

## Mini-batch gradient descent

→ Batch gradient descent: Use all  $m$  examples in each iteration

→ Stochastic gradient descent: Use 1 example in each iteration

Mini-batch gradient descent: Use  $b$  examples in each iteration

$b = \text{mini-batch size. } b = 10.$

Get  $b = 10$  examples  $(x^{(i)}, y^{(i)}), \dots, (x^{(i+9)}, y^{(i+9)})$

$$\theta_j := \theta_j - \alpha \frac{1}{10} \sum_{k=i}^{i+9} (h_{\theta}(x^{(k)}) - y^{(k)}) \cdot x_j^{(k)}$$

$j := i + 10$

相当于每次选取  $b$  个数据，而不是所有数据，进行加和，梯度下降处理

## Mini-batch gradient descent

Say  $b = 10, m = 1000.$

Repeat {

→ for  $i = 1, 11, 21, 31, \dots, 991$  {

$$\theta_j := \theta_j - \alpha \frac{1}{10} \sum_{k=i}^{i+9} (h_{\theta}(x^{(k)}) - y^{(k)}) x_j^{(k)}$$

(for every  $j = 0, \dots, n$ )

}

}

$m = 300, 600, 900$

↑

→  $b$  examples

→ 1 example

Vectorization

$b = 10$   
↑

优点：便于实现向量化，并行计算

缺点：多了一个参数  $b$  需要确认