

1. Introduction

一、监督学习 : 有正确答案 Supervised Learning

回归问题: 通过回归推出一个, 预测, 连续的有价值的输出 Regression problem

分类问题: classification problem

支持向量算法---无限维的向量

二、非监督学习 Unsupervised Learning : 只有一些数据, 没有给正确答案

Model and Cost Function

聚类算法: 举例: 新闻

鸡尾酒宴会问题 : 音频分类

2. Linear Regression with One Variable 线性回归模型

训练样本: Training set of housing prices

m : 训练样本的数量

x : 输入变量/特征量

y : 输出变量/目标变量

training set----> learning algorithm -----> h(hypothesis 假设)表示一个函数

代价函数 cost function 平方误差代价函数

Hypothesis:

$$h_{\theta}(x) = \theta_0 + \theta_1 x$$

Parameters:

$$\theta_0, \theta_1$$

Cost Function:

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

Goal: minimize $J(\theta_0, \theta_1)$

Contour plot 轮廓图

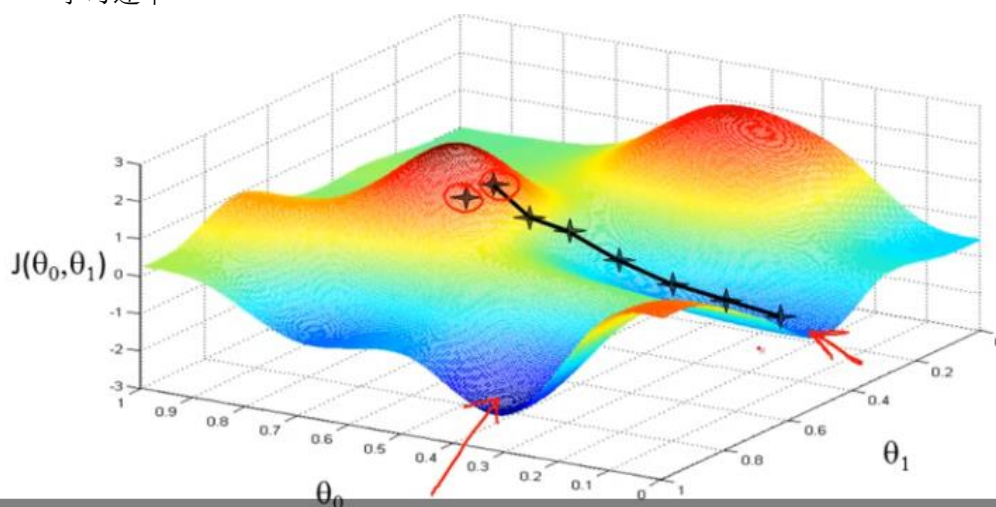
Parameter Learning

梯度下降法： 非常不同的局部最优解 gradient descent

Gradient descent algorithm

repeat until convergence {
 $\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$ (simultaneously update
 $j = 0$ and $j = 1$)
}

α --学习速率



非常不同的局部最优解 这就是梯度下降算法的一个特点

$:=$ 赋值运算

用于线性回归的代价函数 总是 凸函数

凸函数 `convex function`: 没有局部最优解, 只有一个全局最优解

批量梯度下降: `Batch Gradient Descent` 下降时每一步, 都用到了所有的训练样本

Gradient descent algorithm

```
repeat until convergence {  
     $\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})$   
     $\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) \cdot x^{(i)}$   
}
```

不需要迭代的求解, 和线性代数有关的另一种方法: 正规方程 `normal equations`