

```
hbhong@LAPTOP-109R9QT4:/mnt/c/Users/san26$ git clone https://github.com/big-data-europe/docker-spark.git
```

先去 clone spark

```
hbhong@LAPTOP-109R9QT4:/mnt/c/Users/san26/docker-spark$ docker-compose up -d
```

在做 compose 後進入 docker 的環境

寫 python spark

```
bash-5.0# python spark.py
```

```
w_grad = parsedData.map(lambda p: -np.dot((p[1] - sigmoid(np.dot(p[0], w) + b)), p[0])).reduce(lambda x,y: x+y)
```

利用 map 獨立計算每一個人的梯度差值後再用 reduce 加總起來

```
w -= w_grad * lr
```

得到總梯度後往負梯度方向以 learning rate 的速度更新

```
acc = parsedData.filter(lambda p: int(np.where(sigmoid(np.dot(p[0], w) + b) >= 0.5, 1, 0)) == int(p[1])).count() / float(parsedData.count())
```

最後丟進去看正確度多高

```
if(previous_acc>acc):  
    break
```

不斷的做直到收斂或者是 localmin

跑 sgd 去看 W 與 b 還有最後的正確率

```
(array([-0.57204091, -0.31336952, -0.27395643, -0.08597361]), 0.602104559996282)  
0.971574344023  
(array([-0.59902717, -0.32788066, -0.2943518 , -0.08805654]), 0.6612510471877265)  
0.97667638484
```

Git hub: <https://github.com/haobohong/CloudCompSecurity/tree/main/hw4>