Least Square & Gradient Descent

Loss/Cost Function: >>

>> to optimize by minimizing loss function

Least Square >> find min J

Gradient Descent >>

Take partial derivative >>

>> set a step length α

Then

Each time after a step, we need to update our θi

That is

Which means, the current step direction depends on the samples, and

Can also be seen as a constant

In machine learning algorithms, we usually use matrix method to do calculations

Therefore, here we regard the former functions as a matrix function with

**Υ.hats** as a matrix transformed by sample matrix **X** and parameter vector **θ**, that is **Y.hats** = h**θ**(**X**) = **Xθ**

Then, the loss function becomes

We take the gradients/ partial derivatives for the J(**θ**), then get:

With the update of θ, the matrix expression function can be written as

Notes for Gradient Descent to Optimize models

* The choice of step length
* The choice of initial value
* Regularization

Gradient Descent Family

>> Batch GD, Stochastic GD, Mini Batch GD