# Gradient descent

JrPhy

Introduction

- Artificial Intelligent (AI) is famous today, and its fundamental algorithm is gradient descent, the other algorithms base on this method, like conjugate gradient descent, stochastic gradient descent, ada gradient descent, etc.

- The core idea of gradient descent is finding the minimum, it can be found by solving normal equation then uses QR factorized. Although normal equation gets global minimum definitely, it spends much time, so AI uses gradient descent to modify for the efficiency and accuracy.

Concept

- For a square equation, there is a global maximum or minimum. We can use derivative to get it.

- $y = 2x^2+4x+3$ ➔ $y' = 4x+4 = 0$ minimum is (-1,1)

- Since computer can't compute derivative, so we use finite difference.

$$y' = \frac{dy}{dx} \approx \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}$$

- Its meaning is slope, then choose negative direction to find the minimum.
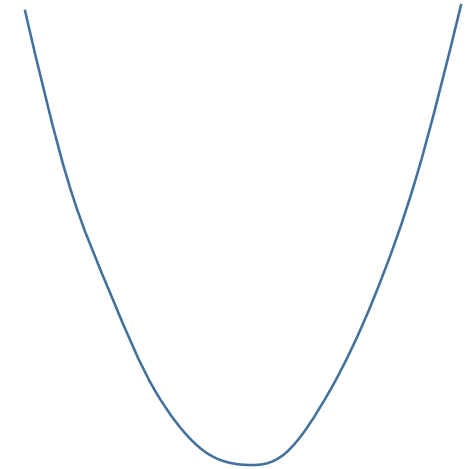
Concept

- Here times a constant λ to try the convergent rate, there is a best constant or function that convergent rate is fastest.

-       too slow   may not converge   the best
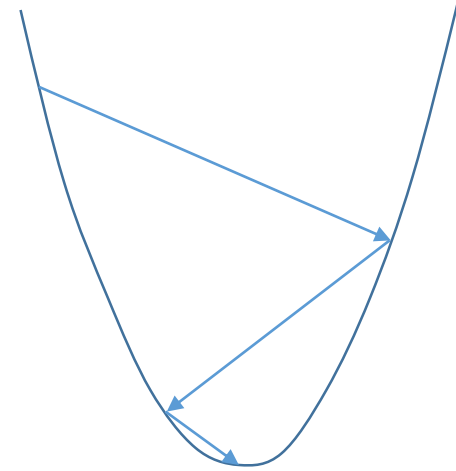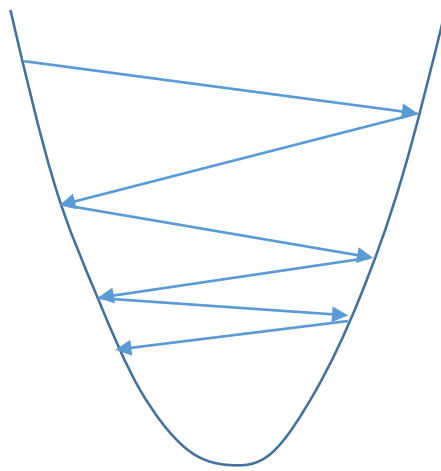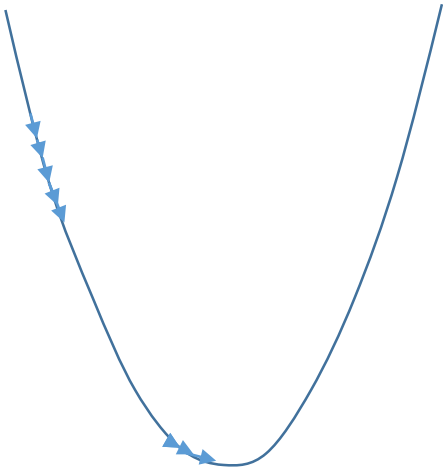
Algorithm

- Guess a initial x_0 and λ, then give x_1 to find y'. So

$$x_1' = x_0 - \lambda y'$$

- Because x is a float number, so remember to define error to set the condition.

- The other detail is in the code.

Result

- $y = f(x) = x^2 - 4x + 2$, global minimum at $(2, -2)$
- $x_0 = 10$, $\lambda = 0.1$ iterators times 57
- The local minimum is: 2.000003, The value is: -2.000000
- You can try my code to change x_0 and λ to see the iterators times