The learning process in machine learning

VIP Machine Learning Course

Learning parameters (in supervised learning)

- The main component of the learning process in machine learning is iteration (which essentially involves trial and error).
- Each model or algorithm in machine learning has a number of optional or mandatory parameters that are used during the model learning process.
- The algorithm manipulates these parameters iteratively to achieve the best output (highest accuracy) and minimize errors.

More precisely

- During learning, a machine learning model aims to satisfy the conditions of an objective function or loss function, which measures the discrepancy between predicted and actual values.
- Simultaneously, it strives to minimize the value of the loss or error computed by the loss function.
- Moreover, each input parameter may be multiplied by a weight (W) and added with a bias (B). The machine learning model tries to learn these weights and biases.

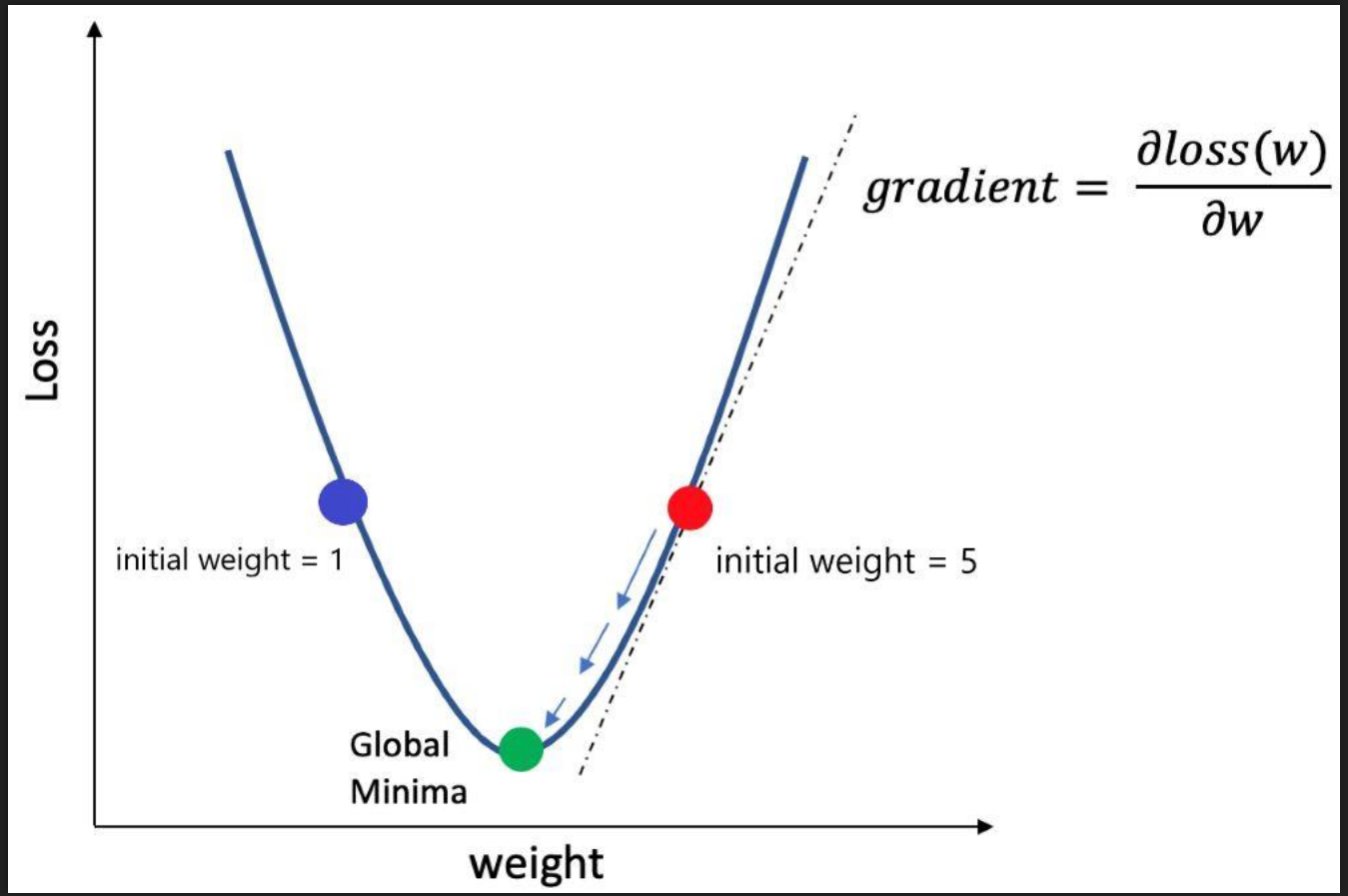
Gradient Descent

- One of the algorithms widely used in machine learning models is Gradient Descent. This algorithm is utilized to find the minimum value of a loss function.
- Gradient Descent is an optimization algorithm (specifically, an iterative optimization algorithm). Its time complexity is of the order $O(kn^2)$, where k is the number of features (or dimensions) and n is the total number of data points.
- To be more precise, Gradient Descent attempts to find a local minimum on a differentiable function by iteratively adjusting parameters based on the negative gradient of the function.

The working method of Gradient Descent

new weight = initial weight - d loss/dw

- If the new value of W is less than the minimum value of W, meaning we are on the left side of the minimum W, the slope or derivative with respect to W is negative. This indicates that the new weight value will increase, and consequently, we are moving towards the minimum W on the left side.
- When the new value of W is greater than the minimum W, and we are on the right side of the minimum W, the derivative of the above expression is positive. This means that the new weight value will decrease, and therefore, we are moving towards the minimum W on the right side.



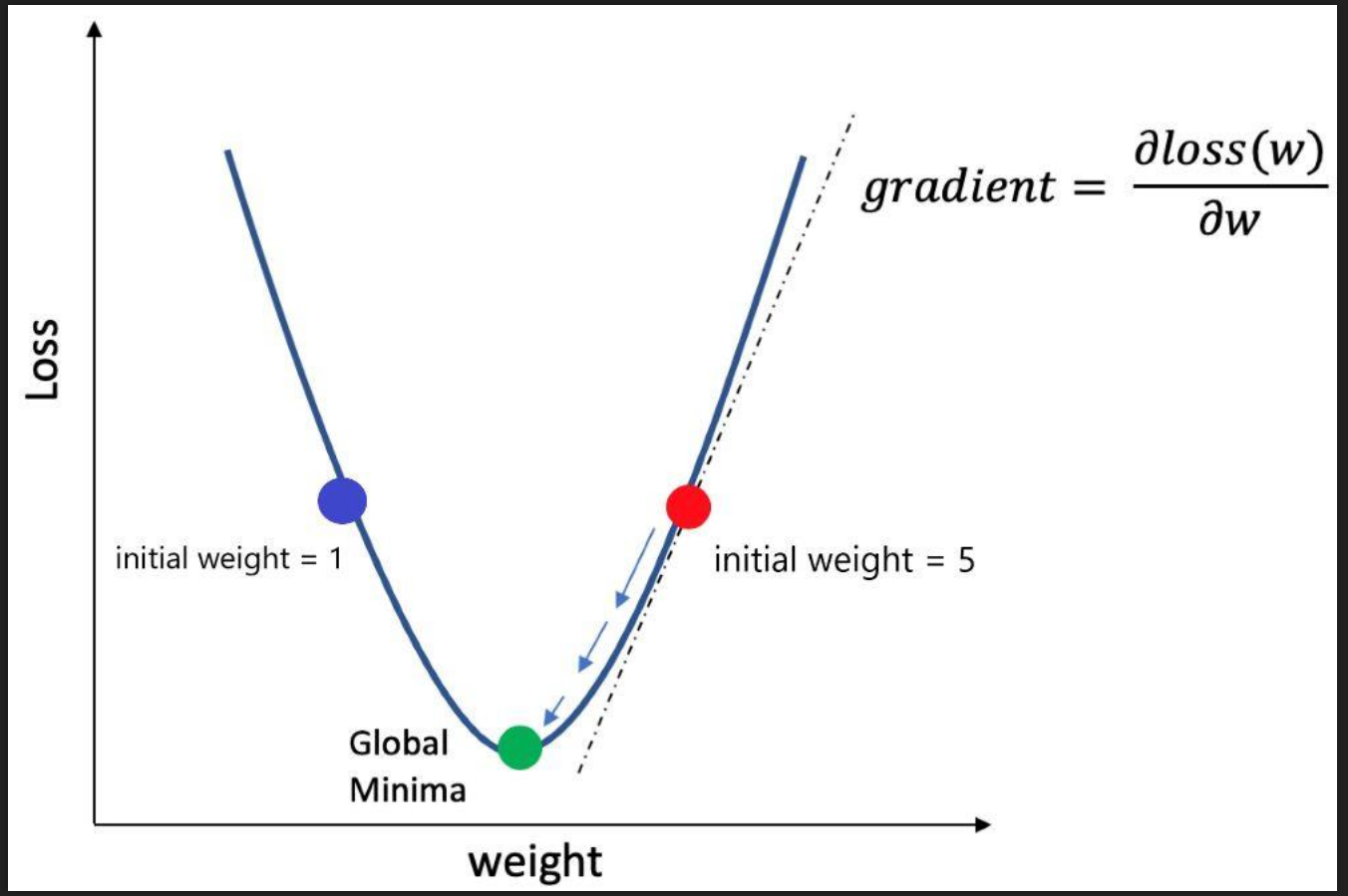
The working method of Gradient Descent

new weight = initial weight - d loss/dw

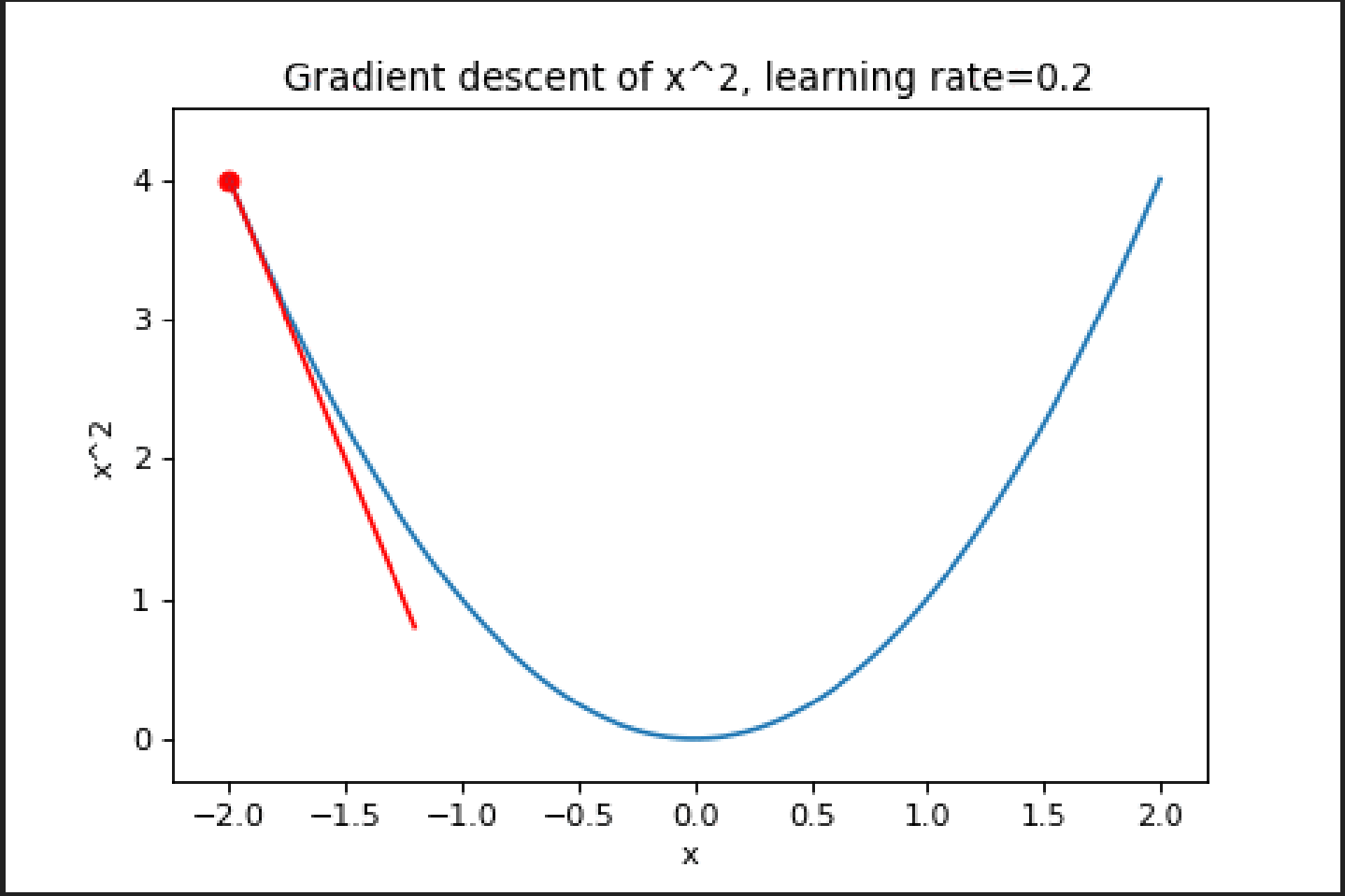
- In fact, to improve the speed of this process, it's better to use a coefficient called the learning rate.

w' = w - alpha * d loss/dw

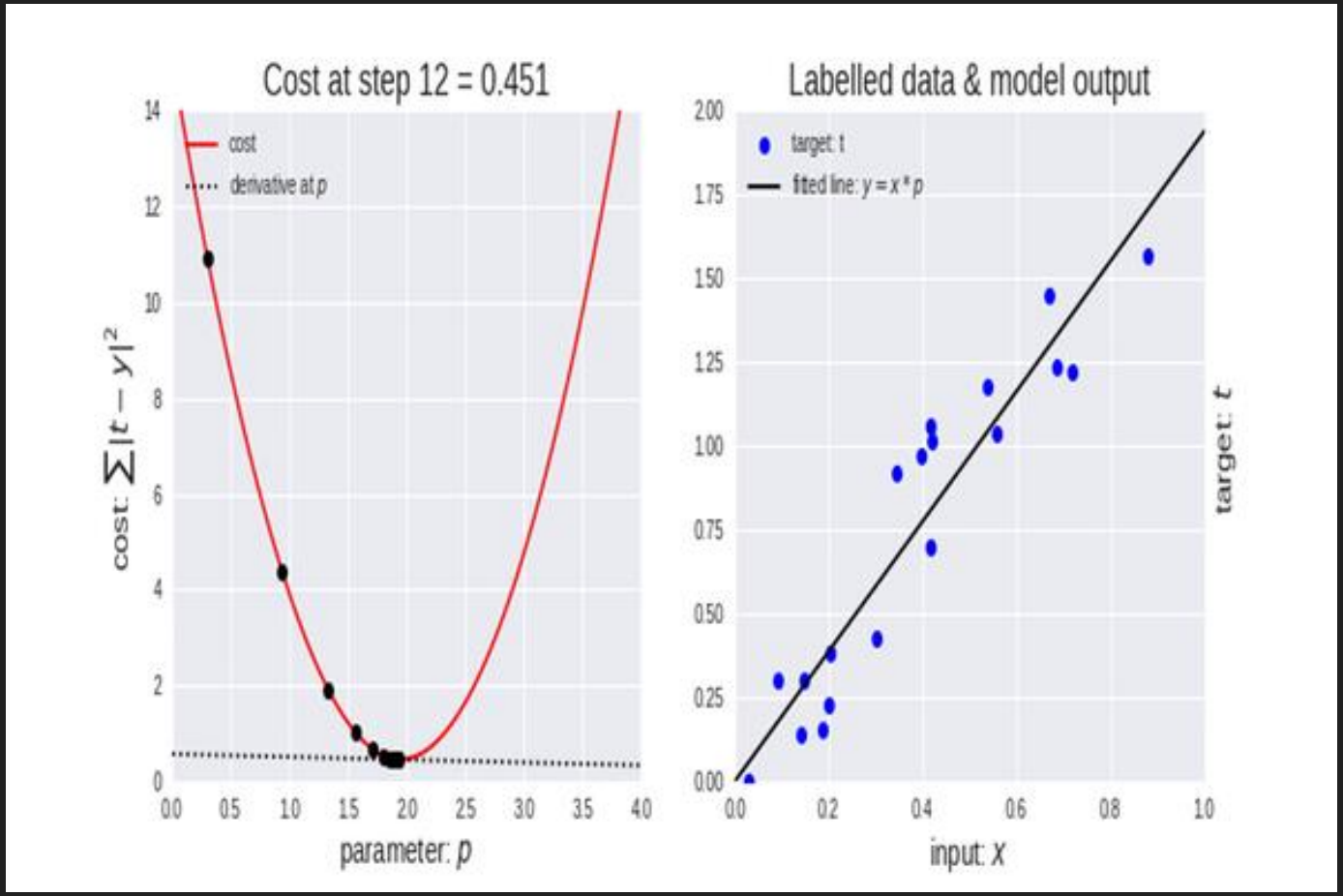
Note: alpha is the learning rate.



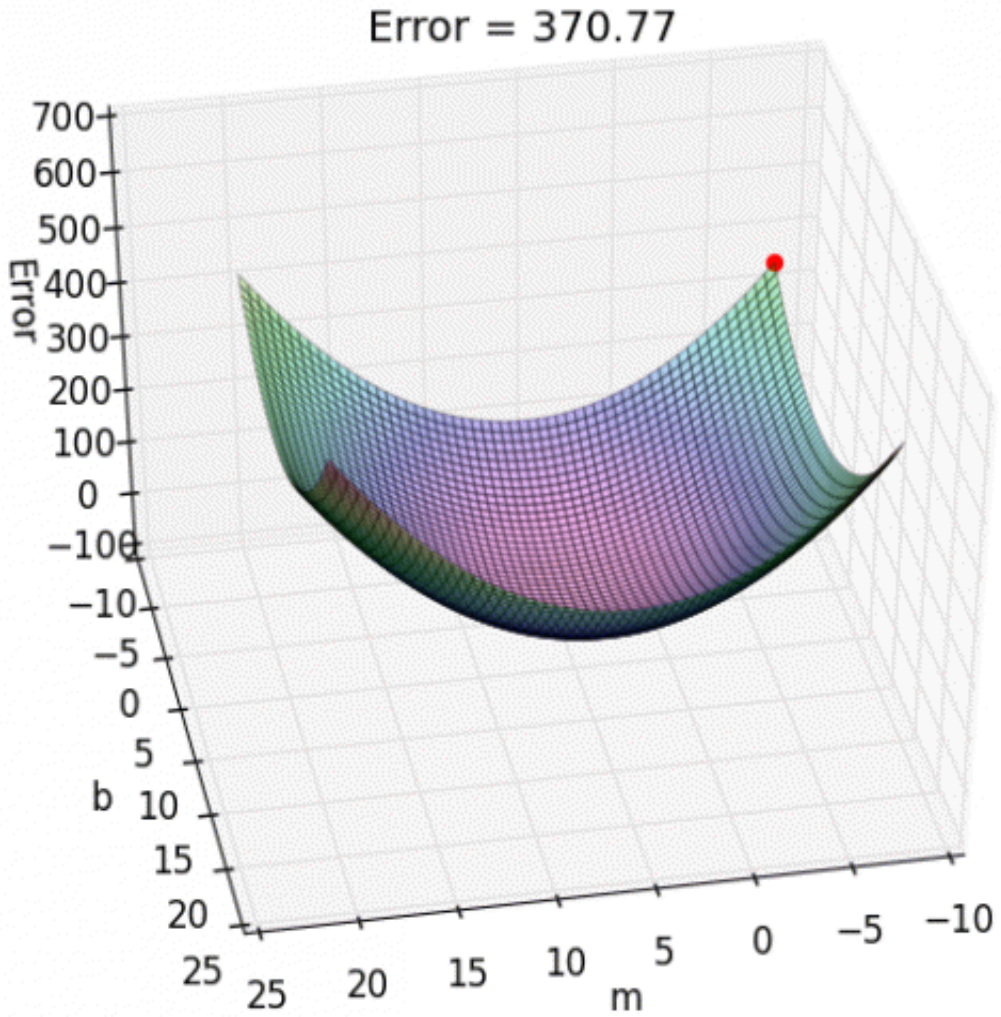
The working method of Gradient Descent

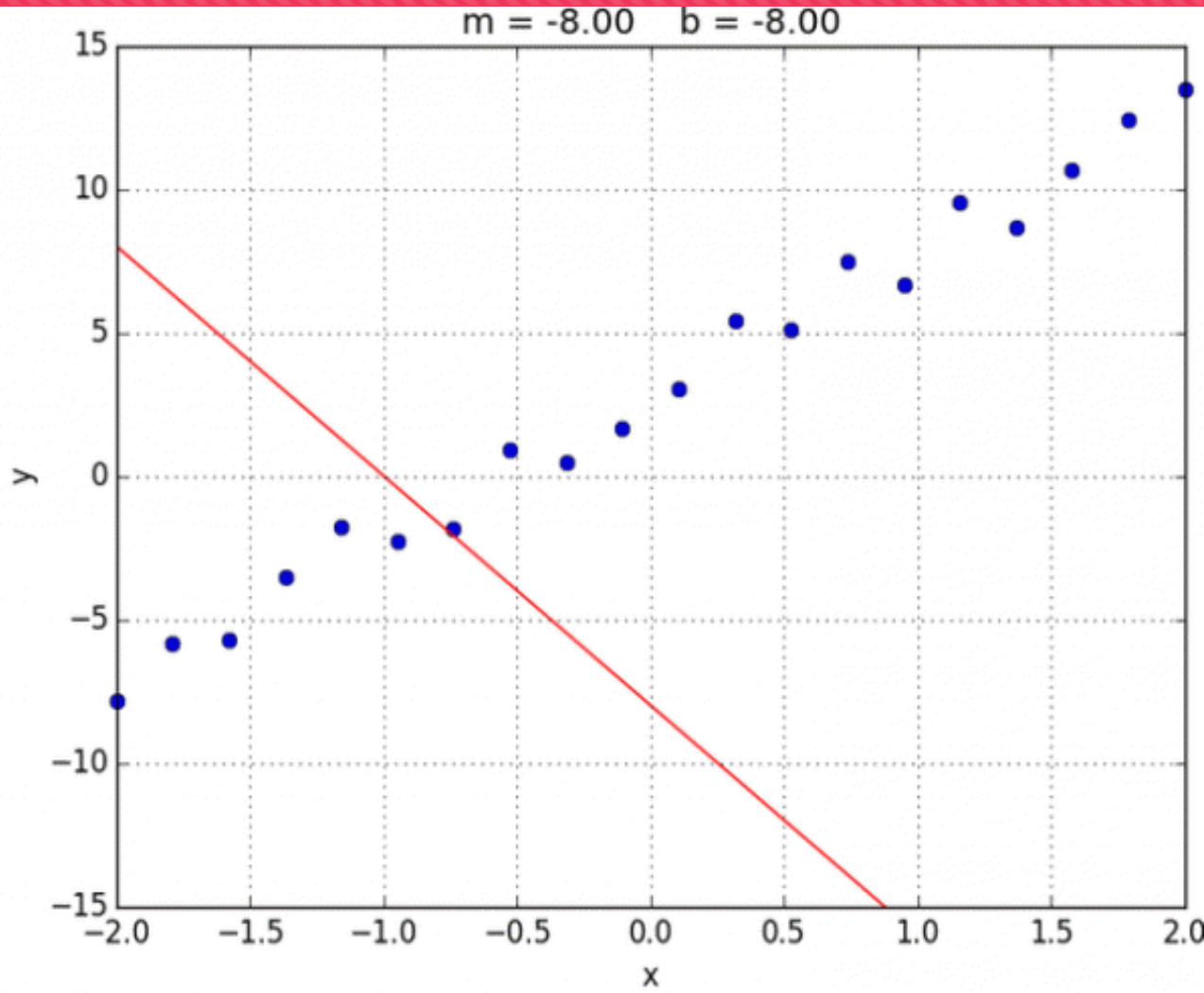


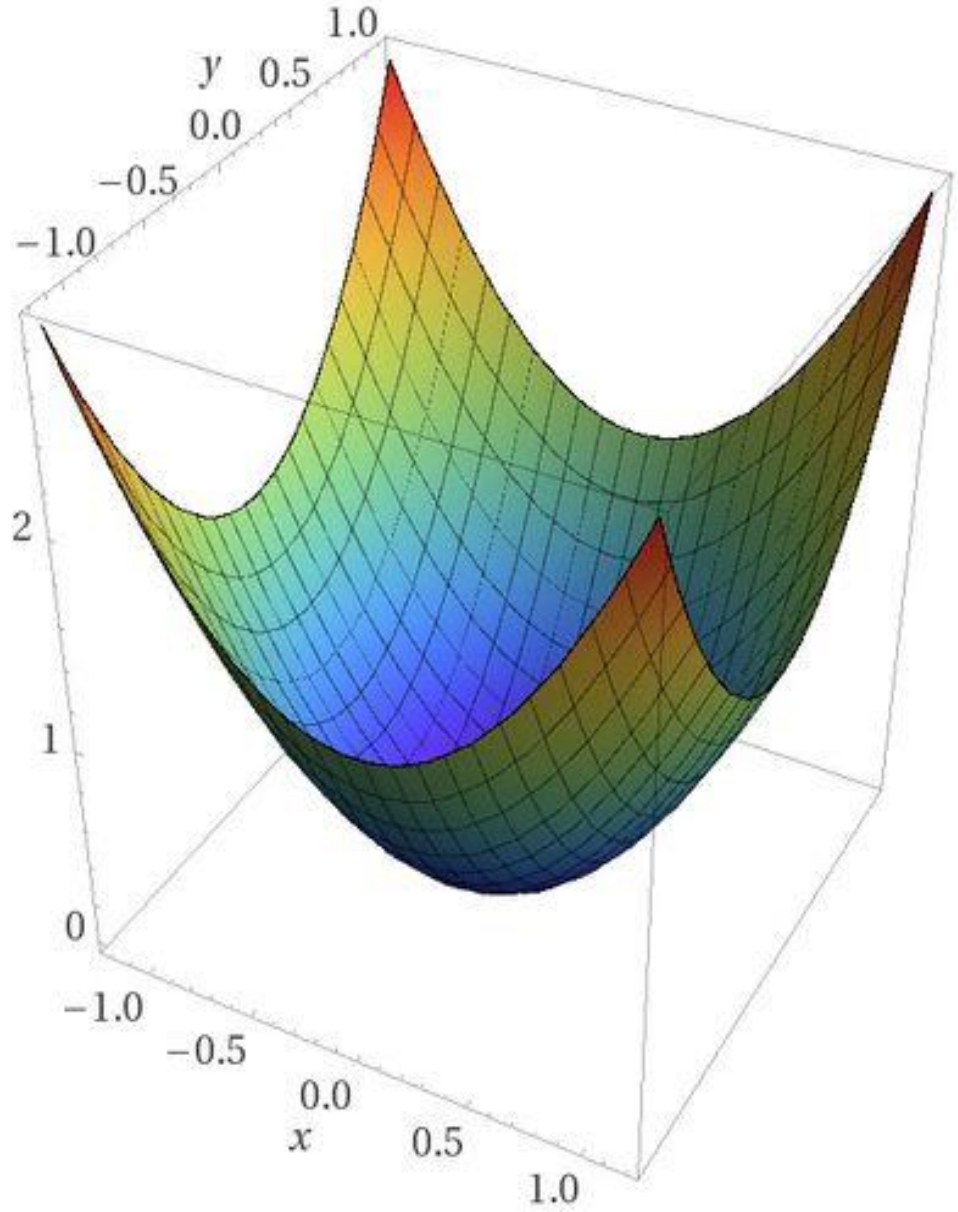
The working method of Gradient Descent



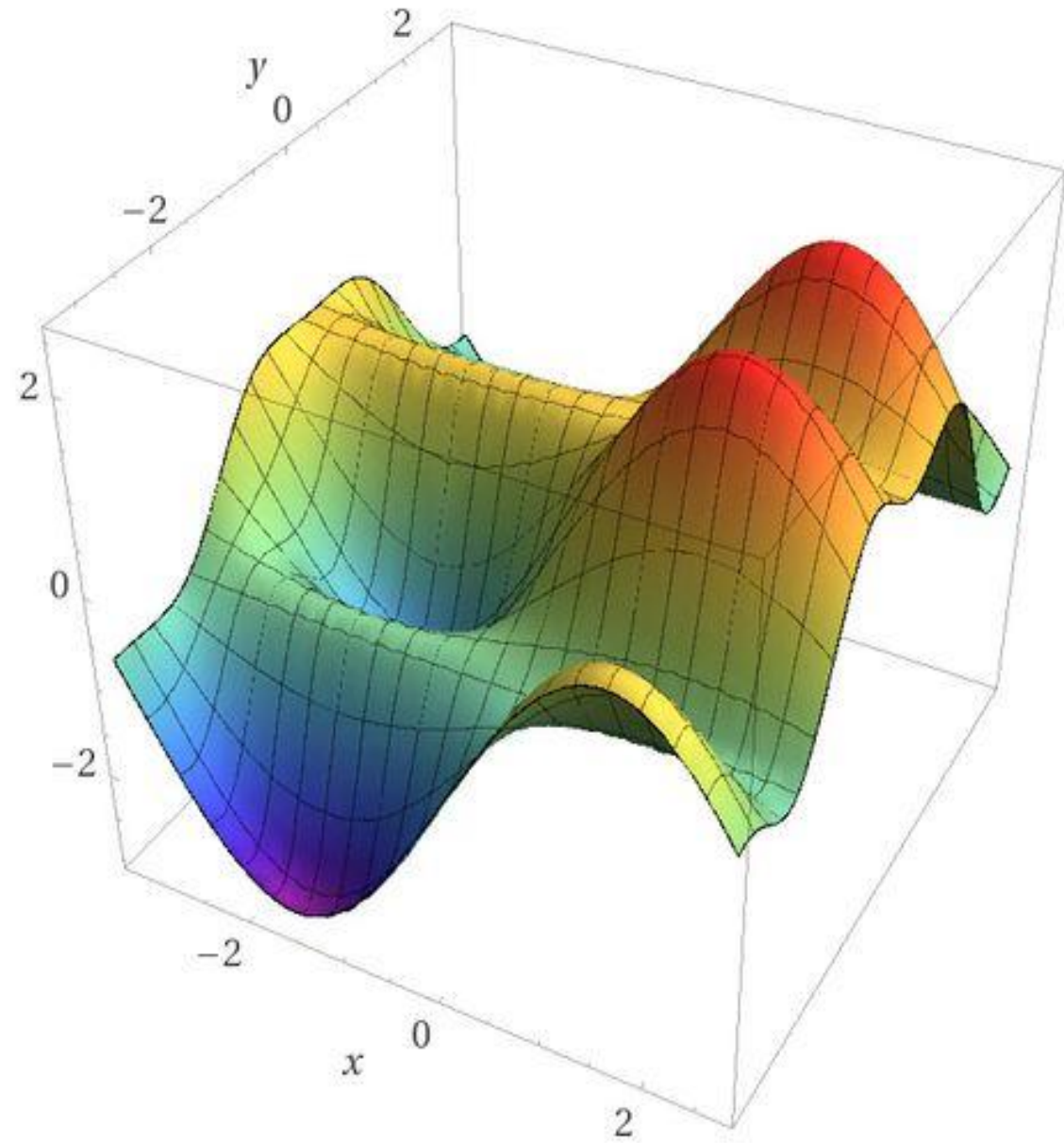
The working method of Gradient Descent





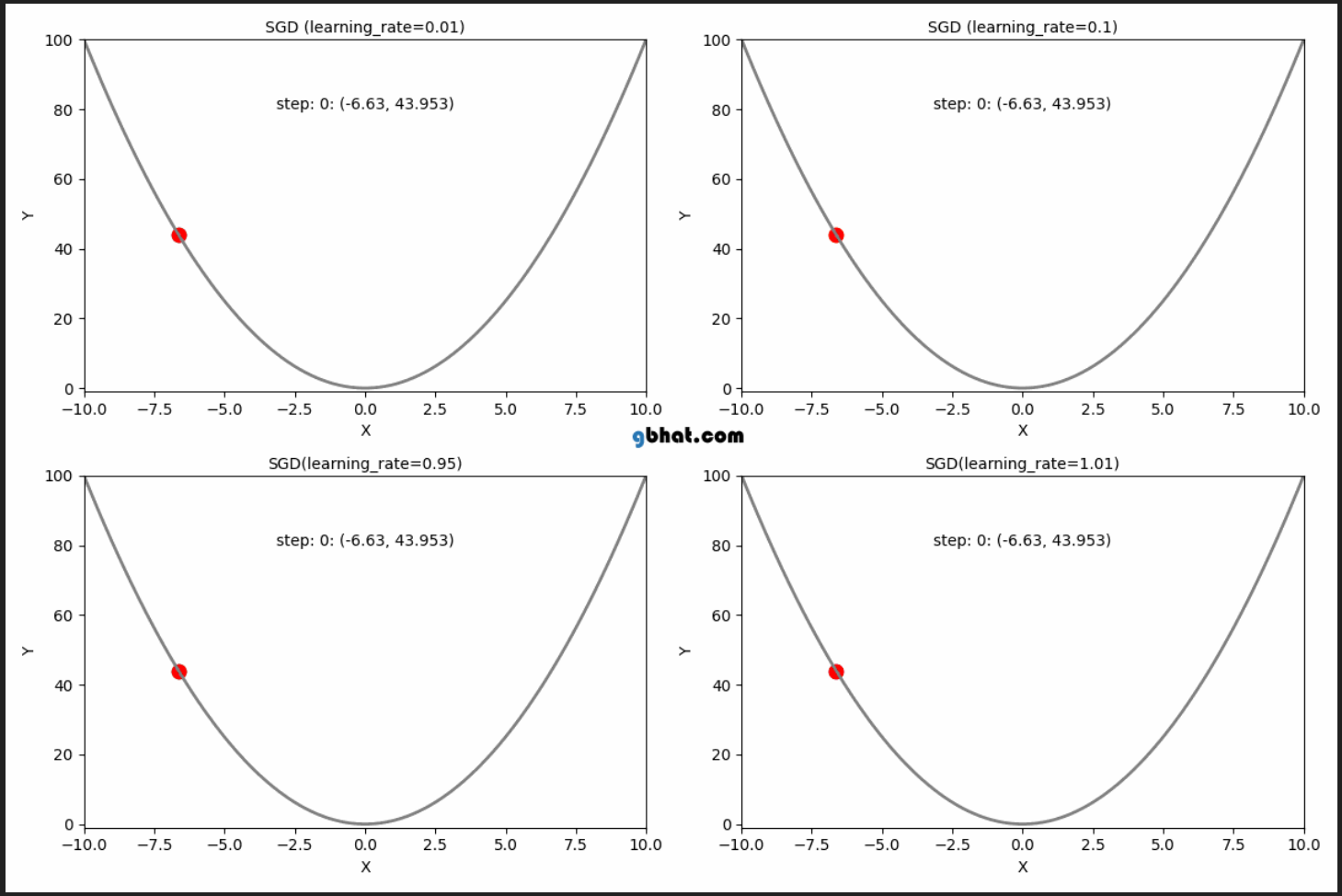


Computed by WolframAlpha



Computed by WolframAlpha

The working method of Gradient Descent

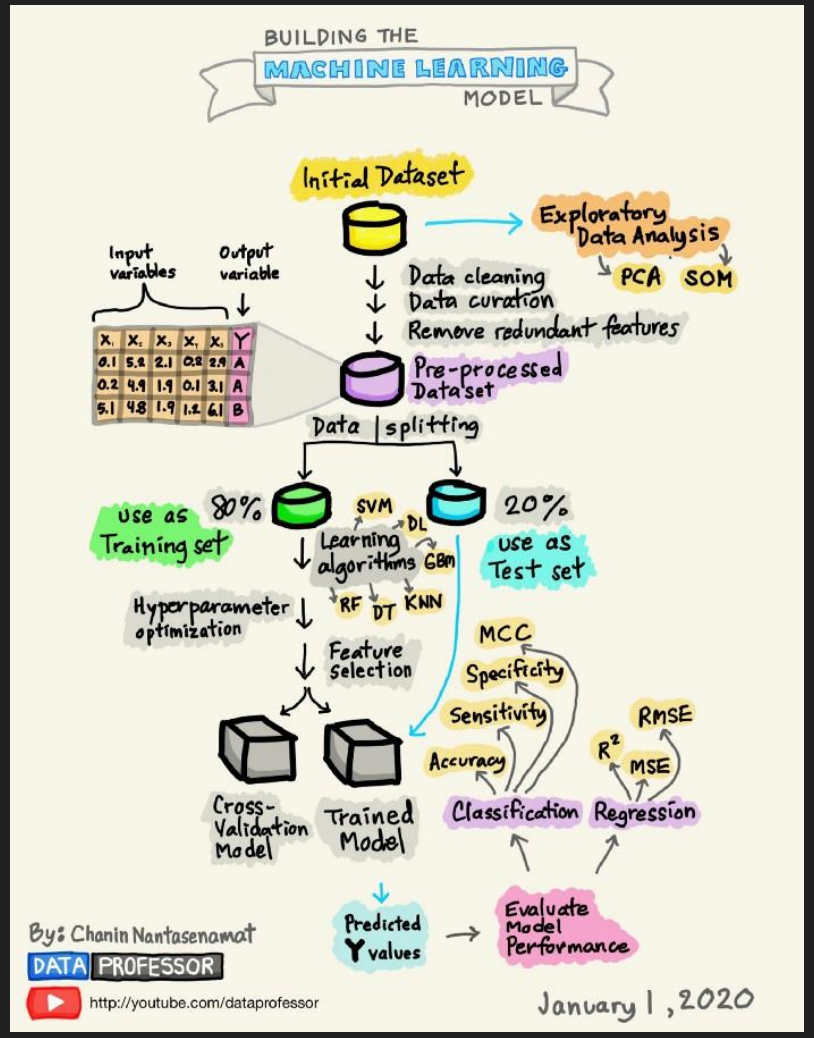


Additional resources

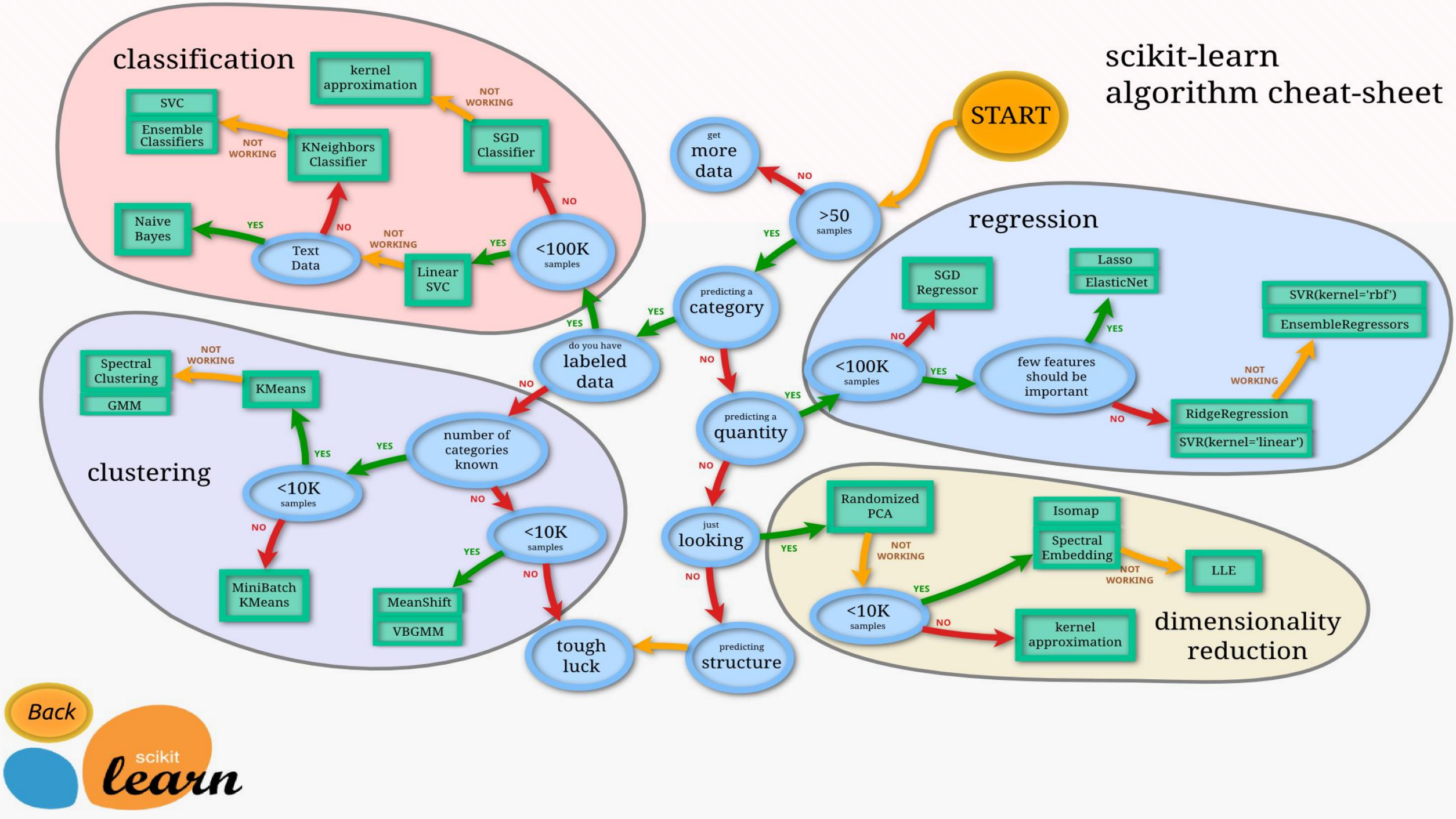
https://towardsdatascience.com/implementing-gradient-descent-in-python-from-scratch-760a8556c31f
https://github.com/Vitomir84/Statistics-and-probability/blob/master/Gradient%20and%20programming%20gradient%20descent.ipynb

https://www.youtube.com/watch?v=sDv4f4s2SB8
https://www.youtube.com/watch?v=vMh0zPT0tLI

Steps to Build a Machine Learning Model



scikit-learn algorithm cheat-sheet



The End

Thank you for your attention. I wish you pleasant times ahead.