Machine Learning crash course

Ahmed Hammad

High Energy Accelerator Research Organization (KEK), Japan.

4Th summer school at CTP, BUE

References

1- Hands-on machine learning with Sikit-Learn and Tensorflow. By Aurélien Géron

2- Probabilistic machine learning. By Kevin P.Murphy

3- Machine learning with Pytorch and scikit-Learn. By Sebastian Raschka, etal

Introduction

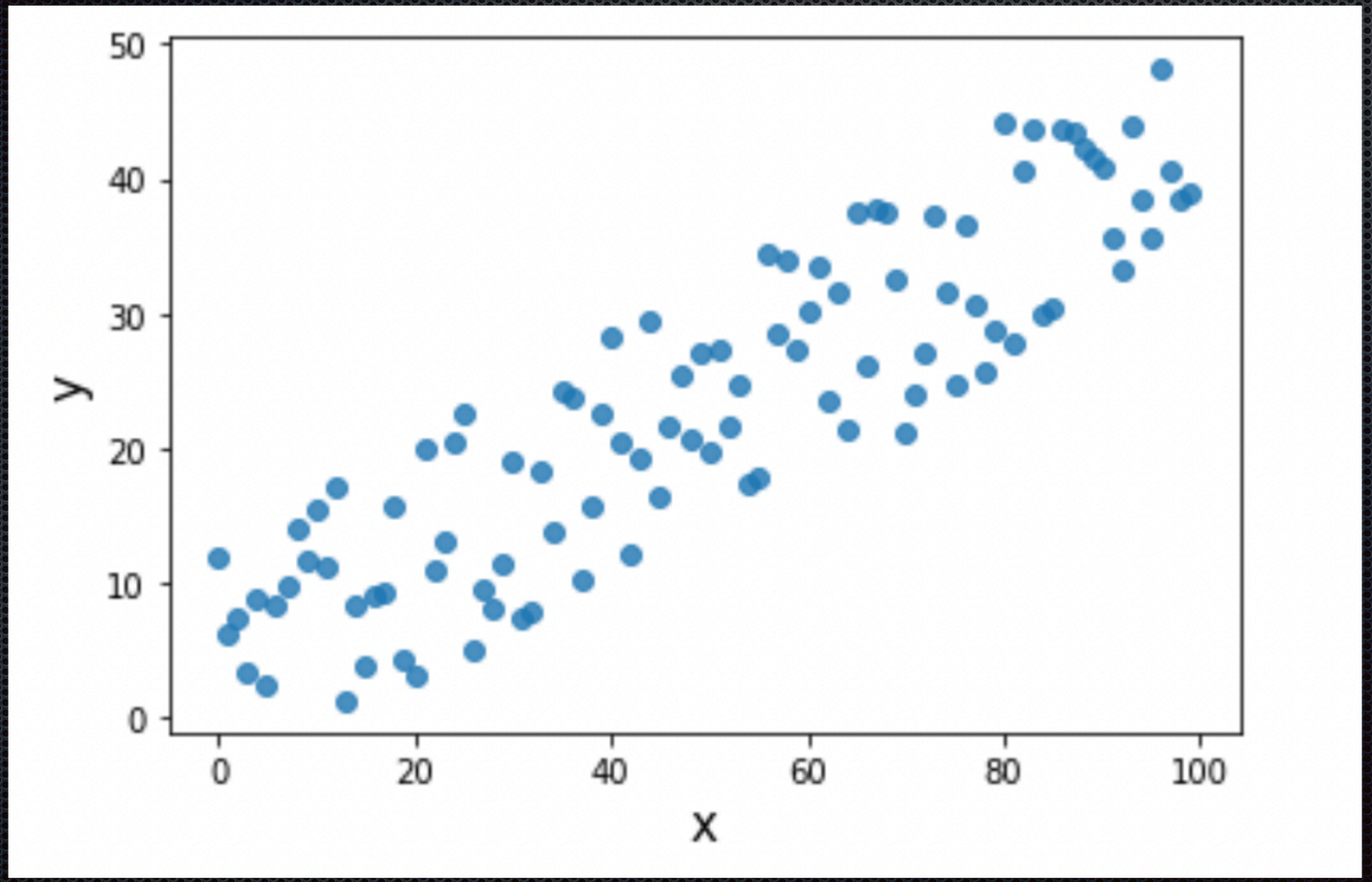
So, what is machine learning ?

$$Y = F(X)$$

Machine learning models approximate the function F for any type of data X

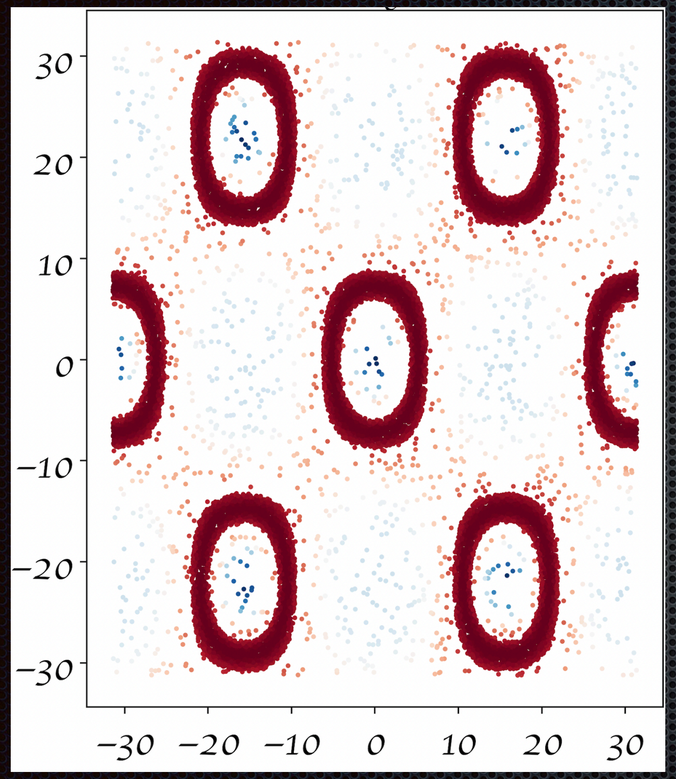
Introduction

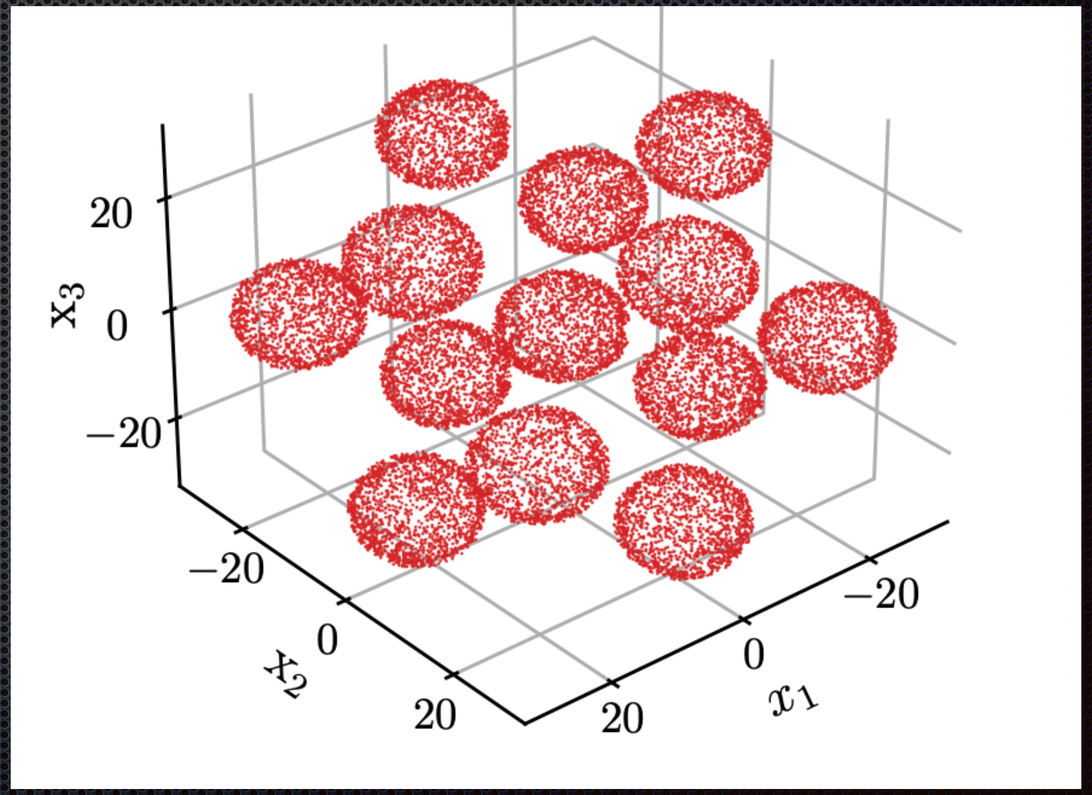
Approximated function can be very simple: One dimension-Linear function



Introduction

Low dimensions - Non-linear function





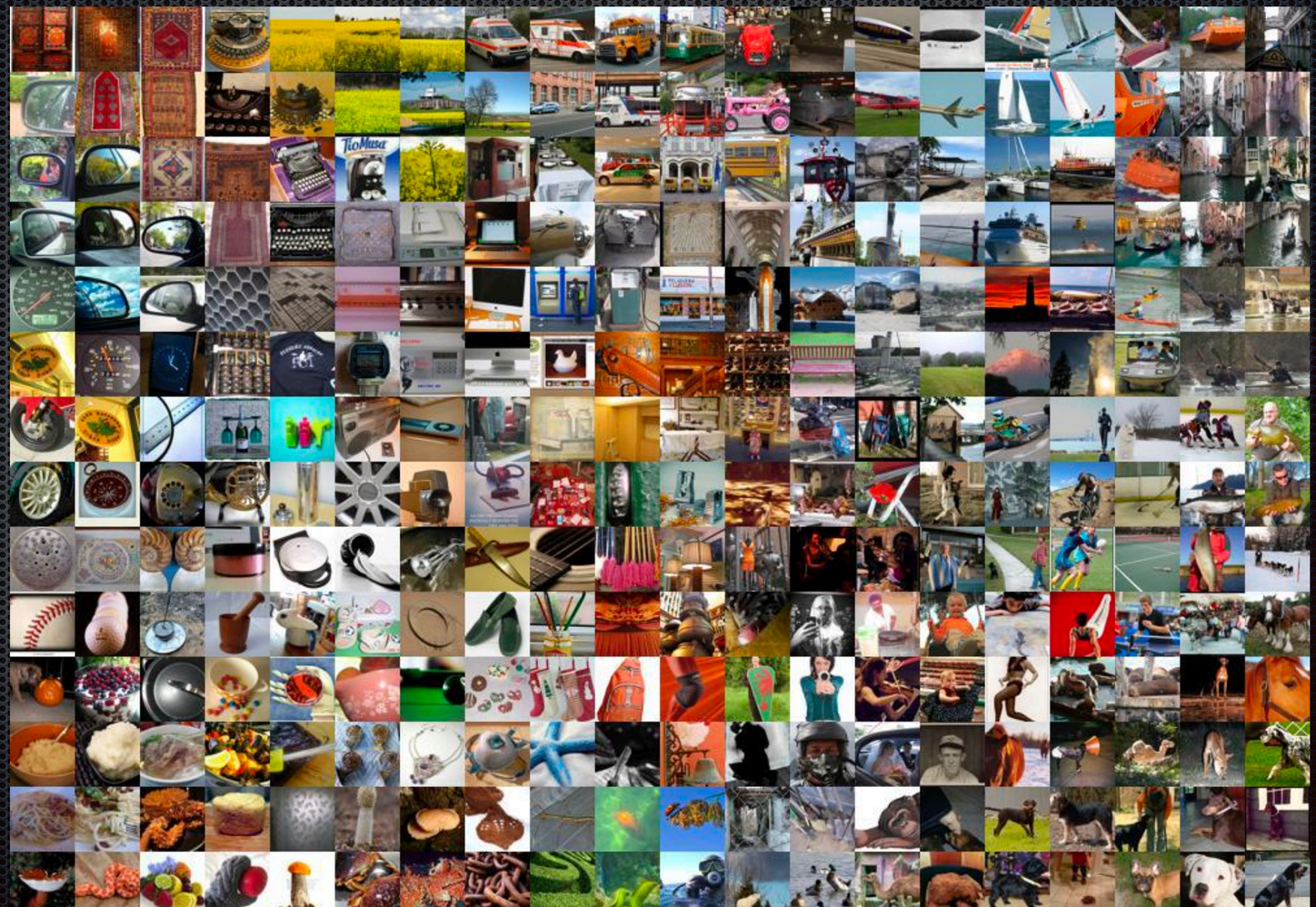
Introduction

ImageNet dataset

Size: 167.6 GB

High dimensions - Non-linear function

$$d = (100000,225,225,3)$$

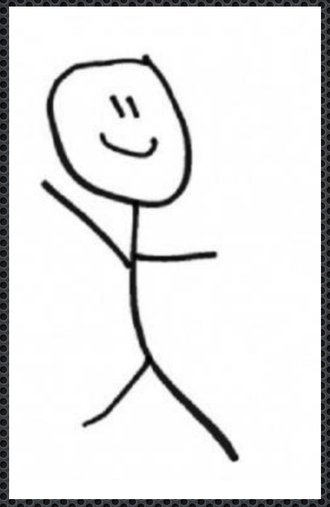


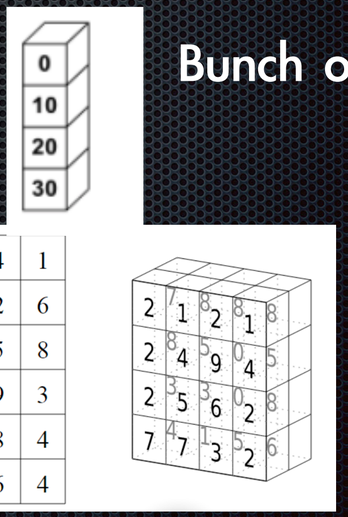
Introduction

How does machine learning approximate F ?

Inputs

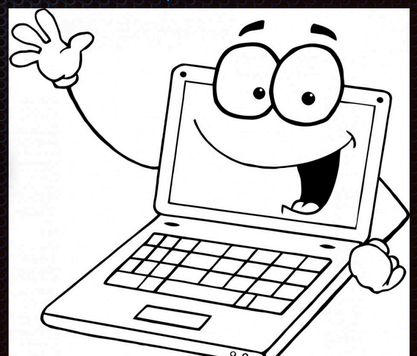
Data scientist





Bunch of numbers

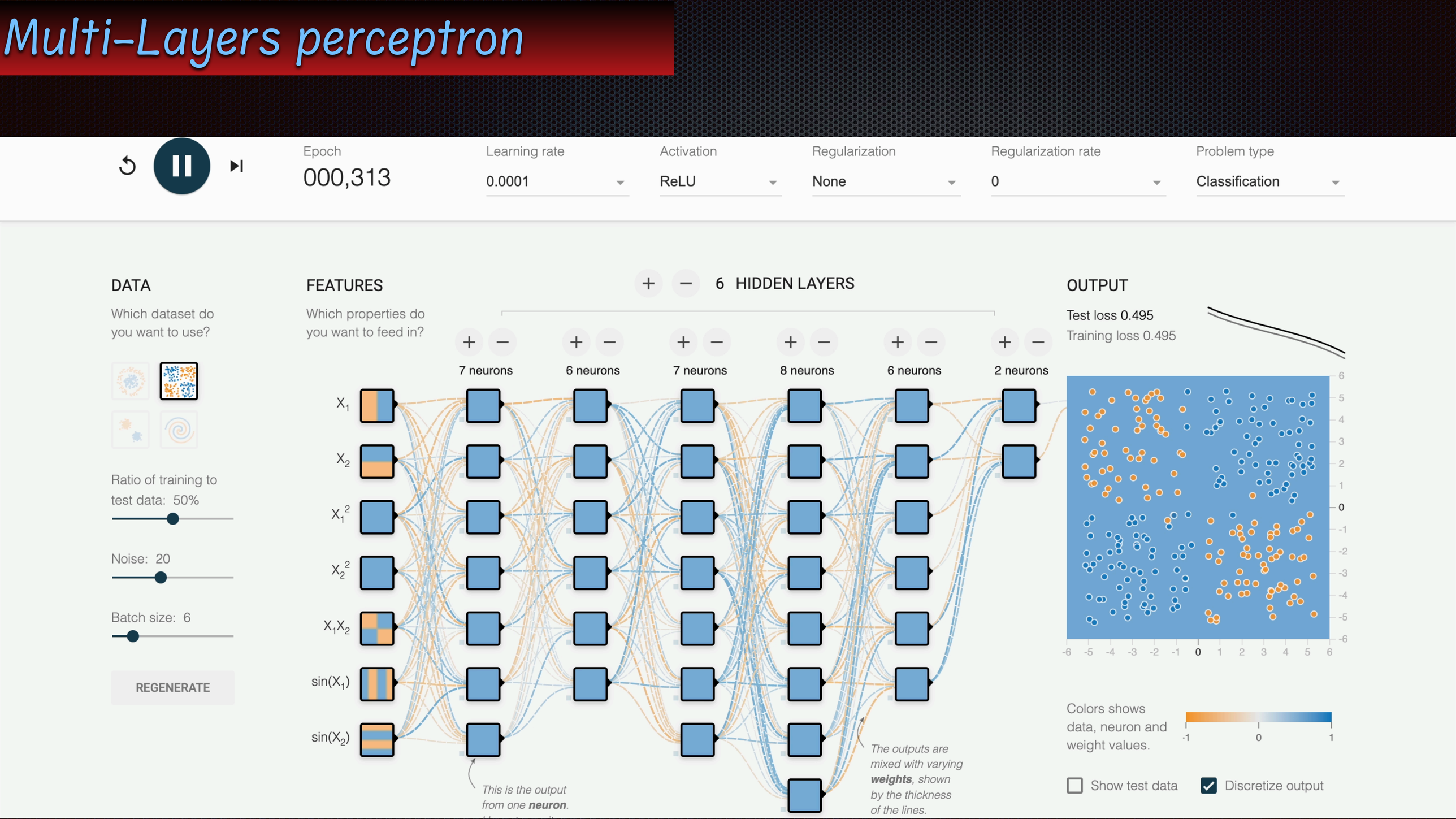
Find a pattern



New data

Prediction

Multi-Layers perceptron

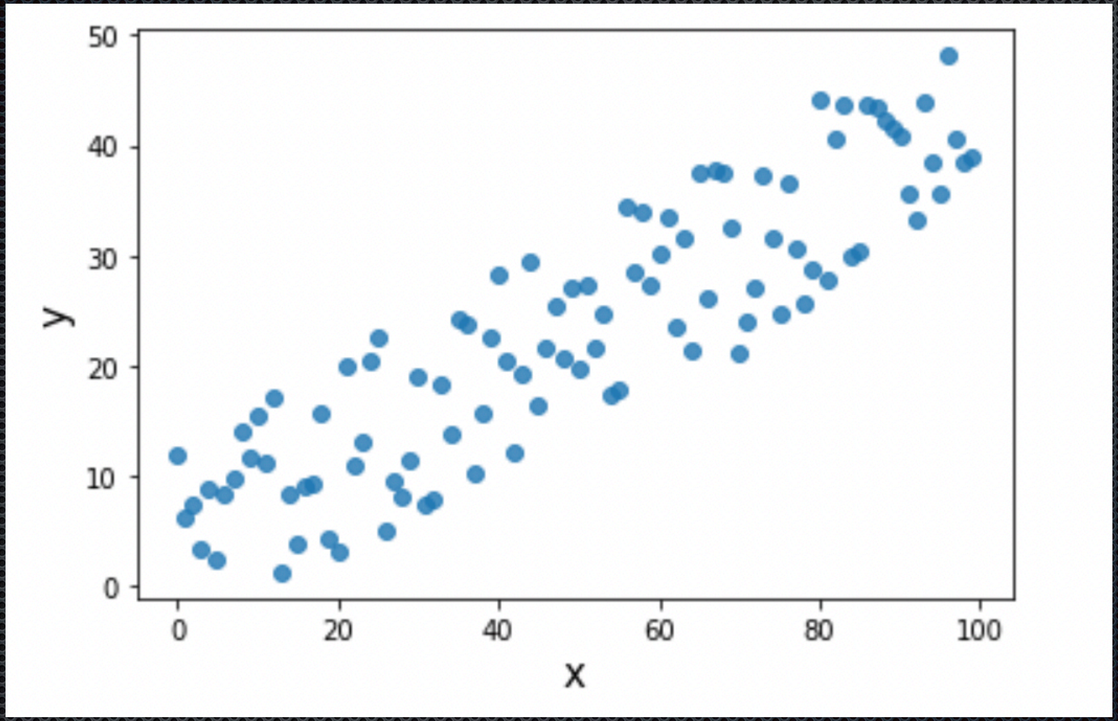


Multi-Layers perceptron

Well, seems good! But how it works ?

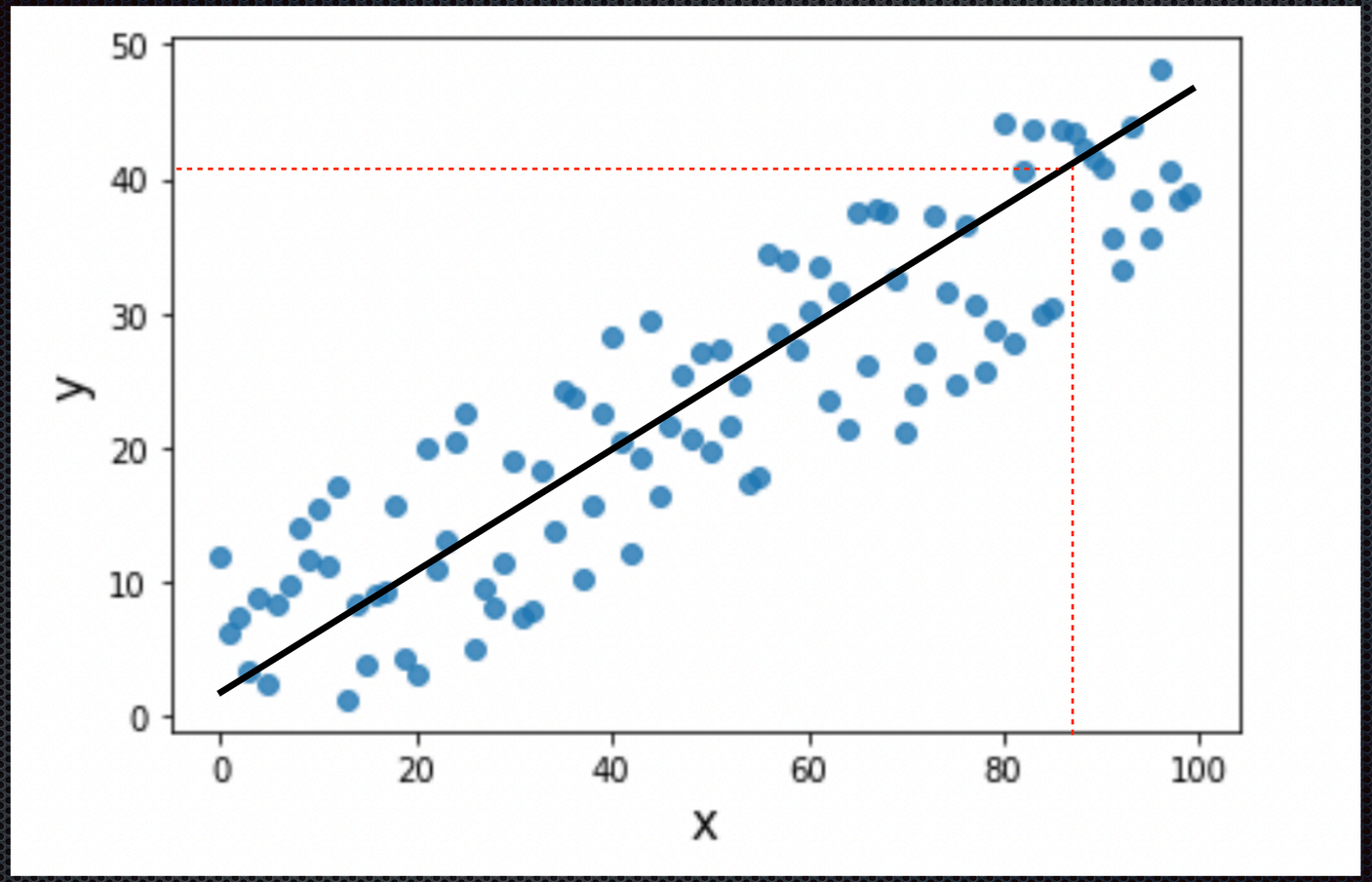
Linear regression

For a continuous data set X



Given new (unseen) x point, can you predict the corresponding Y ?

Linear regression

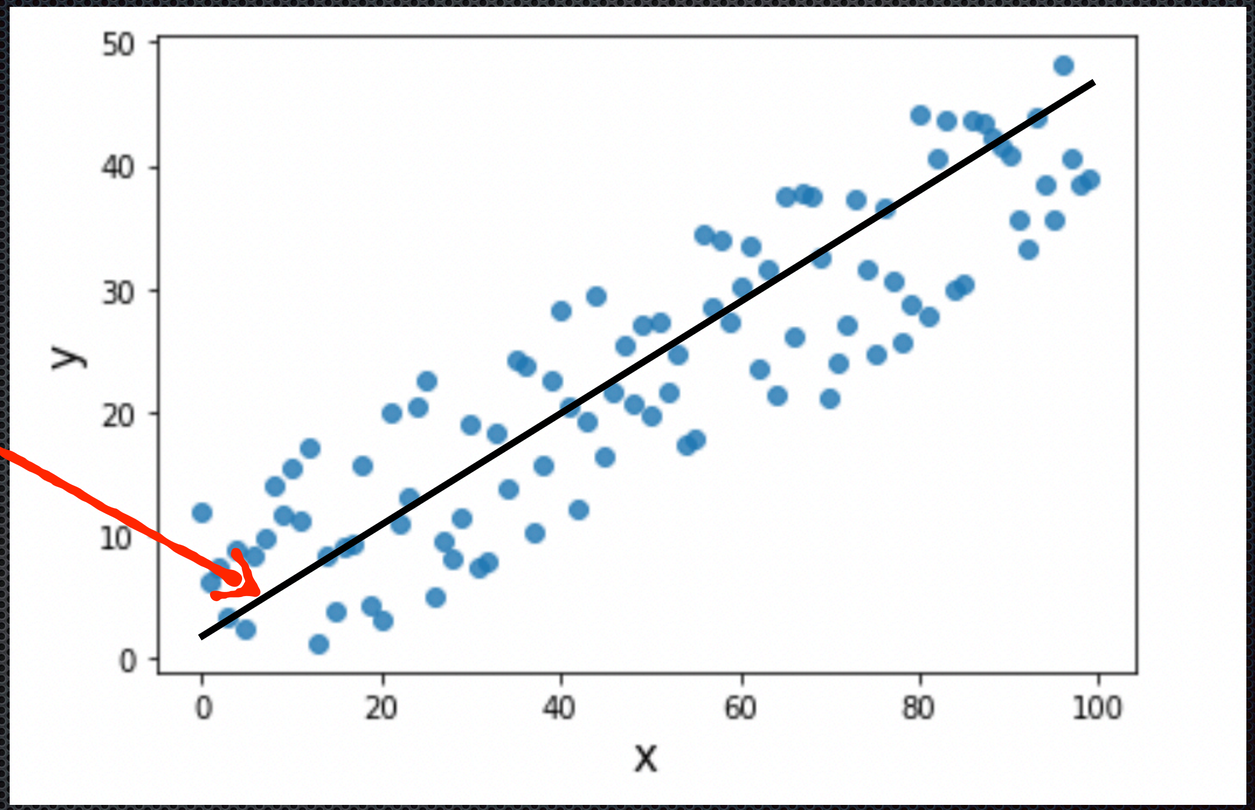


Once we found the line that best fit the data we can interpret the Y for new x points

Linear regression

How to find the line that fit the given data ?

$$\hat{Y} = B0 + B1x$$



With given Y and x we try to find B0 and B1 that fit the line to the data

Linear regression

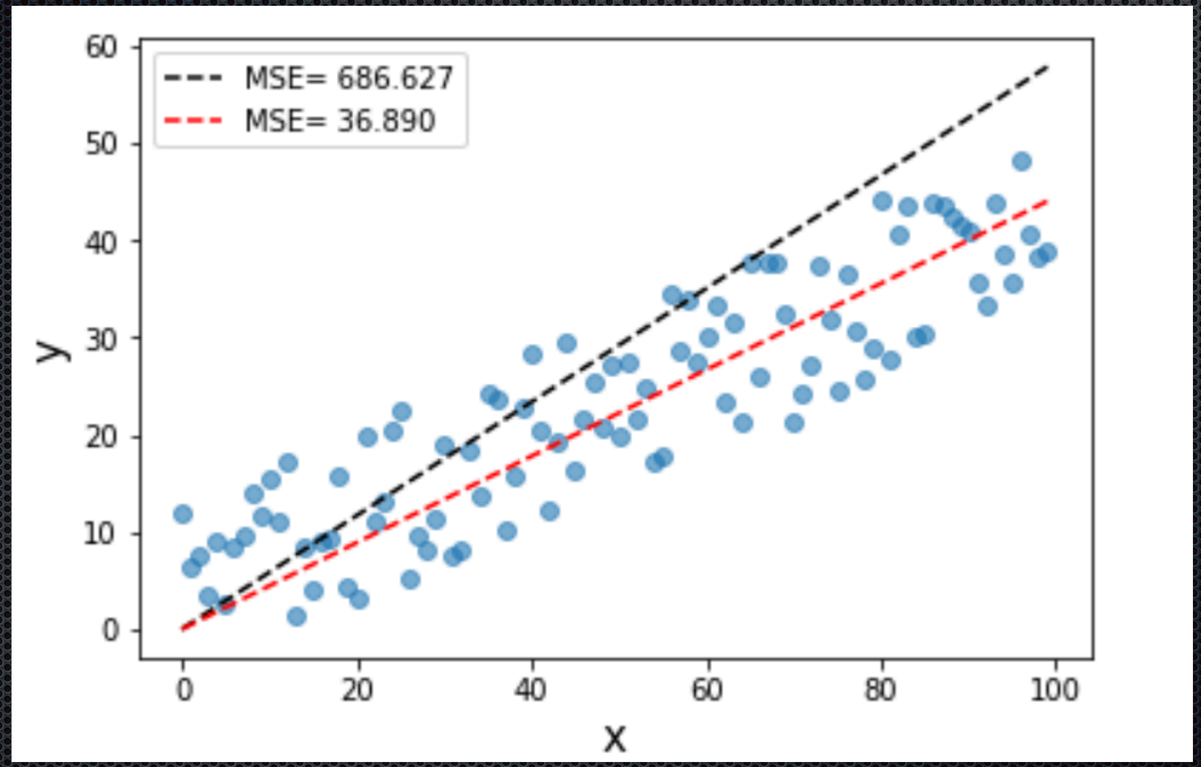
Question:

We can draw infinite number of lines that can fit the data!!
Which line shall we consider ?

Linear regression

Question:

We can draw infinite number of lines that can fit the data!!
Which line shall we consider ?

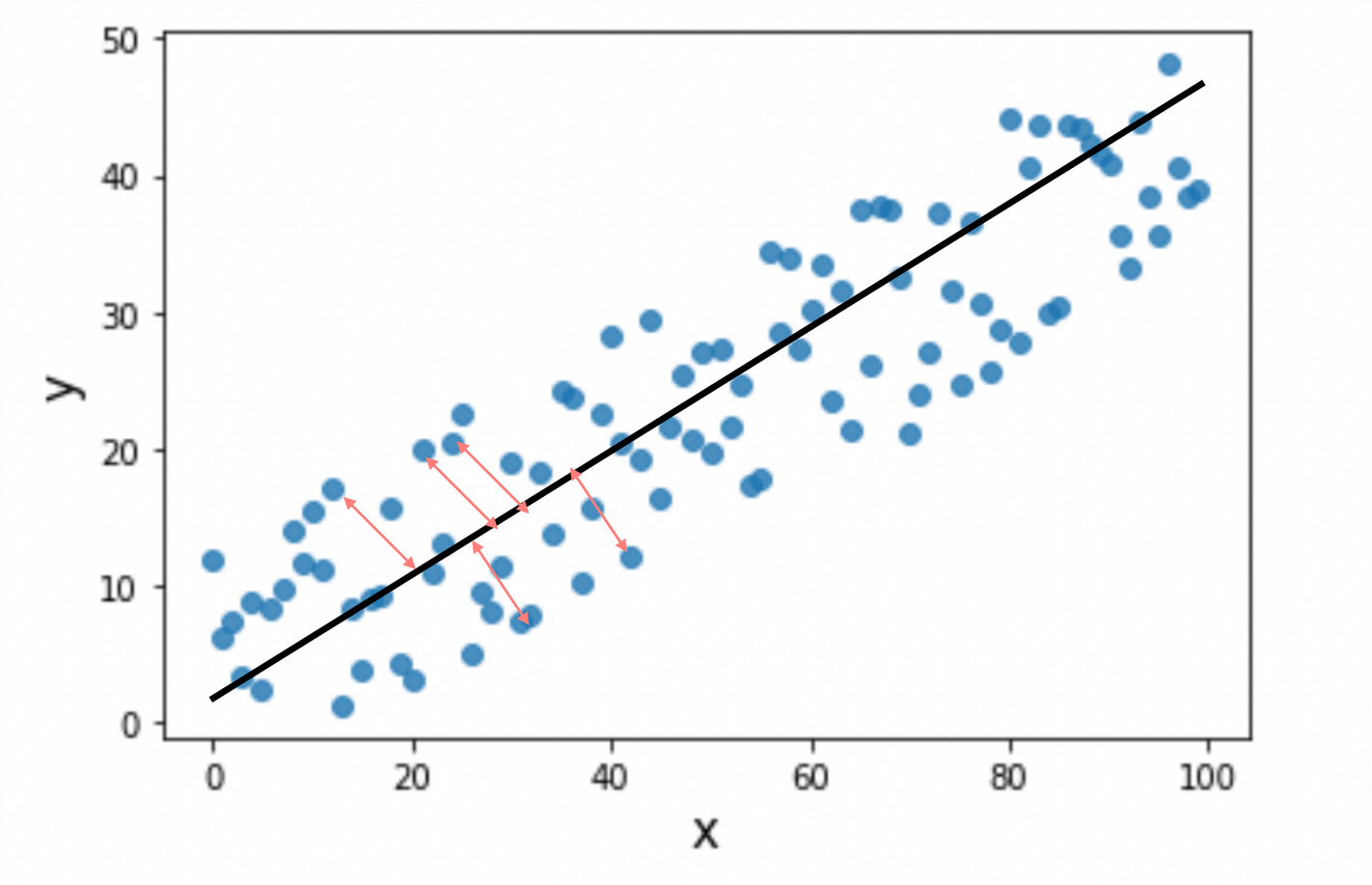


We have to consider the line that minimizes the error function

Linear regression-loss function

Loss function:

It mostly quantifies the difference between the model predictions and the true values



Linear regression-loss function

Types of linear regression loss functions

Absolute mean error:

$$\mathrm{AME} = \frac{1}{m}\sum^{m}\left|Y - \hat{Y}\right| = \frac{1}{m}\sum^{m}|Y - (B0 + B1x)|$$

Mean square error:

$$\mathrm{MSE} = \frac{1}{m}\sum^{m}\left|Y - \hat{Y}\right|^2 = \frac{1}{m}\sum^{m}|Y - (B0 + B1x)|^2$$

Root mean square error:

$$\mathrm{RMSE} = \sqrt{\frac{1}{m}\sum^{m}\left|Y - \hat{Y}\right|^2} = \sqrt{\frac{1}{m}\sum^{m}|Y - (B0 + B1x)|^2}$$

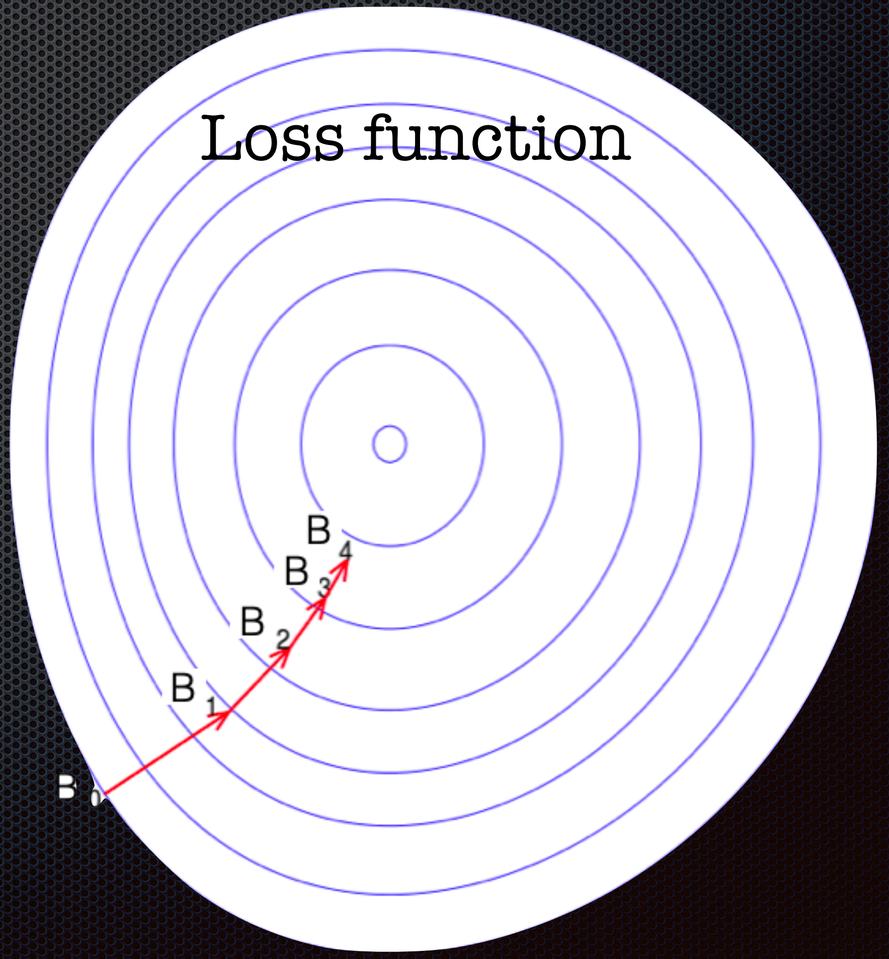
Gradient Descent method

Question:

Well, now we know that to find the best fit line we need to minimize the loss function, but how ??

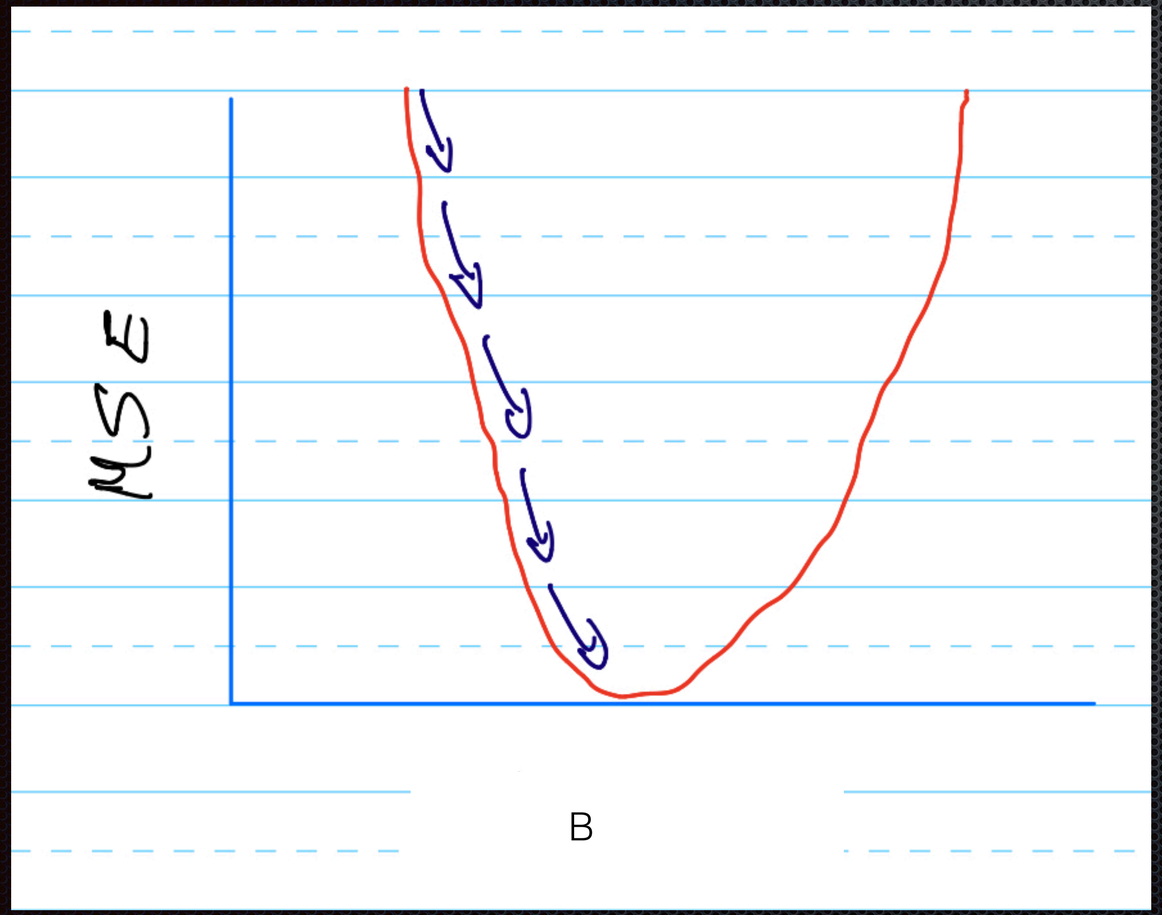
Find the best fit parameters using the gradient descent method

$$B^i_{\text{new}} = B^i_{\text{old}} - \eta \nabla \text{Loss}(B^i_{\text{old}})$$



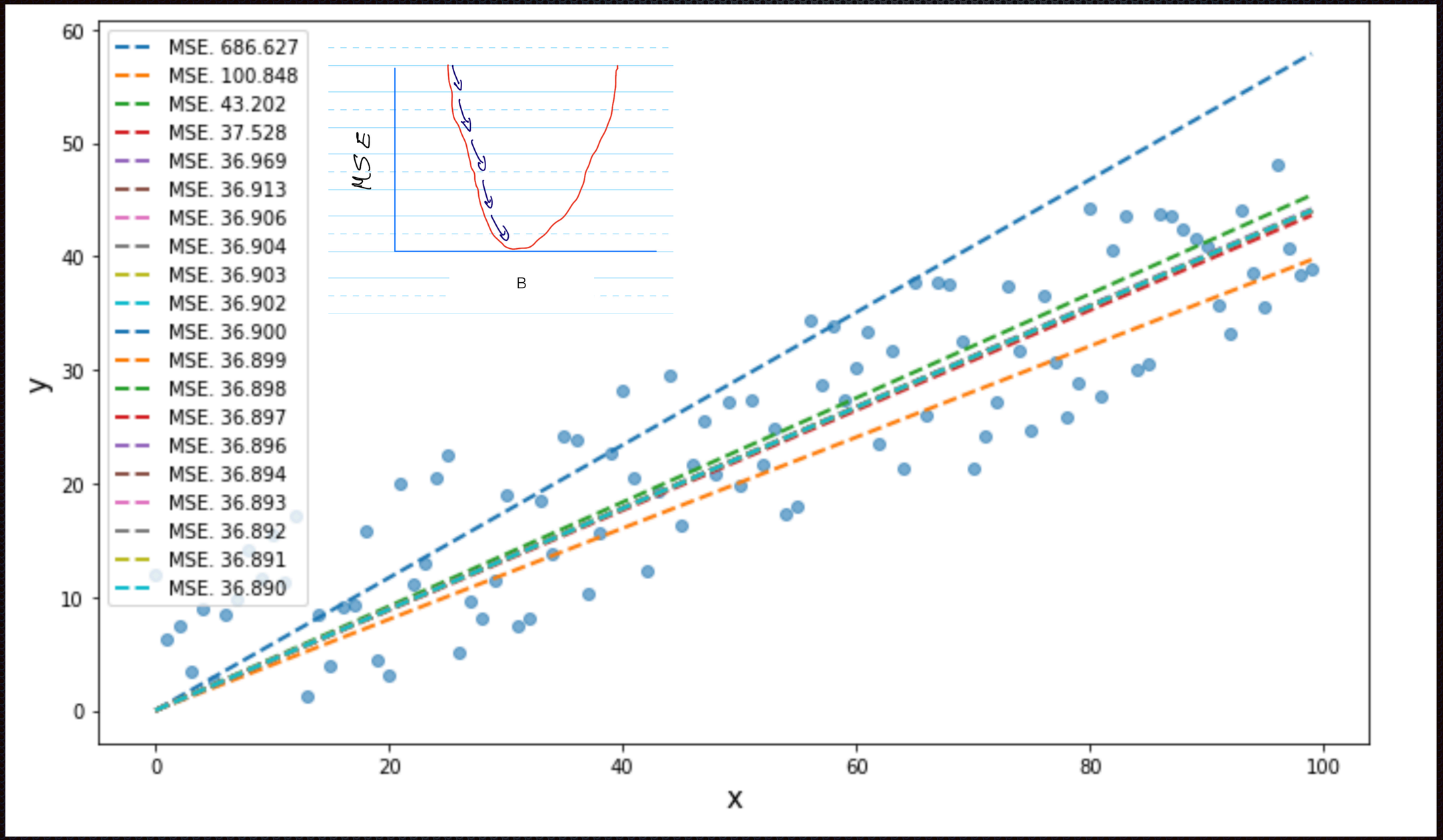
Gradient Descent method

Given a number of iteration, every iteration we update the parameters and calculate the error function until we hit the global minimum



$$B^i_{\text{new}} = B^i_{\text{old}} - \eta \nabla \text{MSE}(B^i_{\text{old}})$$

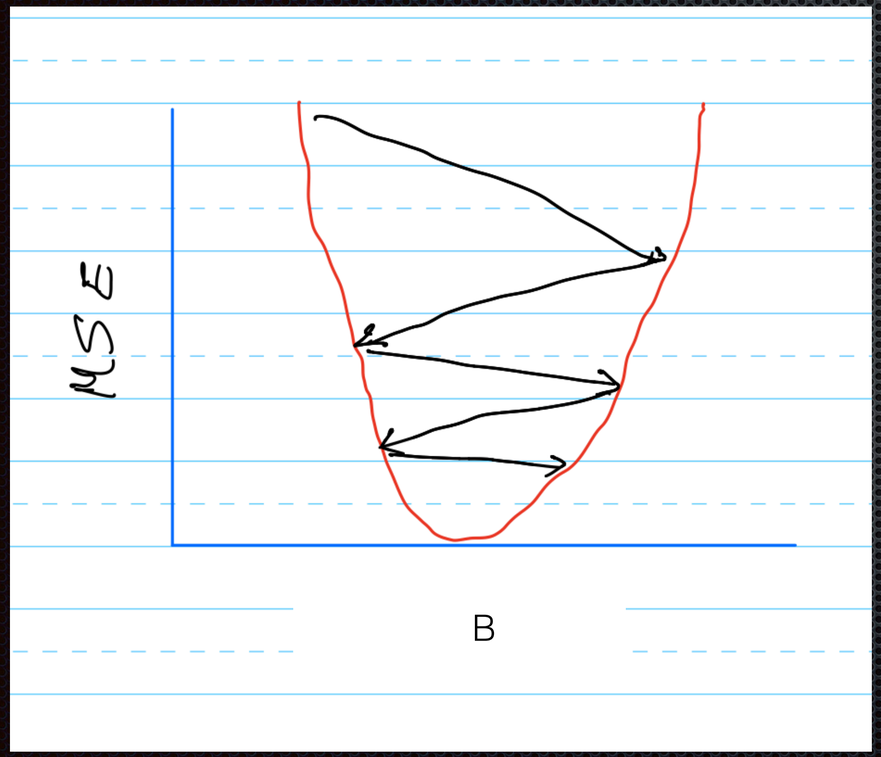
Gradient Descent method



Learning rate

$$B^i_{\text{new}} = B^i_{\text{old}} - \eta \nabla \text{Loss}(B^i_{\text{old}})$$

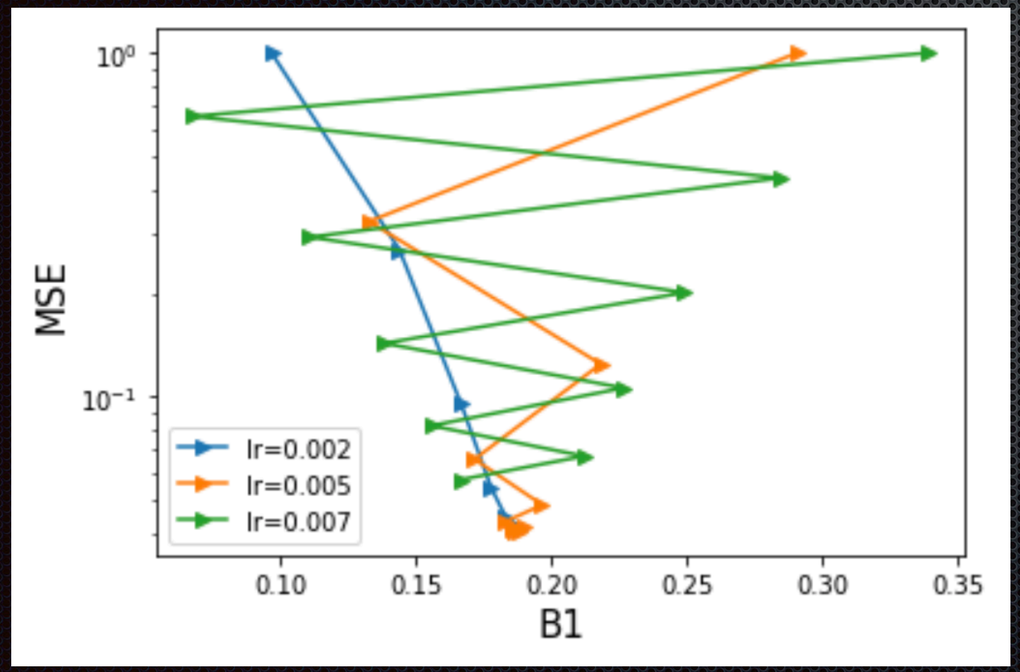
η Is called the learning rate which controls the descent rate of the loss function

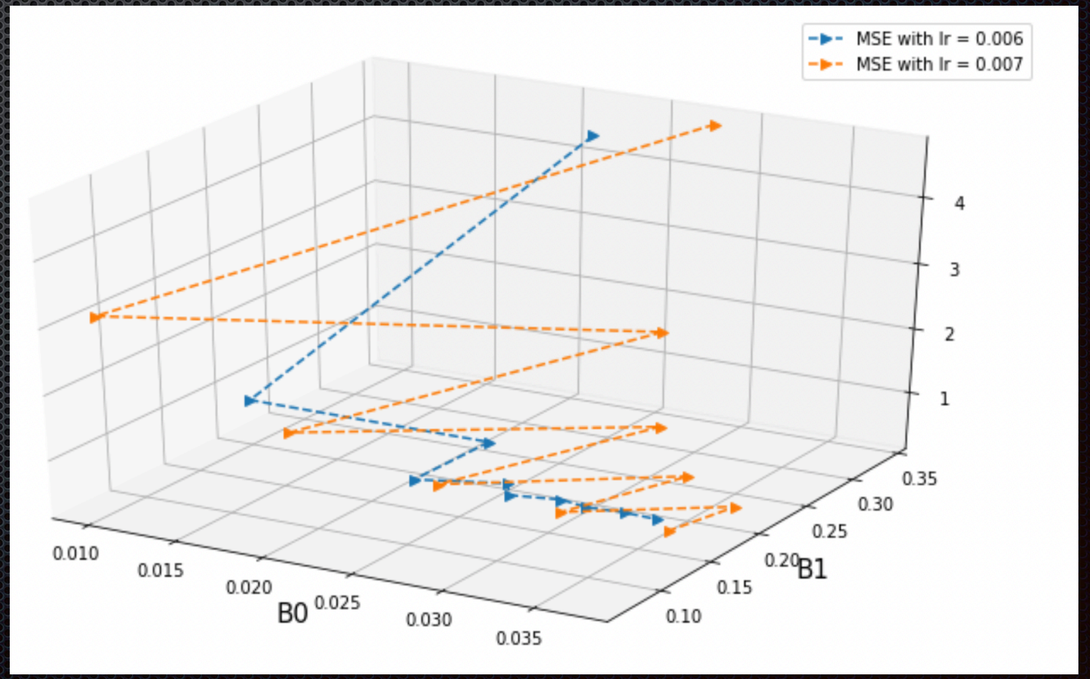


For large learning rate the function oscillates and we will not able to reach the minimum

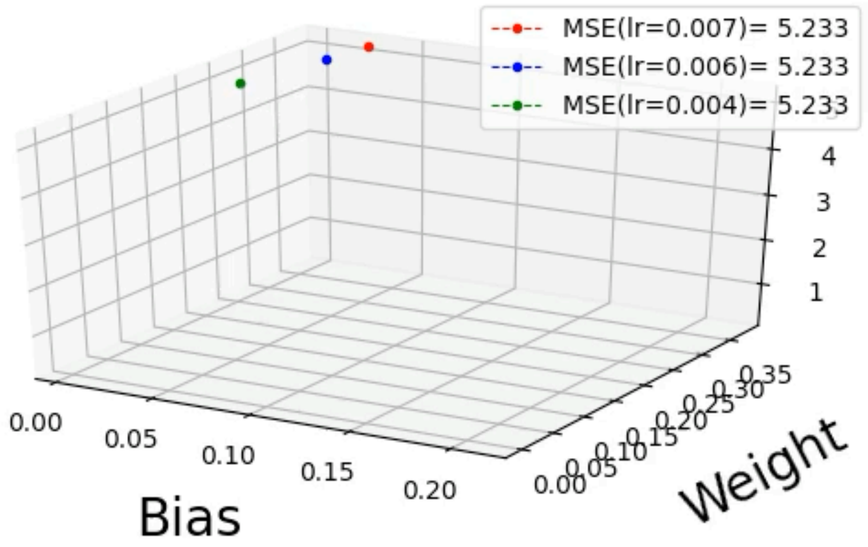
Learning rate

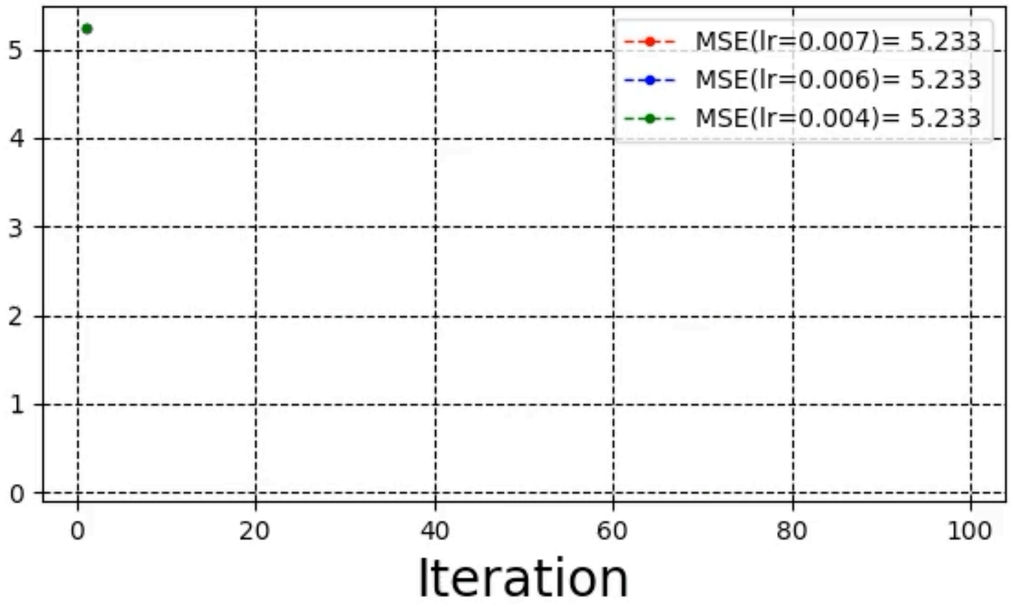
Effect of different learning rates

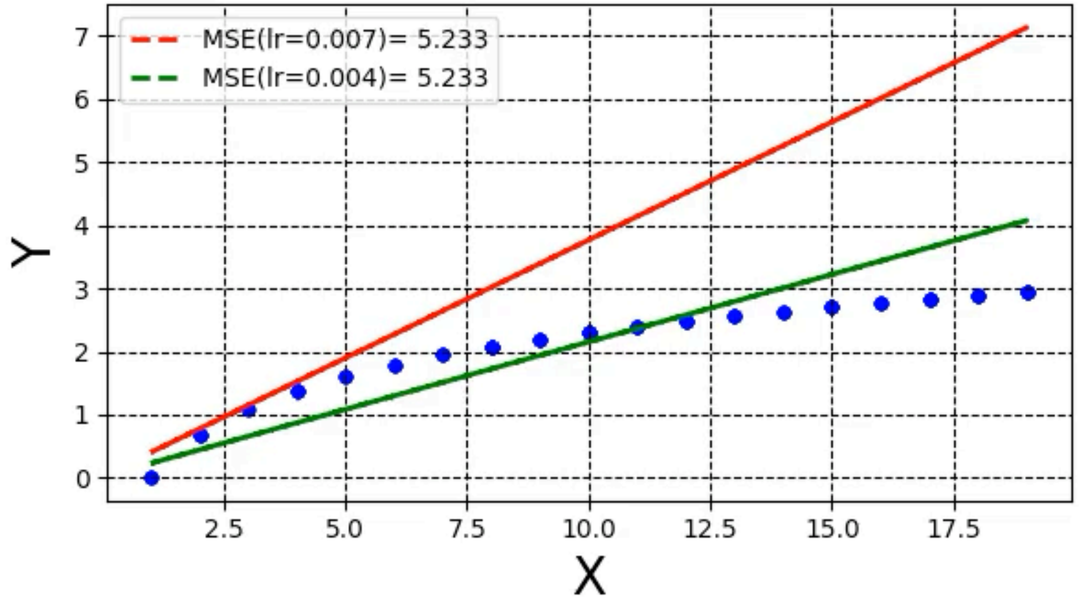


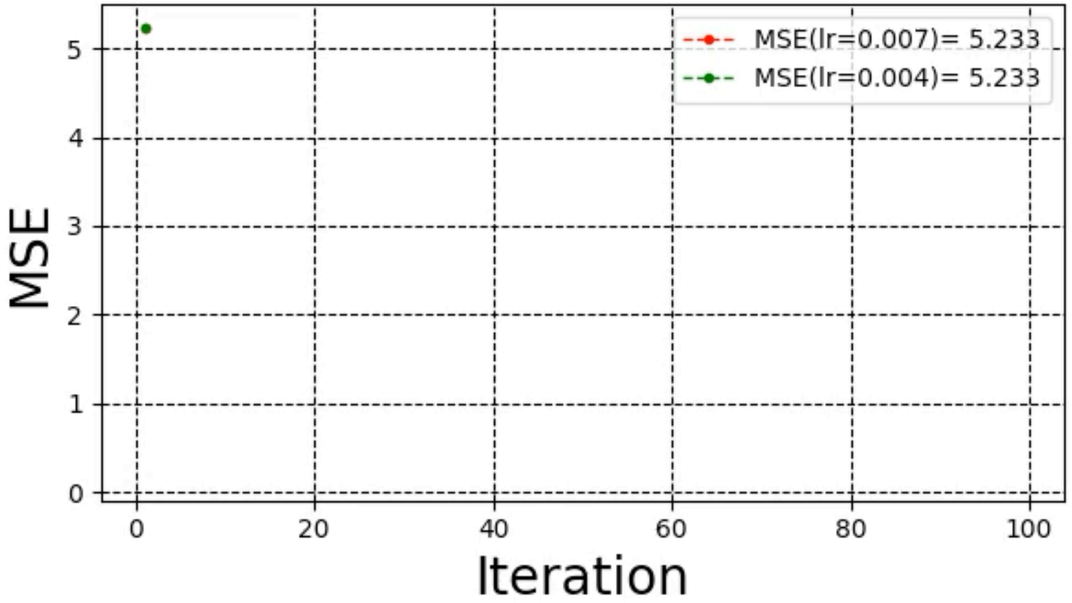


Learning rate









To be continued…