CS229-note1-学习心得

第一部分

线性回归

简介：

1.背景：地区 ***房间大小和价格*** **数据，推进如何执行监督学习，以及确定参数** ***thetaθ*** **。**

将 **y** 假设为关于 **x** 的线性函数，**θ** 代表拟合参数

于是：

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

为了简单起见，假定 **X0 = 1**

于是：

$$h(x) = \sum_{i=0}^{n} \theta_i x_i = \theta^T x,$$

如何确定 ***θ*** ？

于是：

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

注：1/2参考方差

2.LMS算法-最小均方，也称为 Widrow-Hoff 学习规则，也即最小梯度下降

为了使J(θ) 更小，直到我们希望收敛到使 J(θ) **最小化**的 θ 值

于是：

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta).$$

注：α 称为学习率。这是一种非常自然的算法，它会在 J 的最陡下降方向上重复迈出一步

公式2-1

假设我们有一个训练实例（x,y）,我们需要计算出右侧的偏导数

于是：

$$\begin{aligned}
\frac{\partial}{\partial \theta_j} J(\theta) &= \frac{\partial}{\partial \theta_j} \frac{1}{2} \left(h_\theta(x) - y\right)^2 \\
&= 2 \cdot \frac{1}{2} \left(h_\theta(x) - y\right) \cdot \frac{\partial}{\partial \theta_j} \left(h_\theta(x) - y\right) \\
&= \left(h_\theta(x) - y\right) \cdot \frac{\partial}{\partial \theta_j} \left(\sum_{i=0}^{n} \theta_i x_i - y\right) \\
&= \left(h_\theta(x) - y\right) x_j
\end{aligned}$$

公式 2-2

接着，结合**公式2-1和2公式-2**

于是：

$$\theta_j := \theta_j + \alpha \left(y^{(i)} - h_\theta(x^{(i)})\right) x_j^{(i)}.$$

注：LMS更新规则（LMS代表"最小均方"），也称为 Widrow-Hoff 学习规则。

如果在拟合过程中，我们的预测值和实际值有较大出入，以下是两种解决方法：

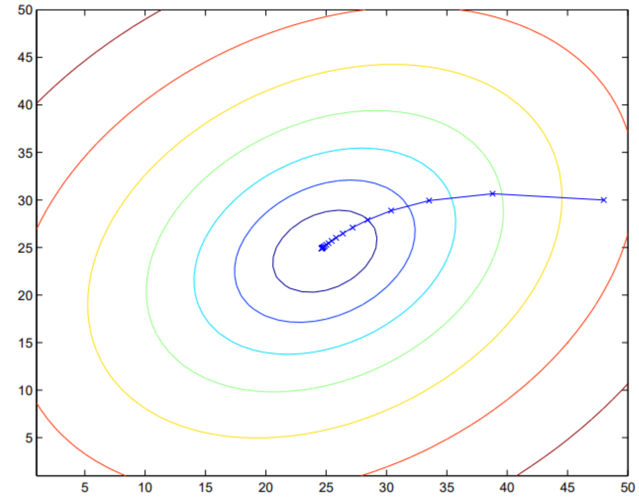
***First*.对于包含多个训练集，替换为以下算法，重复直到收敛：**

$$\begin{array}{l}
\text{Repeat until convergence } \{ \\
\qquad \theta_j := \theta_j + \alpha \sum_{i=1}^{m} \left(y^{(i)} - h_\theta(x^{(i)})\right) x_j^{(i)} \qquad (\text{for every } j). \\
\}
\end{array}$$

注：求和部分为 *∂J(θ)/∂θj (for the original definition of J).*

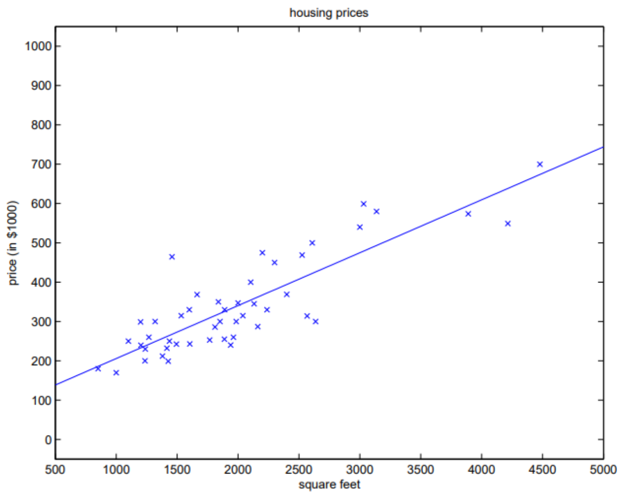
- 此方法被称为**批梯度下降**，每一步都扫描整个**训练集**的每个实例
- 此方法受**局部极小值**影响大，但是**线性回归问题**只有一个全局最优值，因此此梯度下降总是**收敛**
- 假设 **学习率 α** 不太大

下面是对比两种方法的直观图：



The ellipses shown above are the contours of a quadratic function. Also shown is the trajectory taken by gradient descent, which was initialized at (48,30). The x's in the figure (joined by straight lines) mark the successive values of θ that gradient descent went through.

梯度下降实例



If the number of bedrooms were included as one of the input features as well, we get $\theta_0 = 89.60, \theta_1 = 0.1392, \theta_2 = -8.738$.

批梯度下降实例

考虑以下算法：

$$
\begin{aligned}
&\text{Loop } \{ \\
&\qquad \text{for i=1 to m, } \{ \\
&\qquad\qquad \theta_j := \theta_j + \alpha \left(y^{(i)} - h_\theta(x^{(i)})\right) x_j^{(i)} \qquad (\text{for every } j). \\
&\qquad \} \\
&\}
\end{aligned}
$$

- 根据**单个训练样例**相关误差梯度更新参数
- **随机梯度下降**
- 不用每次都扫描**整个训练集**
- 更快 ***接近* θ最小值**，可能永远无法到达最小值，在**最小值附近振荡**

3.正规方程

3.1矩阵导数

3.2最小二乘再访

4.概率解释

5.局部加权线性回归

第二部分

分类和逻辑回归