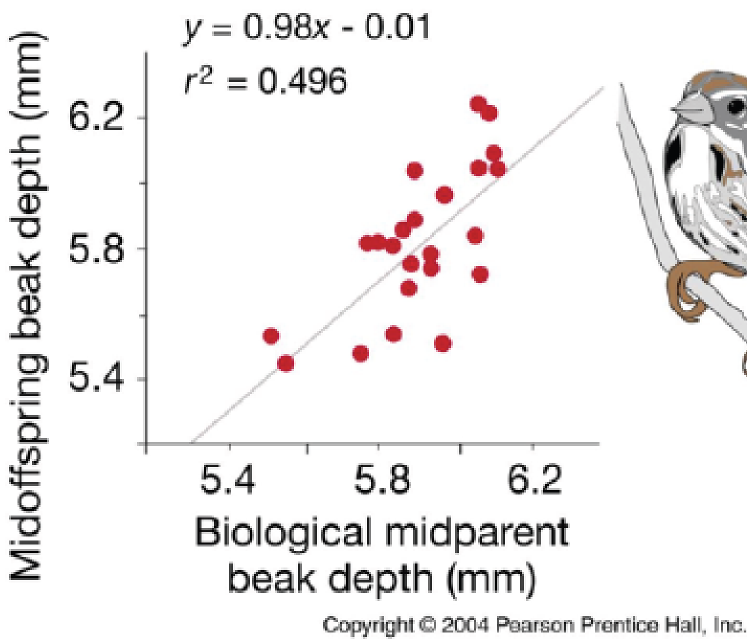
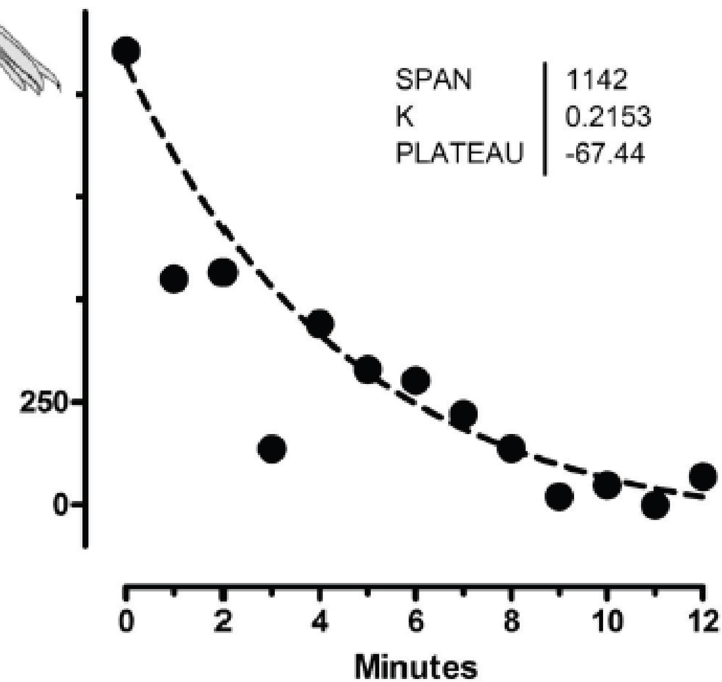
第二讲
线性回归

夏 睿
计算机科学与工程学院
南京理工大学
http://www.nustm.cn/member/rxia

回归



nonlinear



linear

数据, 输入, 输出, 关系

- 训练集

Living area (feet^2)	#bedrooms	Price (1000\$s)
2104	3	400
1600	3	330
2400	3	369
1416	2	232
3000	4	540
⋮	⋮	⋮

一个训练样例 $(x^{(i)}, y^{(i)})$, 其中i表示样例的索引

输入: 特征向量 $x = [x_1, x_2]$

输出: y

- 假设：线性模型

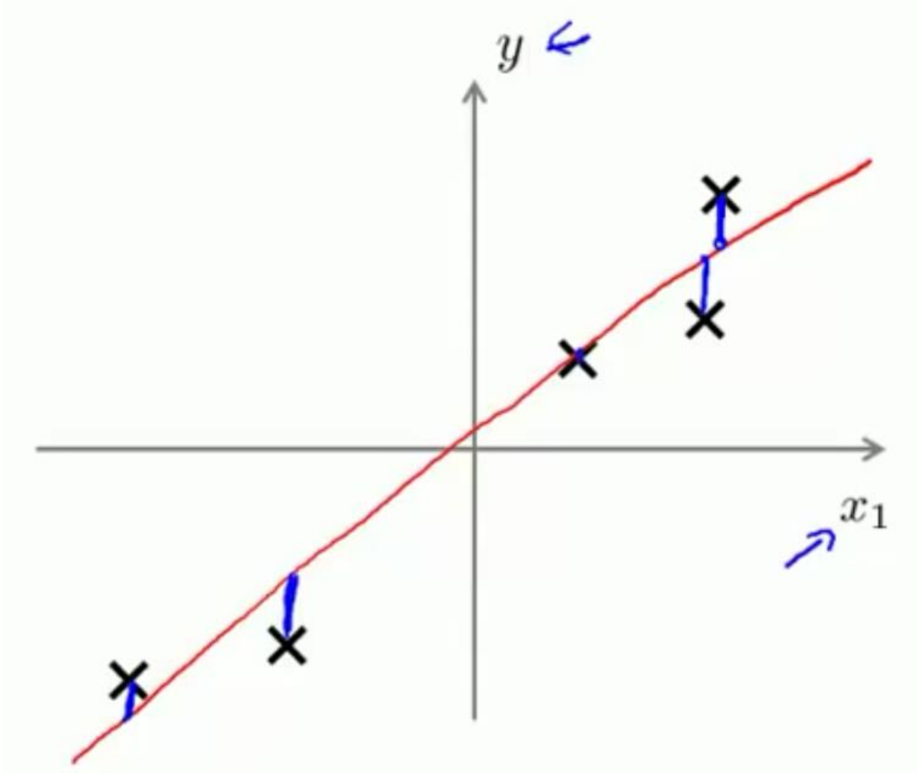
$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

最小均方差（LMS）算法

- 假设

$$h_\theta(x) = \sum_{i=1}^{n} \theta_i x_i = \theta^{\mathrm{T}} x$$

- 参数 θ



- 代价函数

$$J_l(\theta) = \frac{1}{2}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$
$$= \frac{1}{2}\sum_{i=1}^{m}(\theta^{\mathrm{T}} x^{(i)} - y^{(i)})^2$$

- 目标

$$\theta^* = \arg_\theta \min J_l(\theta)$$

LMS问题求解

- 定义

$$X=\begin{bmatrix}-\left(x^{(1)}\right)^{\mathrm{T}}-\\-\left(x^{(2)}\right)^{\mathrm{T}}-\\\vdots\\-\left(x^{(n)}\right)^{\mathrm{T}}-\end{bmatrix}\qquad y=\begin{bmatrix}y^{(1)}\\y^{(2)}\\\vdots\\y^{(n)}\end{bmatrix}$$

- 然后有，

$$X\theta-y=\begin{bmatrix}\left(x^{(1)}\right)^{\mathrm{T}}\theta\\\vdots\\\left(x^{(n)}\right)^{\mathrm{T}}\theta\end{bmatrix}-\begin{bmatrix}y^{(1)}\\\vdots\\y^{(n)}\end{bmatrix}=\begin{bmatrix}h_\theta(x^{(1)})-y^{(1)}\\\vdots\\h_\theta(x^{(n)})-y^{(n)}\end{bmatrix}$$

- LMS 的代价函数

$$J(\theta)=\frac{1}{2}\sum_{i=1}^{n}(h_\theta\left(x^{(i)}\right)-y^{(i)})^2=\frac{1}{2}(X\theta-y)^{\mathrm{T}}(X\theta-y)$$

LMS的闭式解

- 通过矩阵微分计算LMS梯度

$$
\begin{aligned}
\nabla_\theta J(\theta) &= \nabla_\theta \frac{1}{2}(X\theta - y)^{\mathrm{T}}(X\theta - y) \\
&= \frac{1}{2}\nabla_\theta\left(\theta^{\mathrm{T}}X^{\mathrm{T}}X\theta - \theta^{\mathrm{T}}X^{\mathrm{T}}y - y^{\mathrm{T}}X\theta + y^{\mathrm{T}}y\right) \\
&= \frac{1}{2}\nabla_\theta \mathrm{tr}\left(\theta^{\mathrm{T}}X^{\mathrm{T}}X\theta - \theta^{\mathrm{T}}X^{\mathrm{T}}y - y^{\mathrm{T}}X\theta + y^{\mathrm{T}}y\right) \\
&= \frac{1}{2}\nabla_\theta\left(\mathrm{tr}\theta^{\mathrm{T}}X^{\mathrm{T}}X\theta - 2\mathrm{tr}y^{\mathrm{T}}X\theta\right) = X^{\mathrm{T}}X\theta - X^{\mathrm{T}}y
\end{aligned}
$$

- 通过使梯度等于零获得闭式解

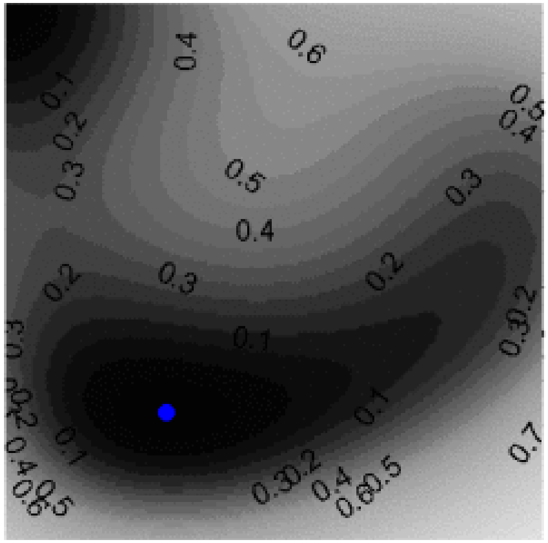
$$\theta^* = \boxed{(X^{\mathrm{T}}X)^{-1}}X^{\mathrm{T}}y$$

有时很难计算

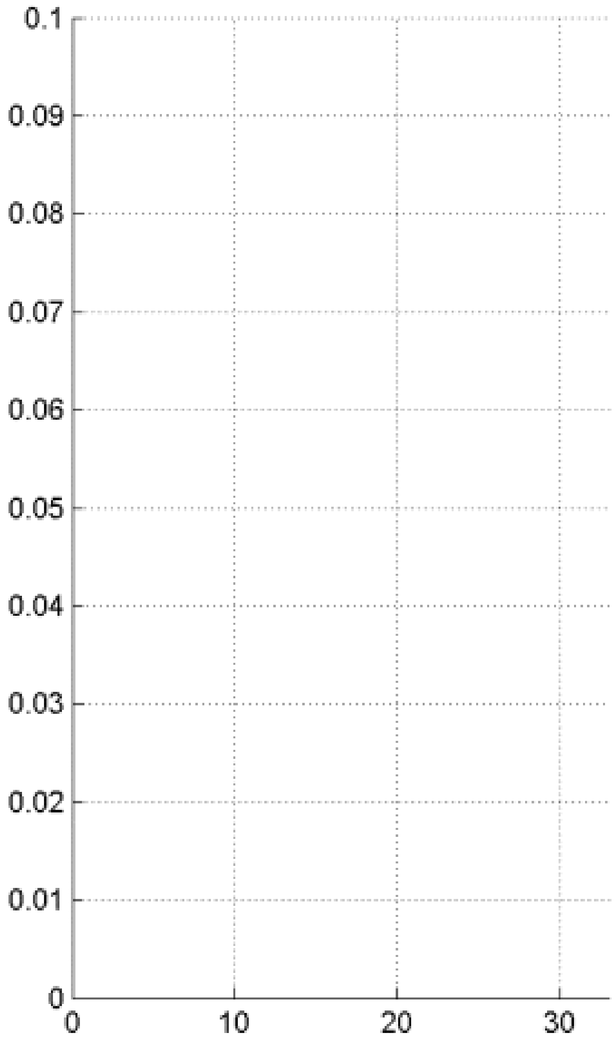
数值优化的梯度下降法

- 梯度下降是求函数$f(\theta)$最小值的一阶迭代优化算法
- 主要思想:
  - 梯度下降方向是函数值增加最快的方向.
- 优化过程:
  - 从初始位置开始(i.e., 初始参数$\theta^{(0)}$)
  - 在当前位置$\theta^{(t)}$, 重复直到收敛
    - 计算当前位置梯度: $\nabla_\theta f(\theta)|_{\theta=\theta^{(t)}}$
    - 沿梯度反方向移动到下一个位置t: $\theta^{(t+1)} = \theta^{(t)} - \alpha \cdot \nabla_\theta f(\theta)|_{\theta=\theta^{(t)}}$, α 是学习率
    - $t = t + 1$

梯度下降算法示意图



0.4
0.6
0.1
0.2
0.3
0.5
0.4
0.5
0.4
0.3
0.2
0.3
0.2
0.1
0.1
0.7



0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
10
20
30

线性回归的梯度下降

- 梯度

$$\frac{\partial J_l(\theta)}{\partial\theta} = \frac{1}{2}\frac{\partial}{\partial\theta}\sum_{i=1}^{n}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

$$= \frac{1}{2}\cdot 2\sum_{i=1}^{n}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)\cdot\frac{\partial}{\partial\theta}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)$$

$$= \sum_{i=1}^{n}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)\frac{\partial}{\partial\theta}\left(\theta^T x^{(i)}\right)$$

$$= \sum_{i=1}^{n}\boxed{\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)x^{(i)}}$$

“误差· 输入”

- 梯度下降(GD) 优化

$$\theta := \theta - \alpha\frac{\partial}{\partial\theta}J_l(\theta) = \theta - \alpha\sum_{i=1}^{n}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)x^{(i)}$$

实践：南京房价预测

- 给出历史数据

  Year x = [2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013]

  Price y = [2.000, 2.500, 2.900, 3.147, 4.515, 4.903, 5.365, 5.704, 6.853, 7.971, 8.561, 10.000, 11.280, 12.900]
- 假设: 价格和年度呈线性关系，因此可以通过线性回归建模
- 任务
  - 通过线性回归得到x和y的关系，基于 1)闭式解 和 2)梯度下降;
  - 预测2014年南京的房价.

欢迎提问！