

# Machine Learning Homework 1.2

## 一、实验题目及实验要求：

- 1、Generate  $n = 2,000$  points uniformly at random in the two-dimensional unit square. Which point do you expect the centroid to be?
- 2、What objective does the centroid of the points optimize?
- 3、Apply gradient descent (GD) to find the centroid.
- 4、Apply stochastic gradient descent (SGD) to find the centroid. Can you say in simple words, what the algorithm is doing?

## 二、实验过程：

- 1、编程语言：python
- 2、IDE：PyCharm
- 3、实验思路：

1) 在二维单位正方形内随机生成 2000 个点，通过梯度下降法和随机梯度下降法在二维单位正方形找到一个点使得这个点到随机生成的 2000 个点的距离最小（图心）。

2) 梯度下降法设置损失函数为：

$$f(x, y) = \sum_{i=1}^N \sqrt{(x - x_i)^2 + (y - y_i)^2}$$

3) 对上述损失函数求偏导数：

$$dx = \sum_{i=1}^N \frac{x - x_i}{\sqrt{(x - x_i)^2 + (y - y_i)^2}}$$

$$dy = \sum_{i=1}^N \frac{y - y_i}{\sqrt{(x - x_i)^2 + (y - y_i)^2}}$$

从  $[0, 1]$  出发寻找图心，通过梯度下降法更新要寻找的点：

$$x = x - \alpha * dx$$

$$y = y - \alpha * dy$$

直到前后更新的点的损失函数之差  $f_n - f_{n-1}$  的绝对值小于阈值

0.00001。

4) 随机梯度下降法设置损失函数为：

$$f(x, y) = \sqrt{(x - x_i)^2 + (y - y_i)^2}$$

5) 对上述损失函数求偏导数：

$$dx = \frac{x - x_i}{\sqrt{(x - x_i)^2 + (y - y_i)^2}}$$

$$dy = \frac{y - y_i}{\sqrt{(x - x_i)^2 + (y - y_i)^2}}$$

从【0, 1】出发寻找图心，通过随机梯度下降法更新要寻找的点：

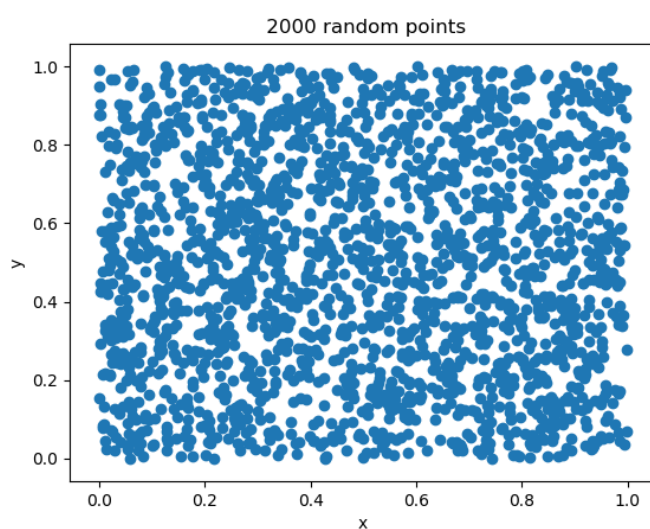
$$x = x - \alpha * dx$$

$$y = y - \alpha * dy$$

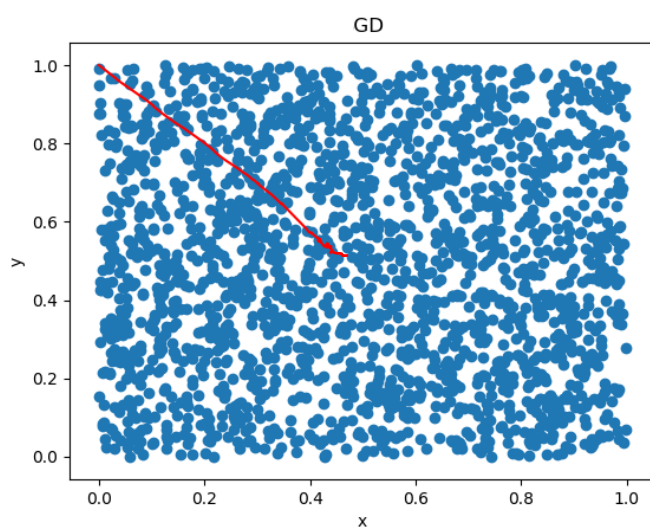
直到前后更新的点的损失函数之差  $f_n - f_{n-1}$  的绝对值小于阈值 0.00001。

### 三、实验结果

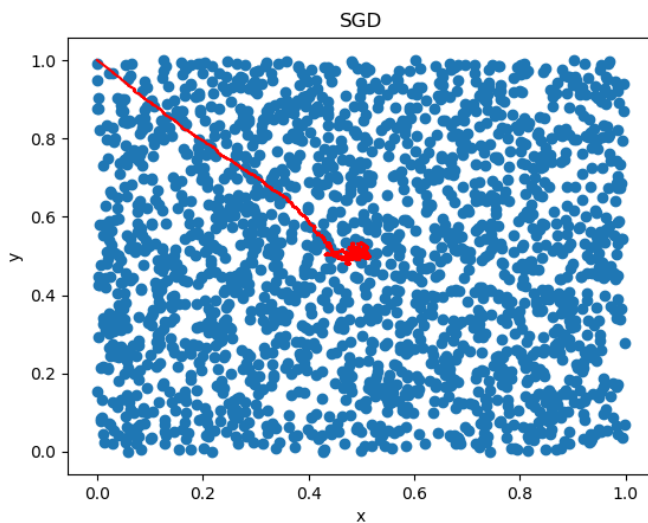
1、在二维单位正方形内随机生成 2000 个点，结果如下：



2、通过梯度下降法获得二维单位正方形的中心点（图心）：[0.50174, 514556]



3、通过随机梯度下降法获得二维单位正方形的中心点（图心）：[0.50033, 0.50879]



#### 四、实验结果分析：

1、梯度下降方法在迭代 1583 次后收敛到 **【0.50174， 514556】**，然而在每一次迭代过程中都用到所有 2000 个数据，每一次和总体迭代速度相对慢。

2、随机梯度下降方法在迭代 843543 次后收敛到 **[0.50033, 0.50879]**，在每一次迭代过程中所用到的数据为在 2000 个数据中随机取到的一个数据，总体迭代速度相较梯度下降法快。

3、梯度下降法在较少次迭代后基本可以迭代到最优值，然而随机梯度下降方法由于其随机性，在收敛到最优值附近会有一定程度的震荡。

#### 五、实验总结：

1、理论上二维单位正方形的几何中心为 **【0.5， 0.5】**。

2、可以利用随机生成的 2000 个样本通过梯度算法和随机梯度算法取得二维单位正方形的图心。

3、随机梯度下降算法从所有样本中随机取一个样本计算损失函数更新每次迭代所得的“图心“，直到前后两次迭代的损失函数小于阈值时停止迭代，输出图心。

4、在使用梯度下降法寻找最优值问题时，如果样本数据很大，每一次迭代都需要使用所有数据及逆行迭代，需要消耗较长的迭代时间。而在使用随机梯度下降方法寻找最优值问题时，其在每一次迭代过程中只是随机从所有数据中选择一组数据，消耗的迭代时间更少，但是其收敛到最优值附近时会有一定程度的震荡。所以在使用梯度下降或者随机梯度下降方法时，往往根据其样本数据量进行选择。