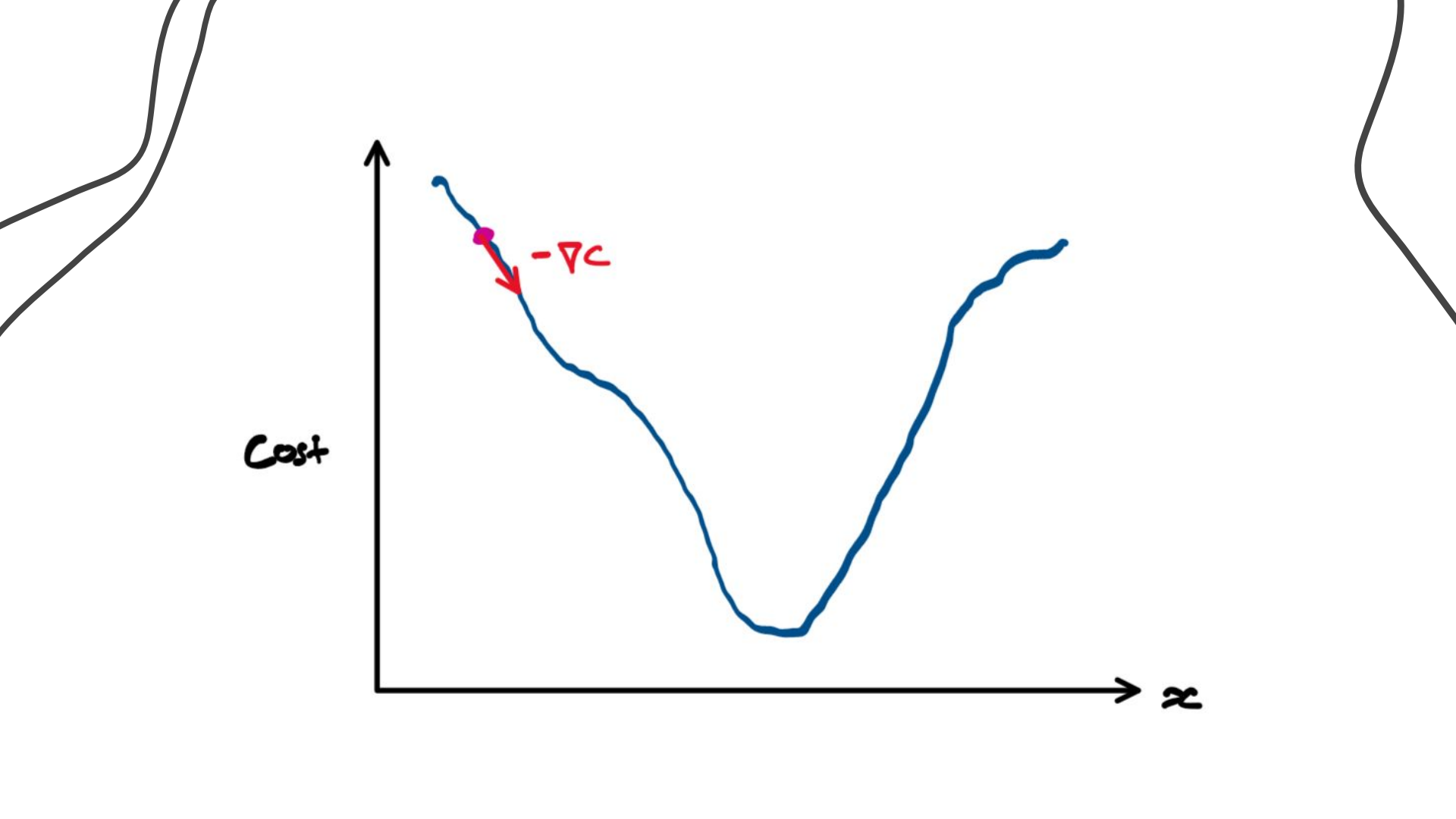
$$\frac{\partial C}{\partial x} \qquad \nabla C$$

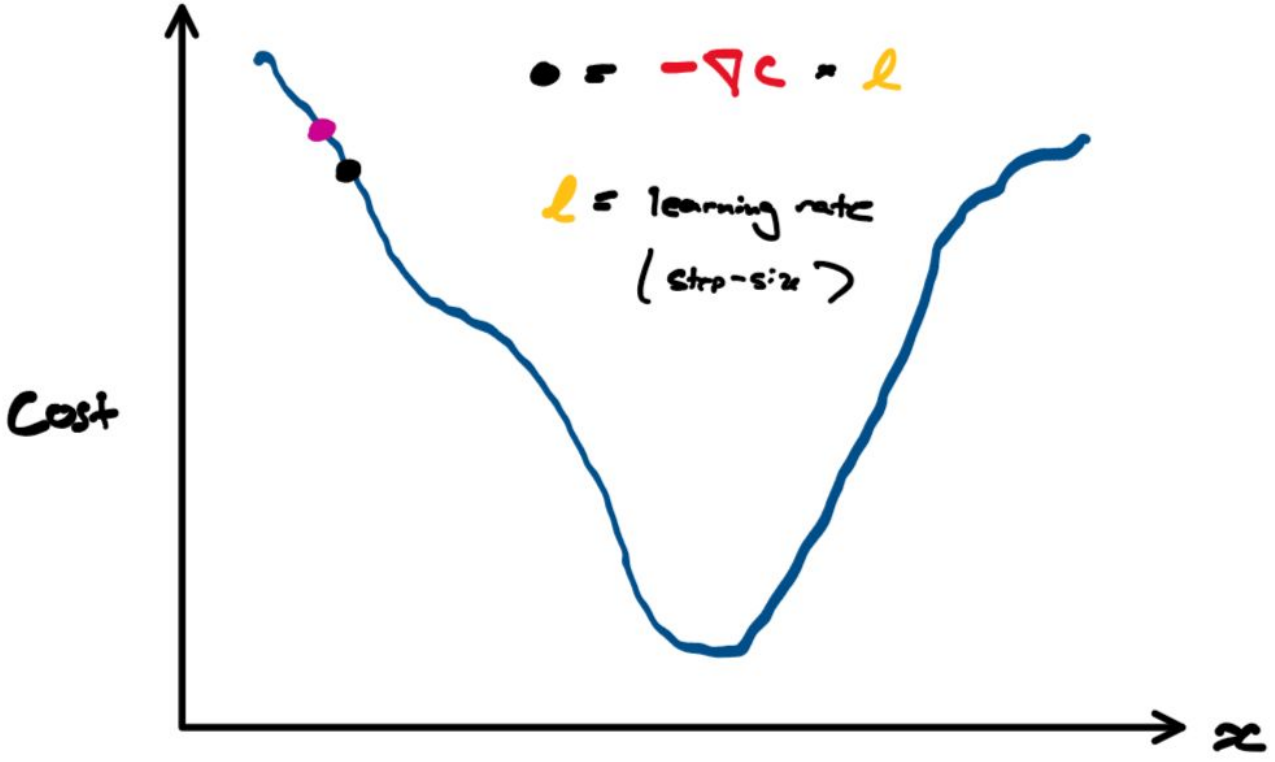
Note:

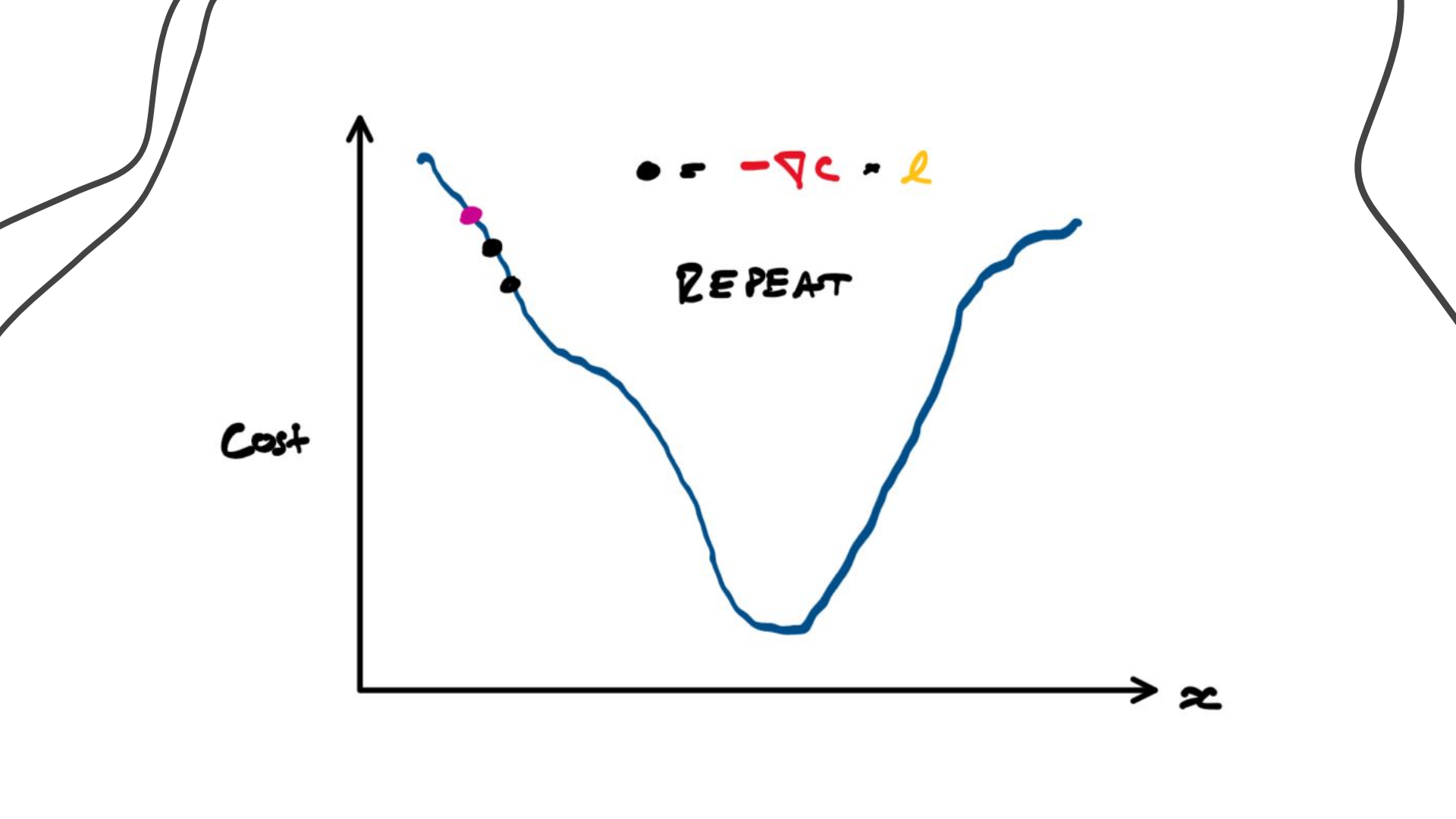
To understand how gradients are calculated, and how they relate to one another, knowledge of calculus (specifically, derivatives) is required.

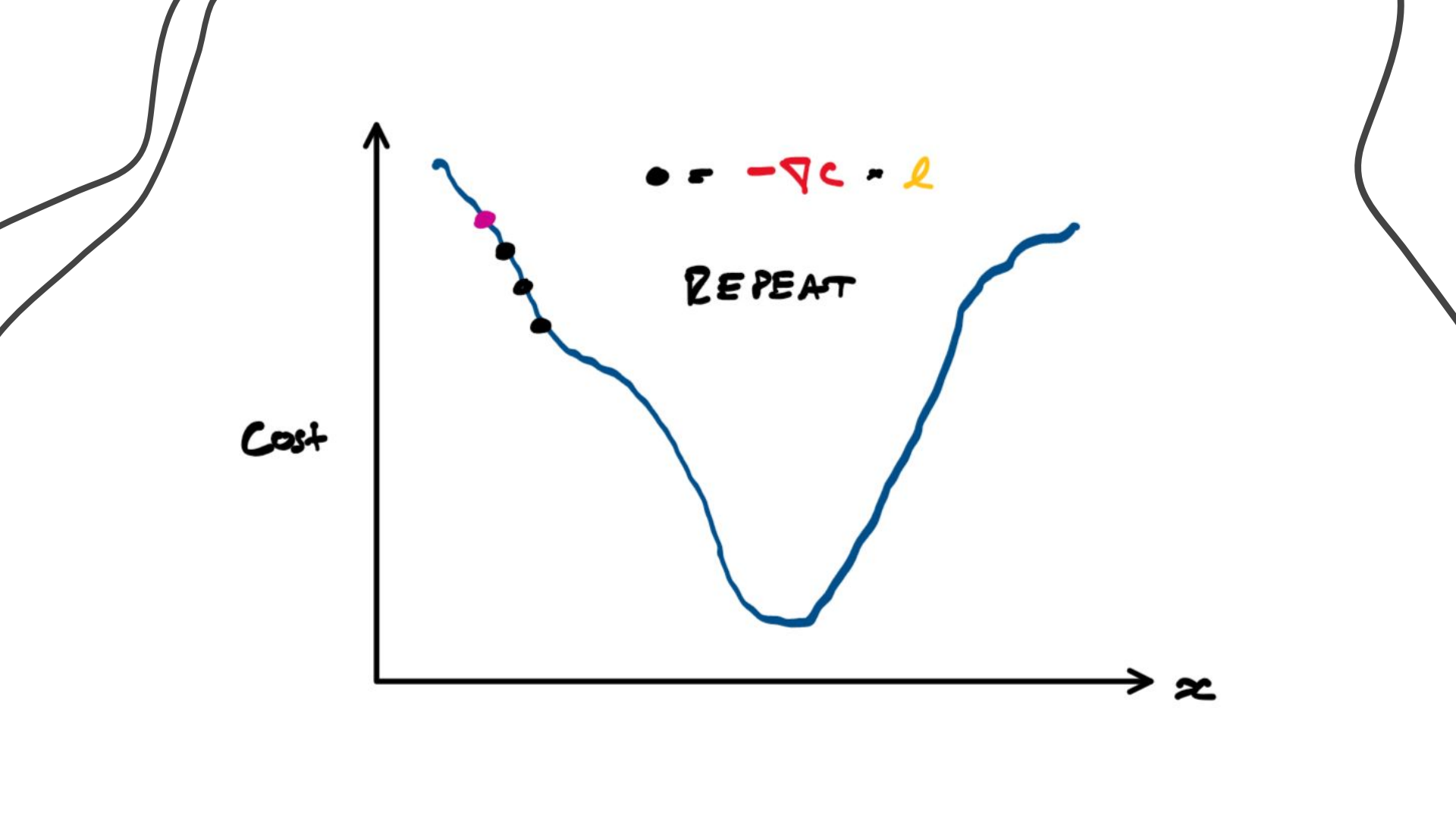

● = random point
Cost
x

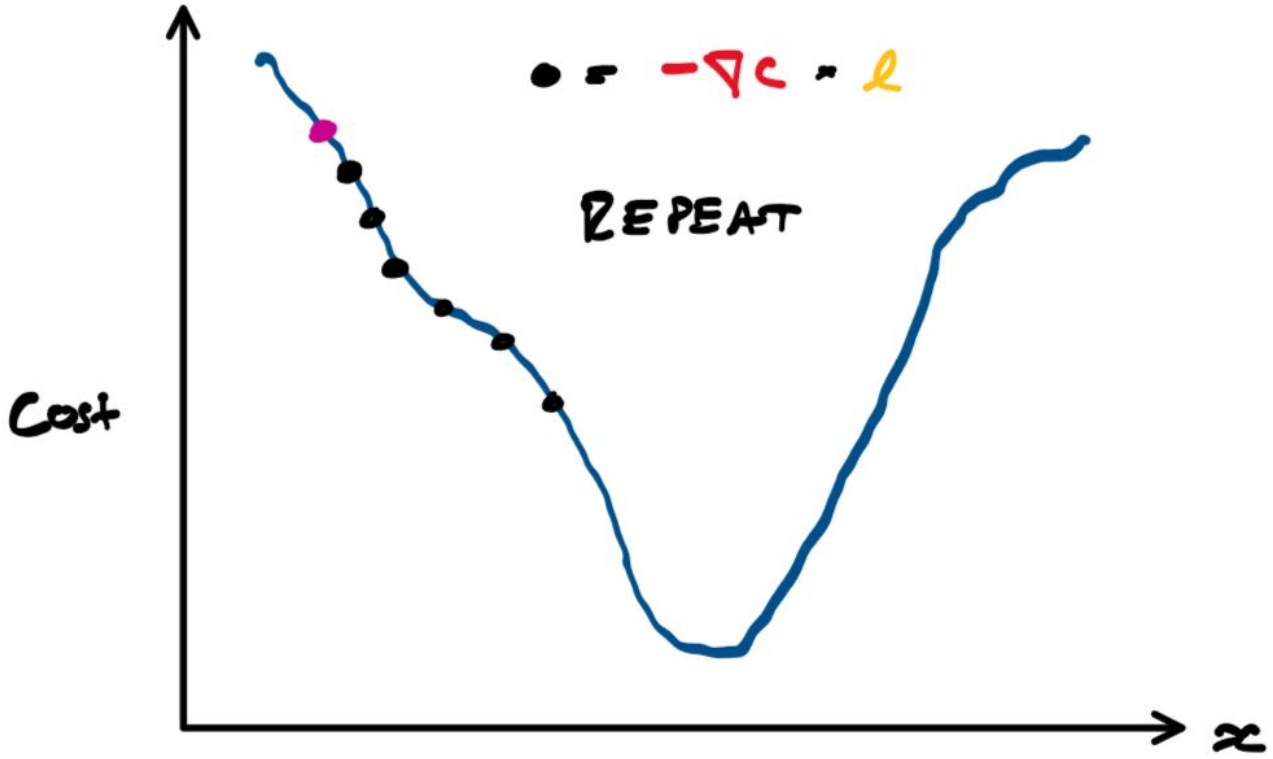

∇C
● = random point
Cost
x

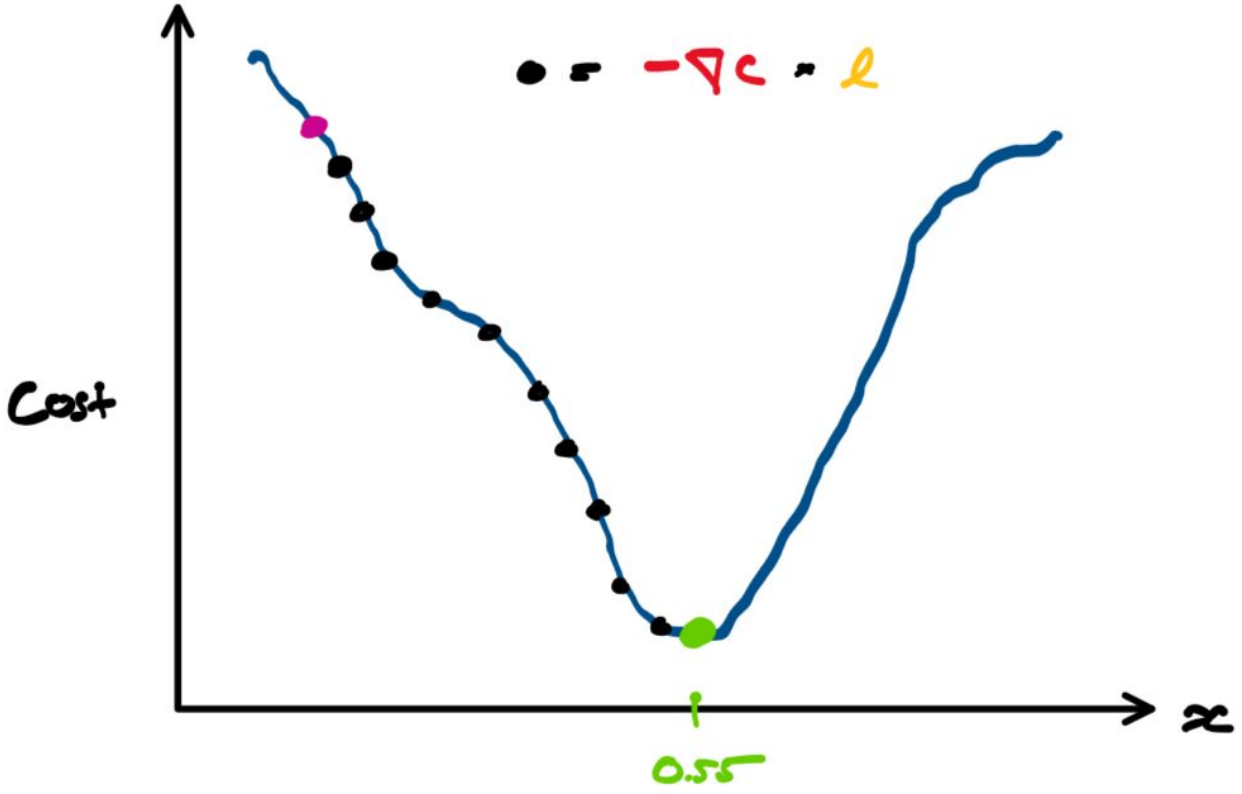

Cost
$-\nabla C$
x

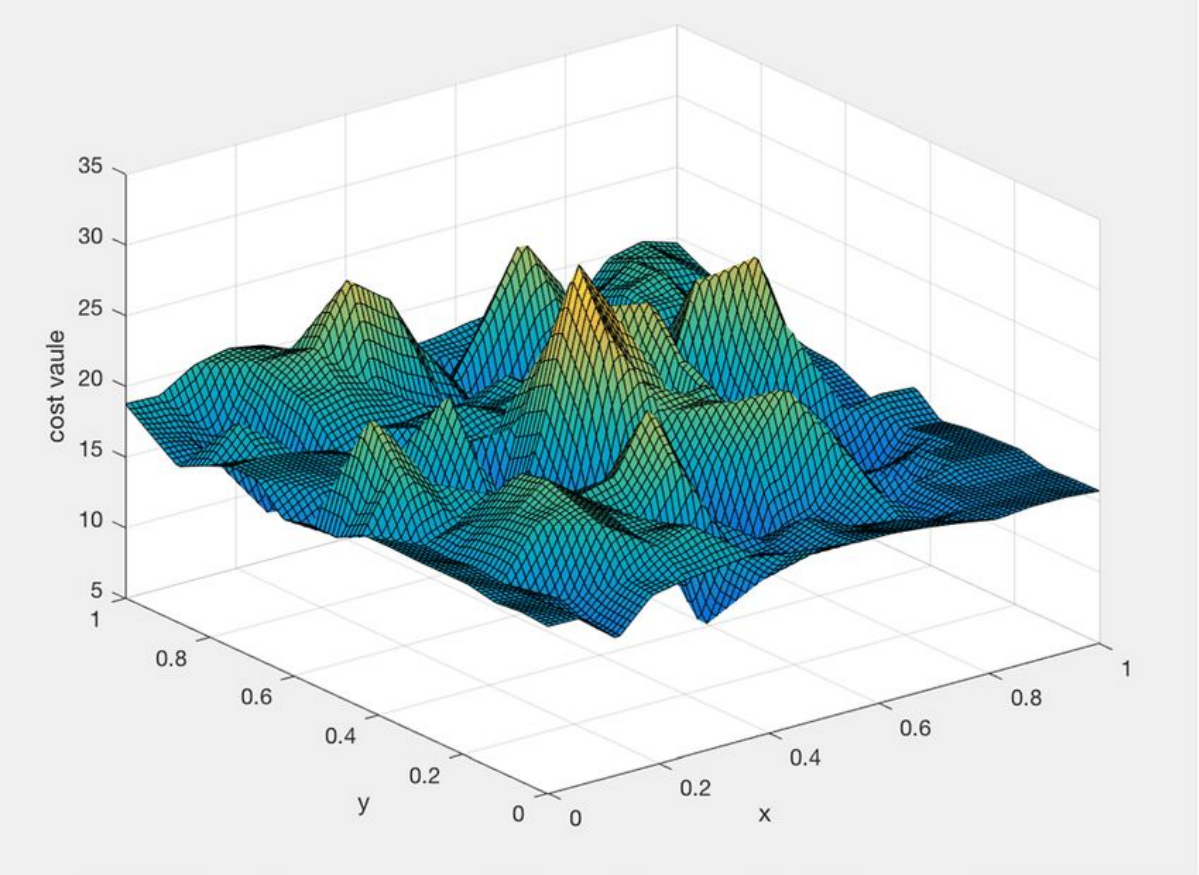

● = −∇C × ℓ
ℓ = learning rate
(step-size)
Cost
x


Cost
x
REPEAT


Cost
x
REPEAT


Cost
x
REPEAT


Cost
x
0.55

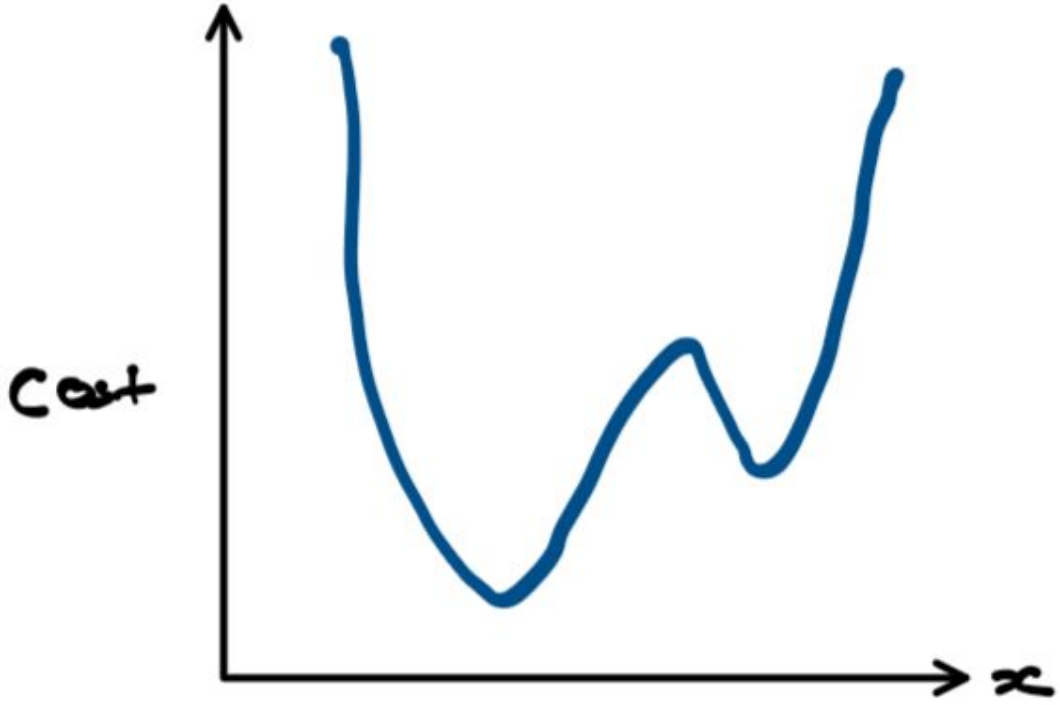


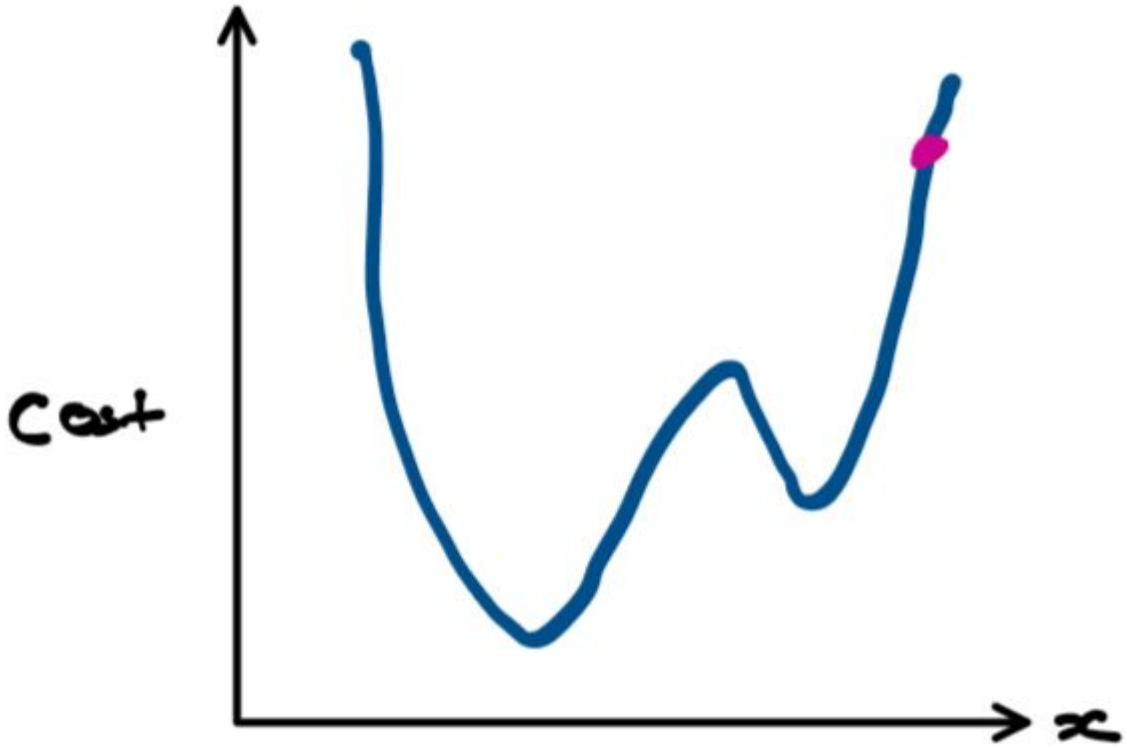
two parameters

?????

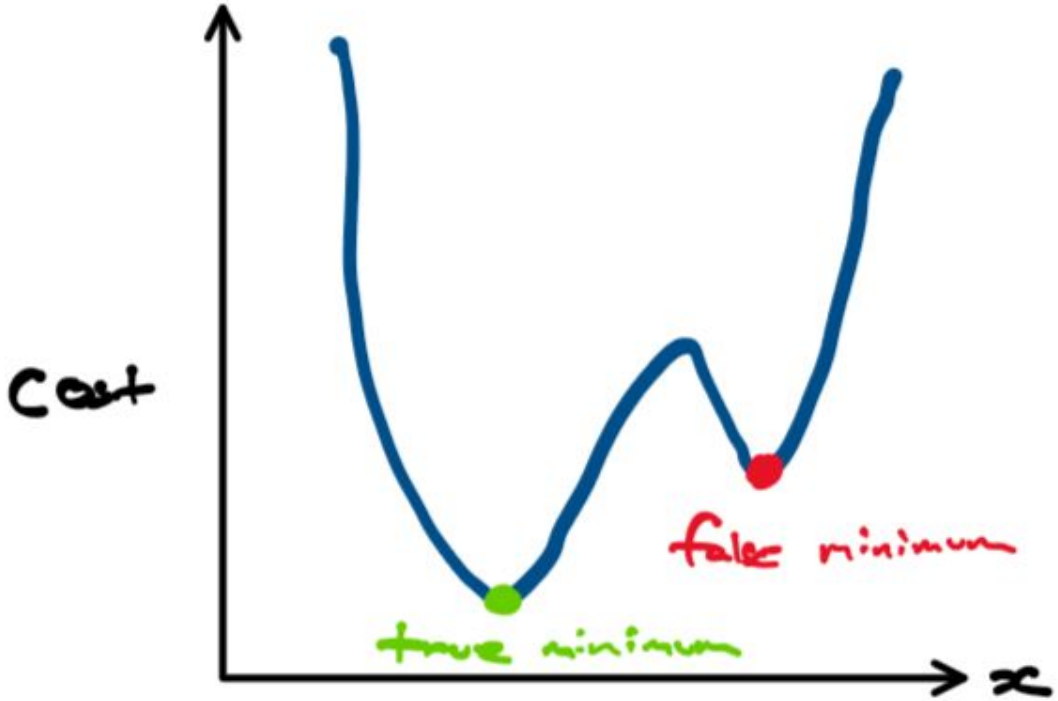
too many
dimensions...

Problem with Gradient Descent...




Cost
x

false minimums are common...



Questions?