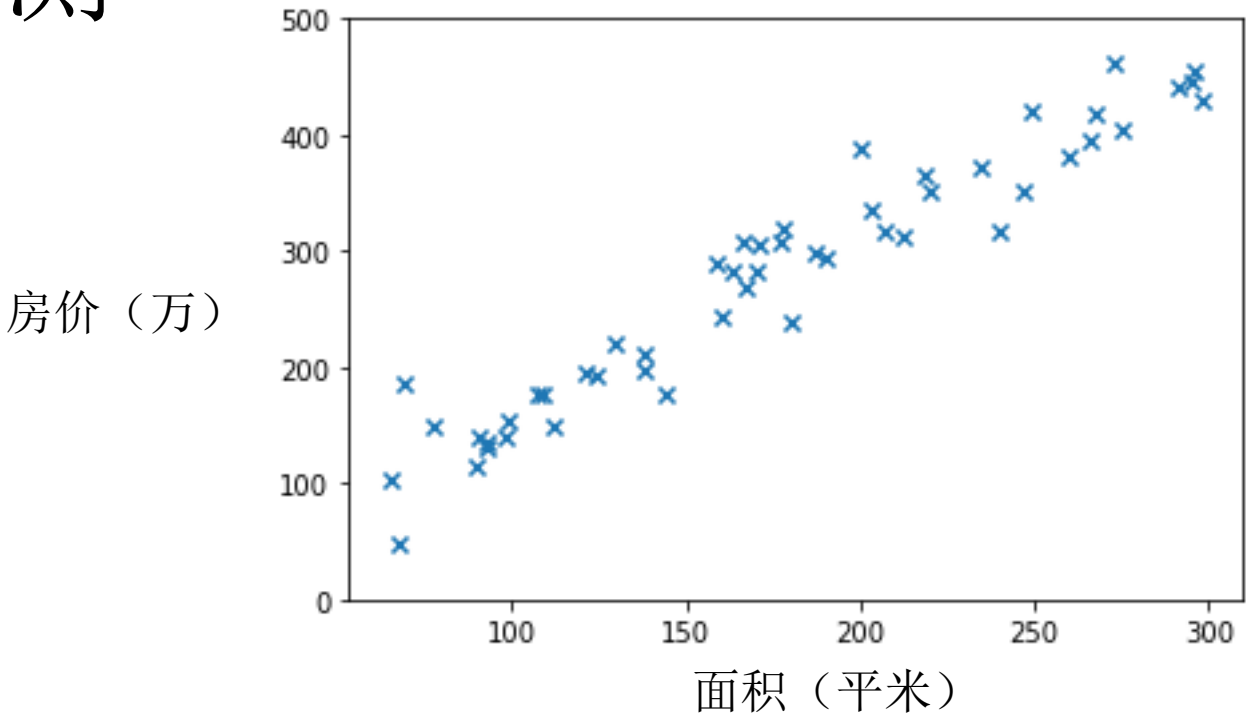
回归算法及其应用

2019年7月20日

房价预测



监督式学习　　回归问题

单变量线性回归

- 房价数据

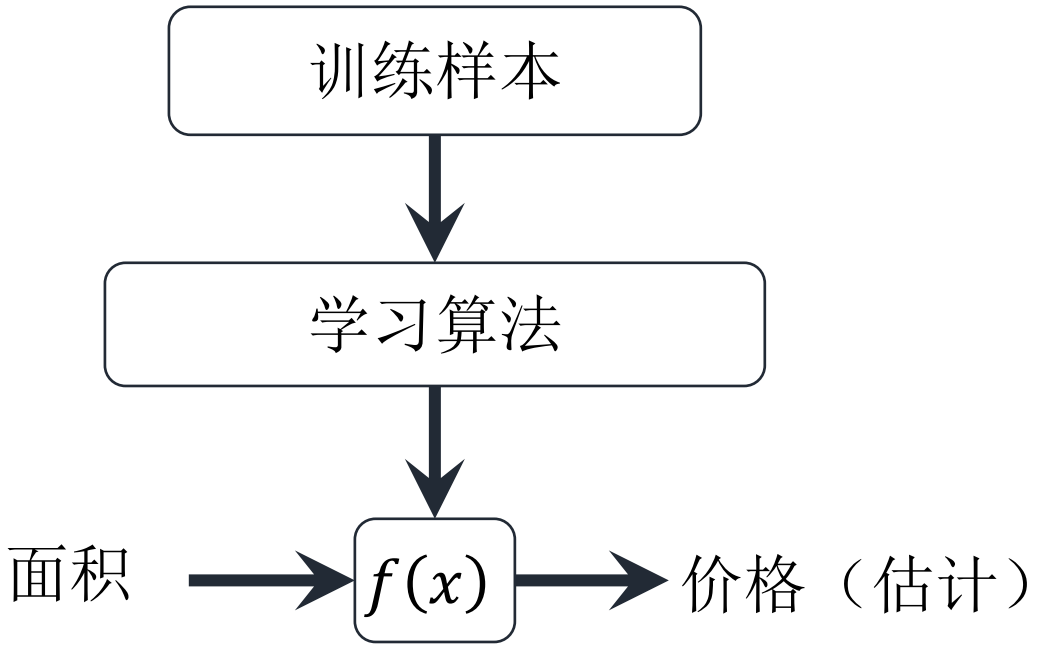
面积（x）	房价（y）
286.126	428.054
203.052	314.231
294.9	454.869
65.7734	102.841
…	…

- 训练样本

$$(x, y)$$

$$\left(x^{(i)}, y^{(i)}\right)$$

单变量线性回归



如何表示$f(x)$

$$f(x) = wx + b$$

单变量线性回归

- 训练样本

面积（x）	房价（y）
286.126	428.054
203.052	314.231
294.9	454.869
65.7734	102.841
…	…

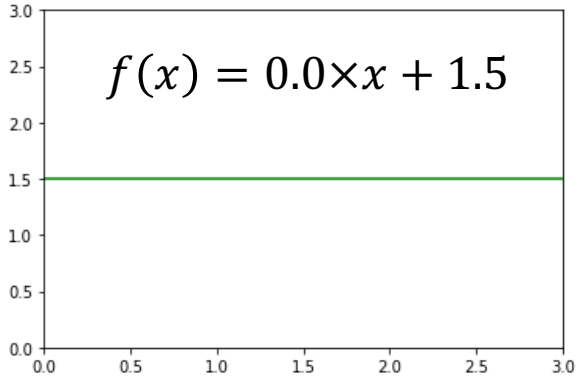
- $f(x) = wx + b$

参数：w, b

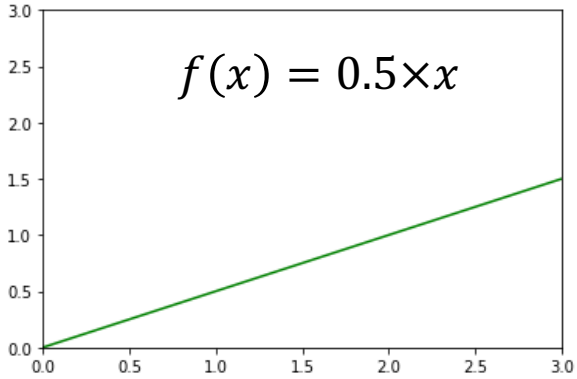
如何确定？

单变量线性回归

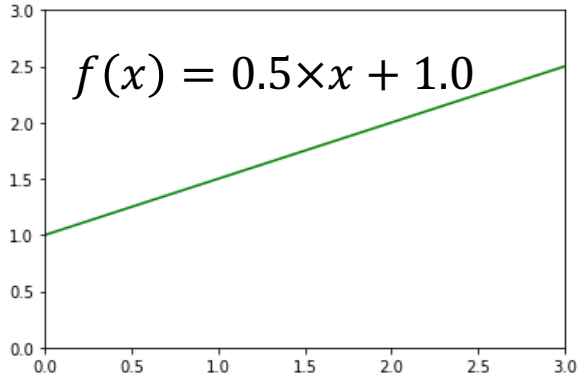
- $f(x) = wx + b$



$w = 0.0$
$b = 1.5$

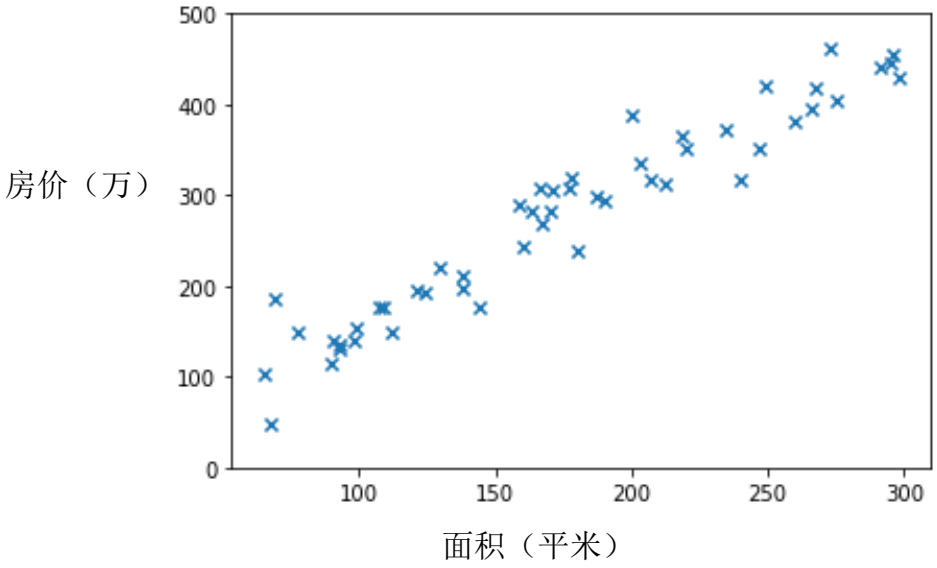


$w = 0.5$
$b = 0.0$



$w = 0.5$
$b = 1.0$

单变量线性回归



准则：挑选w, b使得对训练样本(x, y)的预测$f(x)$和y越接近越好

$$\min_{w,b} \frac{1}{m} \sum_{i=1}^{m} (f(x_i) - y_i)^2$$

目标公式

$$J(w, b) = \frac{1}{m} \sum_{i=1}^{m} (f(x_i) - y_i)^2$$

代价函数

单变量线性回归

问题简化

- 映射

$$f(x) = wx + b$$

$$f(x) = wx$$

- 参数

$$w, b$$

$$w$$

- 代价函数

$$J(w,b) = \frac{1}{m}\sum_{i=1}^{m}(f(x_i) - y_i)^2$$

$$J(w) = \frac{1}{m}\sum_{i=1}^{m}(f(x_i) - y_i)^2$$

- 目标

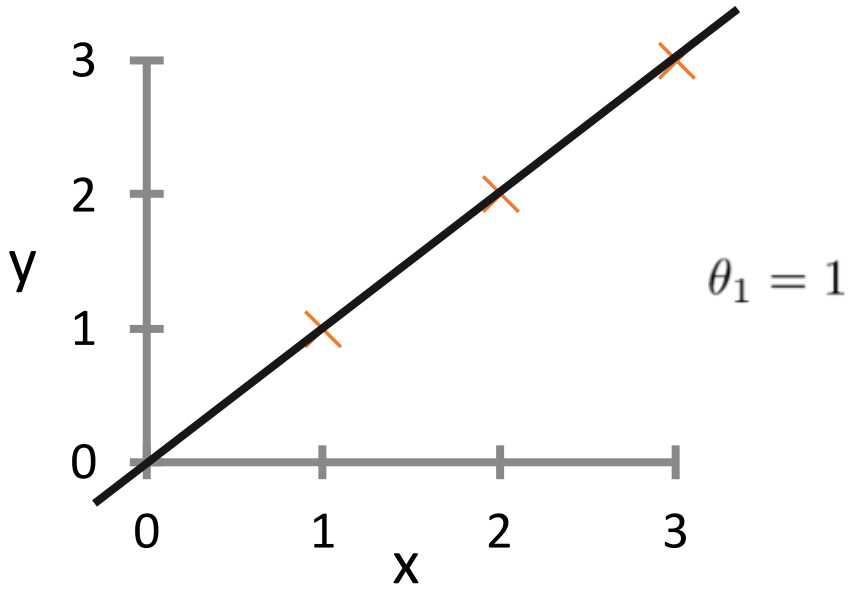
$$\min_{w,b} \frac{1}{m}\sum_{i=1}^{m}(f(x_i) - y_i)^2$$

$$\min_{w} \frac{1}{m}\sum_{i=1}^{m}(f(x_i) - y_i)^2$$

单变量线性回归

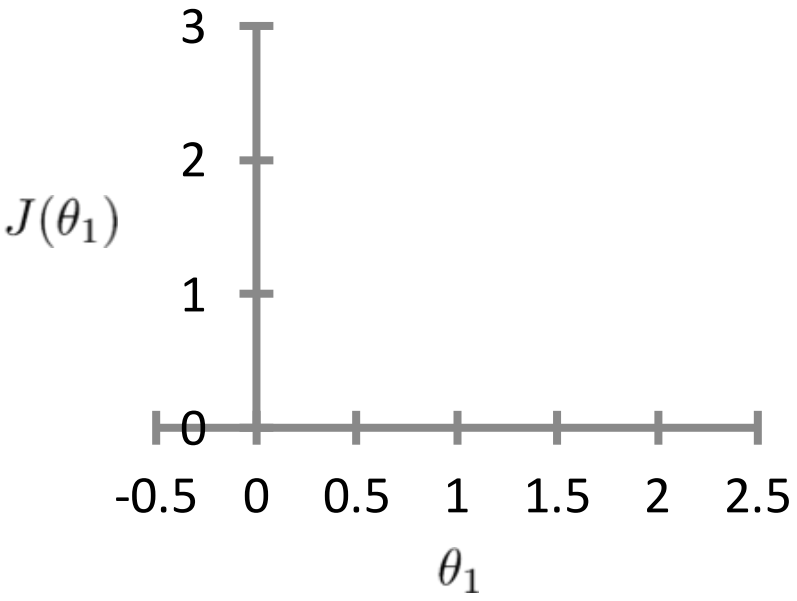
$f(x)$

(固定$w=1, f(x)$是一个关于x的函数)



$J(w)$

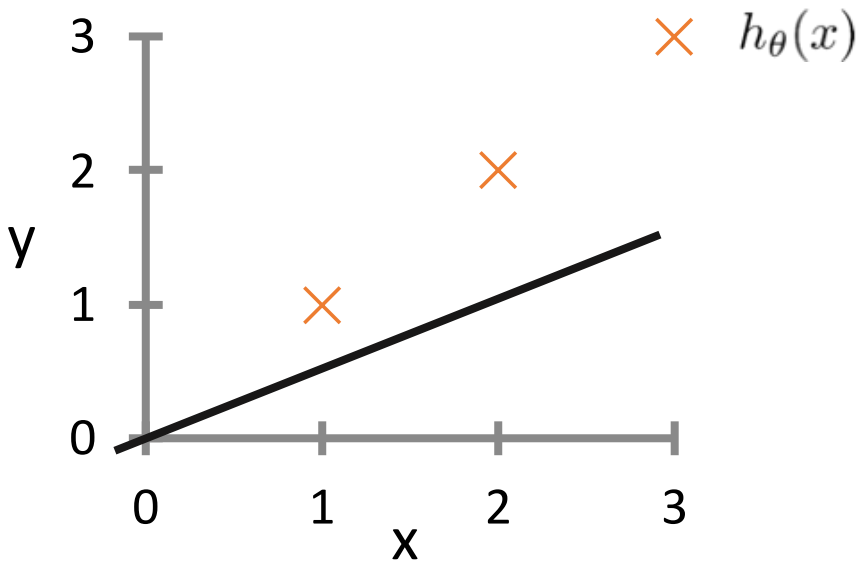
(固定$x, J(w)$是一个关于w的函数)



单变量线性回归

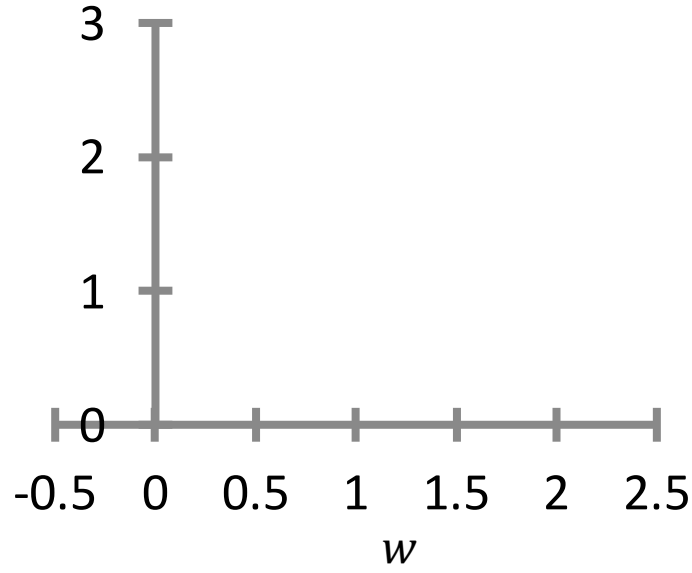
$f(x)$

(固定$w = 0.5$, $f(x)$是一个关于x的函数)



$J(w)$

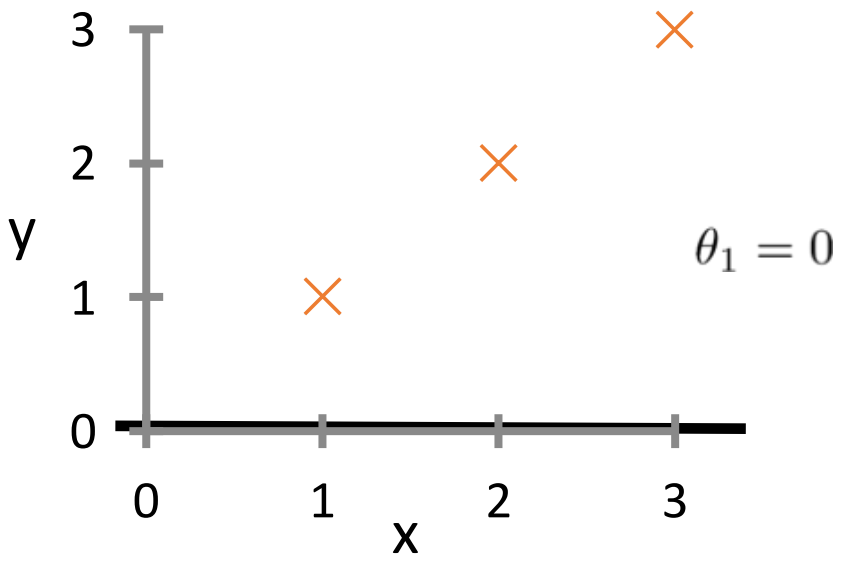
(固定x, $J(w)$是一个关于w的函数)



单变量线性回归

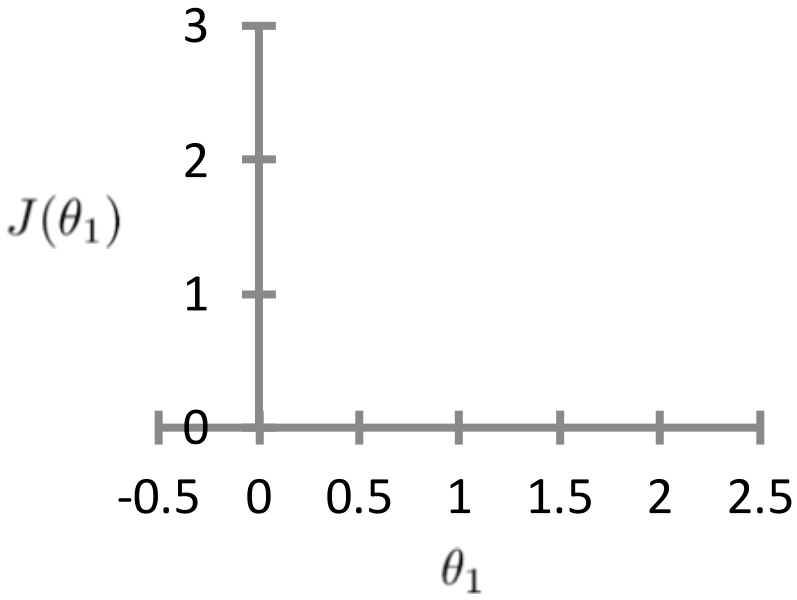
$f(x)$

(固定$w=0, f(x)$是一个关于x的函数)



$J(w)$

(固定$x, J(w)$是一个关于w的函数)



单变量线性回归

- 映射

$$f(x) = wx + b$$

- 参数

$$w, b$$

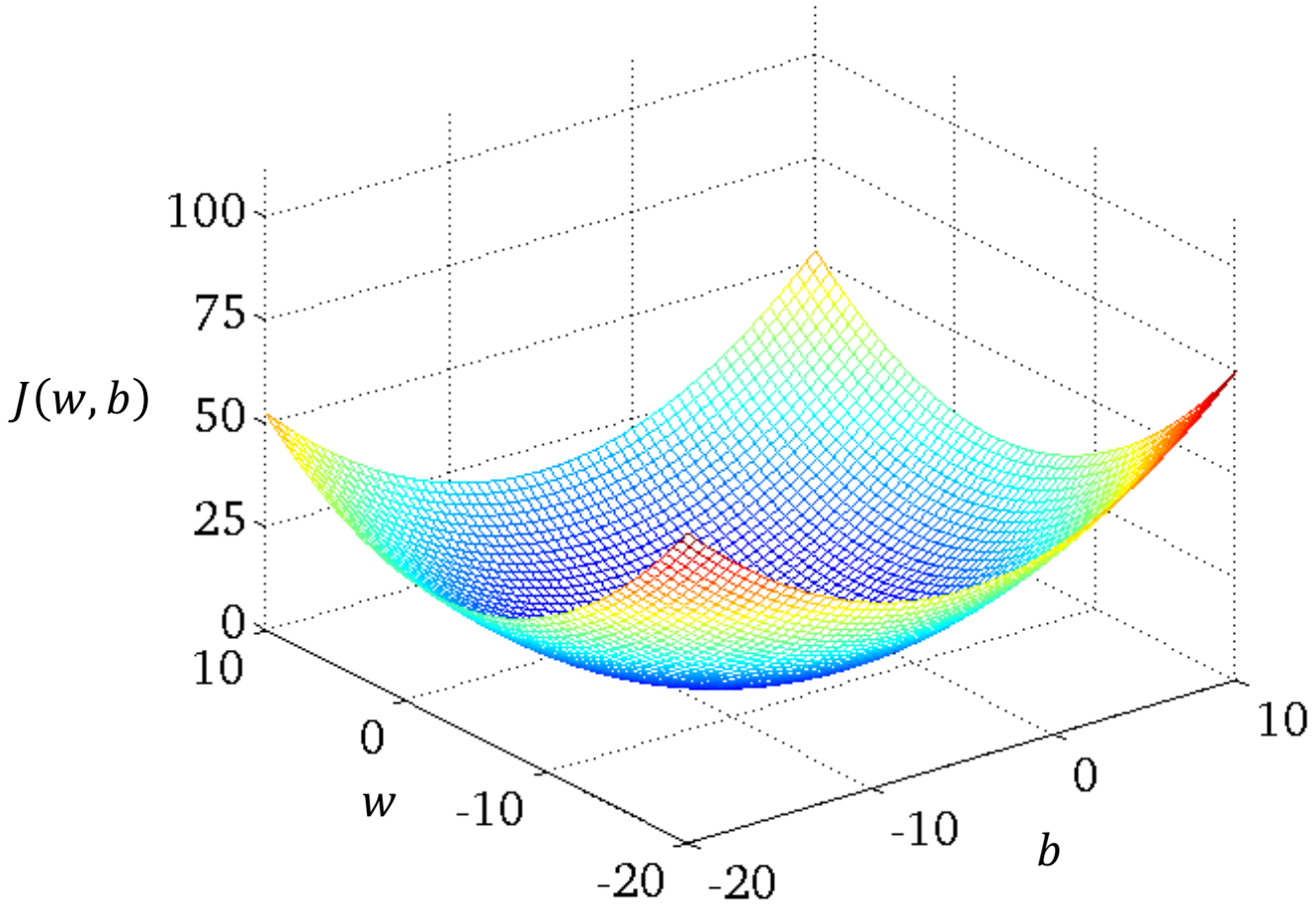
- 代价函数

$$J(w,b) = \frac{1}{m}\sum_{i=1}^{m}(f(x_i) - y_i)^2$$

- 目标

$$\min_{w,b} \frac{1}{m}\sum_{i=1}^{m}(f(x_i) - y_i)^2$$

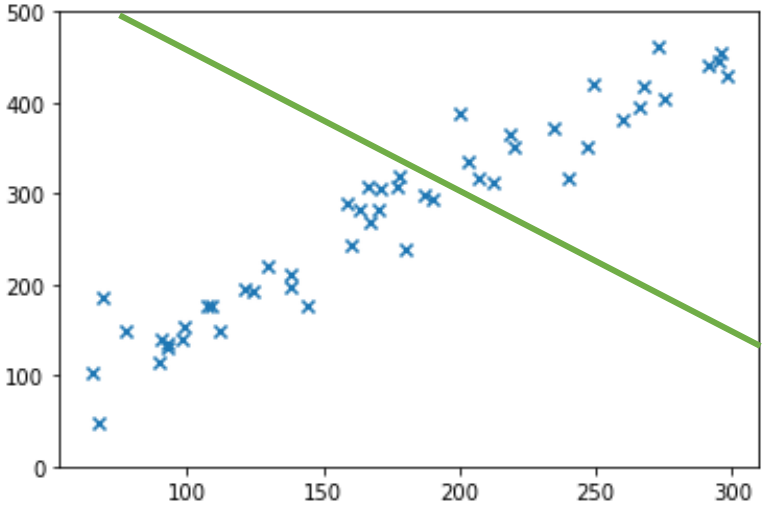
单变量线性回归



单变量线性回归

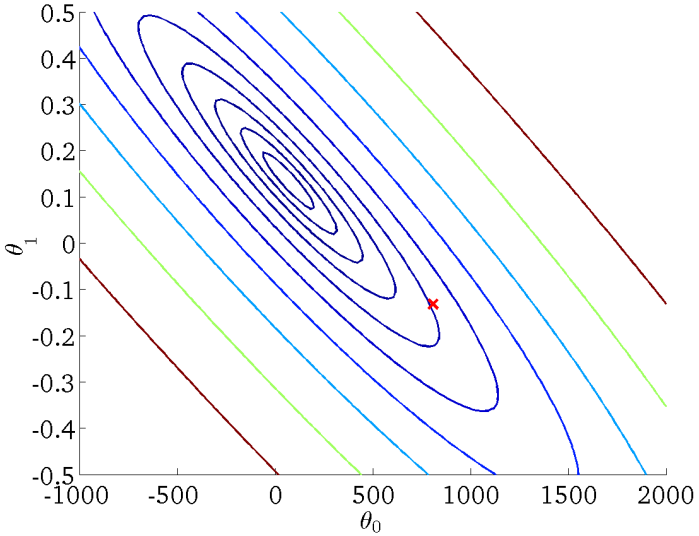
$f(x)$

(固定w和b, $f(x)$是一个关于x的函数)



$J(w,b)$

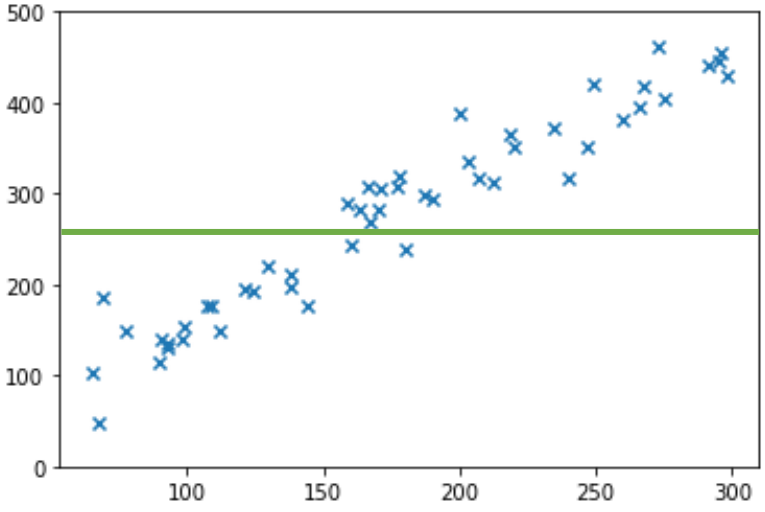
(固定x, $J(w,b)$是一个关于w,b的函数)



单变量线性回归

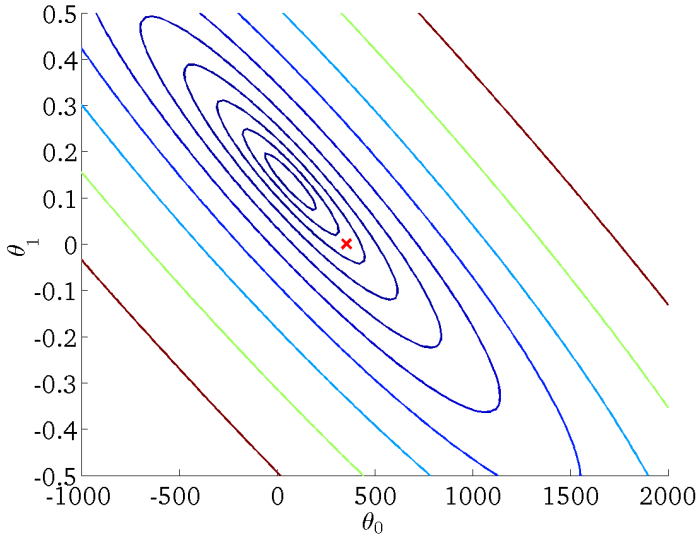
$f(x)$

(固定w和b, $f(x)$是一个关于x的函数)



$J(w,b)$

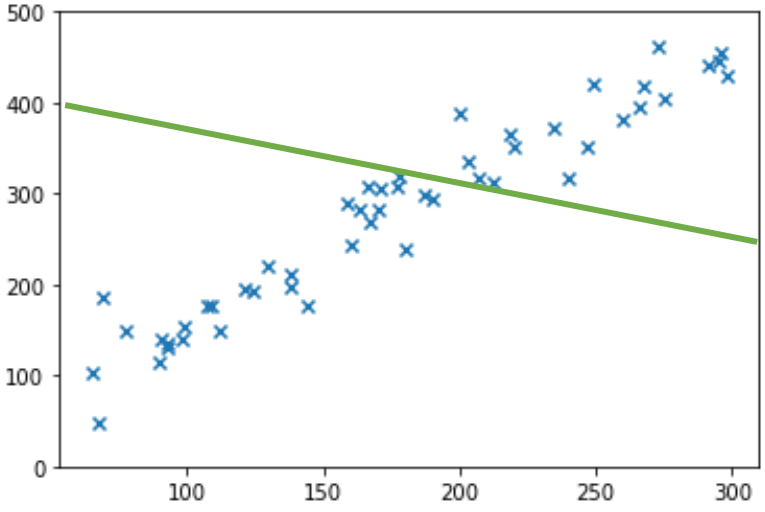
(固定x, $J(w,b)$是一个关于w,b的函数)



单变量线性回归

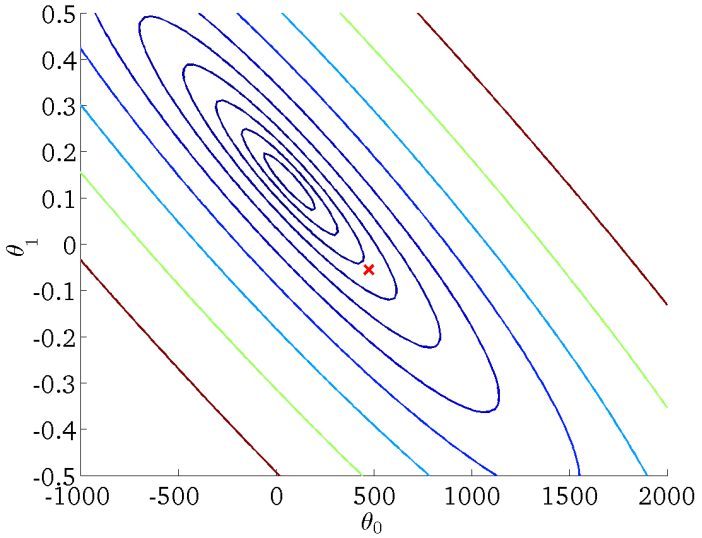
$f(x)$

(固定w和b, $f(x)$是一个关于x的函数)



$J(w,b)$

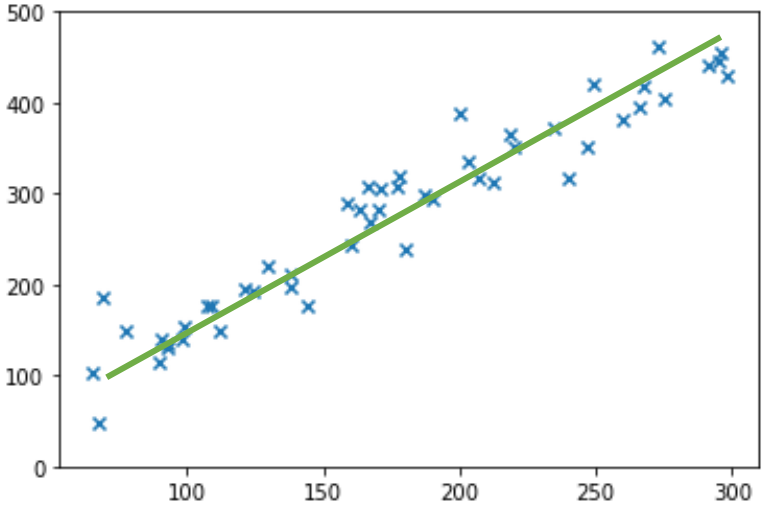
(固定x, $J(w,b)$是一个关于w,b的函数)



单变量线性回归

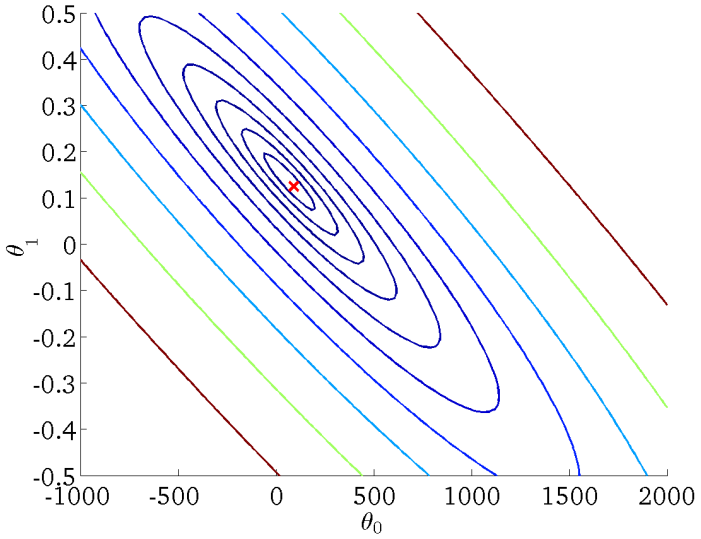
$f(x)$

(固定w和b, $f(x)$是一个关于x的函数)



$J(w,b)$

(固定x, $J(w,b)$是一个关于w,b的函数)



单变量线性回归

- 假设已有代价损失函数

$$J(w,b)=\frac{1}{m}\sum_{i=1}^{m}(f(x_i)-y_i)^2$$

- 目标

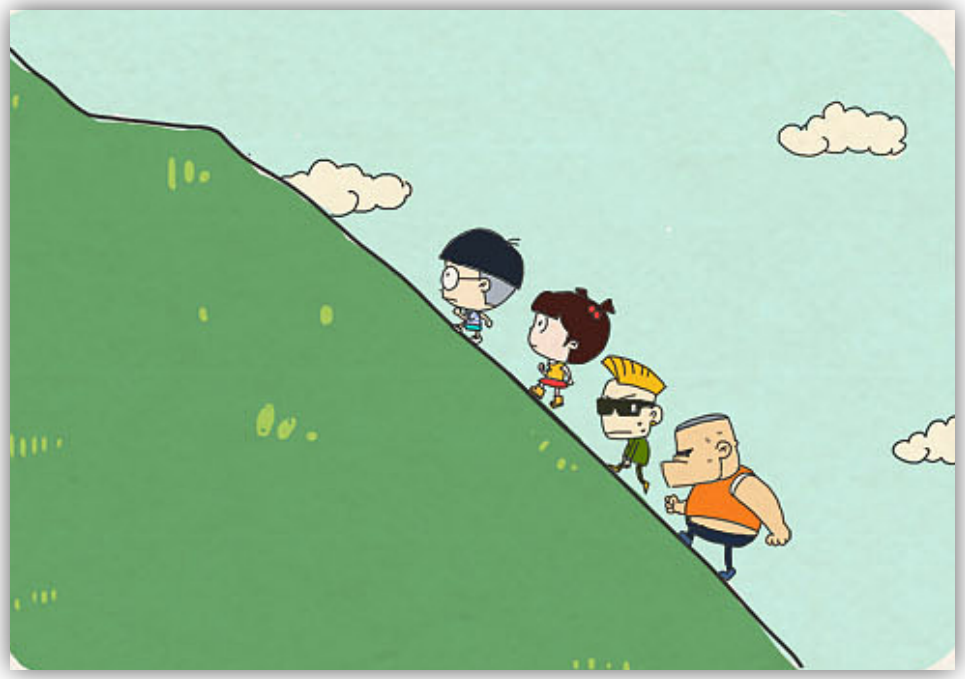
$$\min_{w,b}\frac{1}{m}\sum_{i=1}^{m}(f(x_i)-y_i)^2$$

w,b

- 求解框架
  - 变量w,b随机赋值
  - 变化w,b以减少$J(w,b)$的值，直到寻找到最优参数

单变量线性回归

- 梯度下降算法



单变量线性回归

- 梯度方向

$$\frac{\partial J(w,b)}{\partial w} = \frac{2}{m}\sum_{i=1}^{m}\left(wx^{(i)} + b - y^{(i)}\right)x^{(i)}$$

$$\frac{\partial J(w,b)}{\partial b} = \frac{2}{m}\sum_{i=1}^{m}\left(wx^{(i)} + b - y^{(i)}\right)$$

- 更新变量w, b

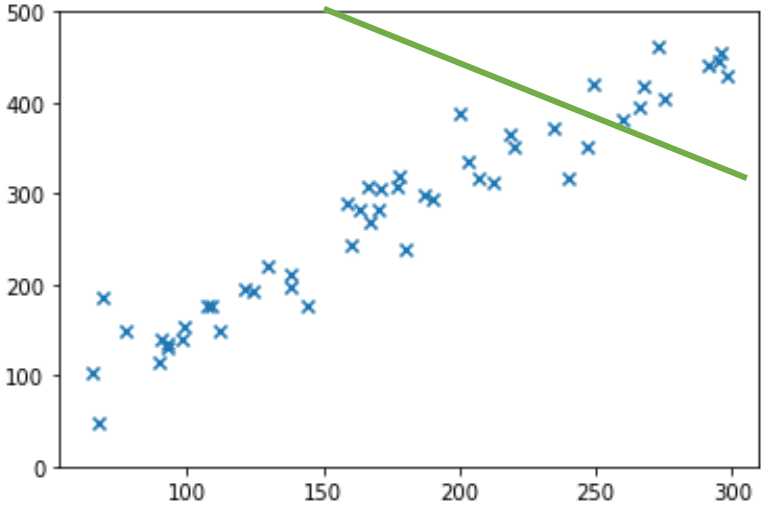
$$w = w - \alpha\frac{\partial J(w,b)}{\partial w}$$

$$b = b - \alpha\frac{\partial J(w,b)}{\partial b}$$

单变量线性回归

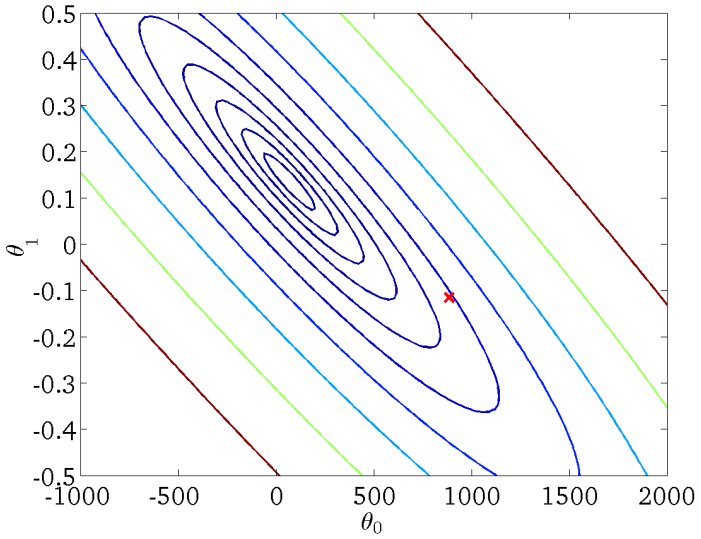
$f(x)$

(固定w和b, $f(x)$是一个关于x的函数)



$J(w,b)$

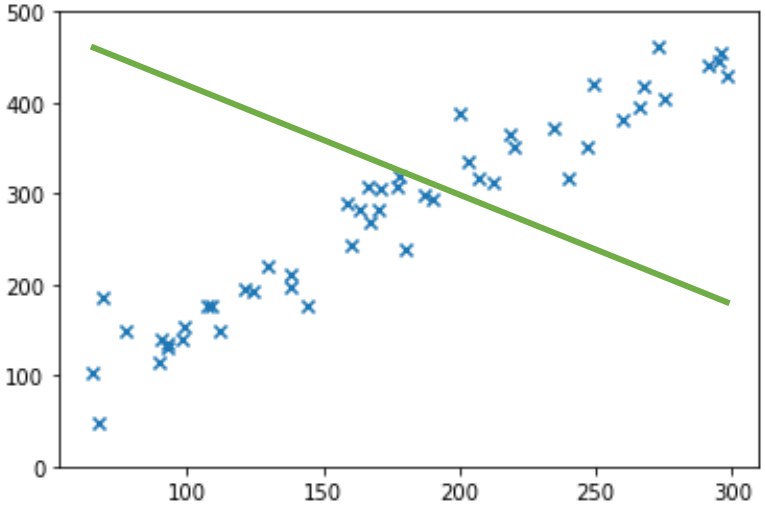
(固定x, $J(w,b)$是一个关于w,b的函数)



单变量线性回归

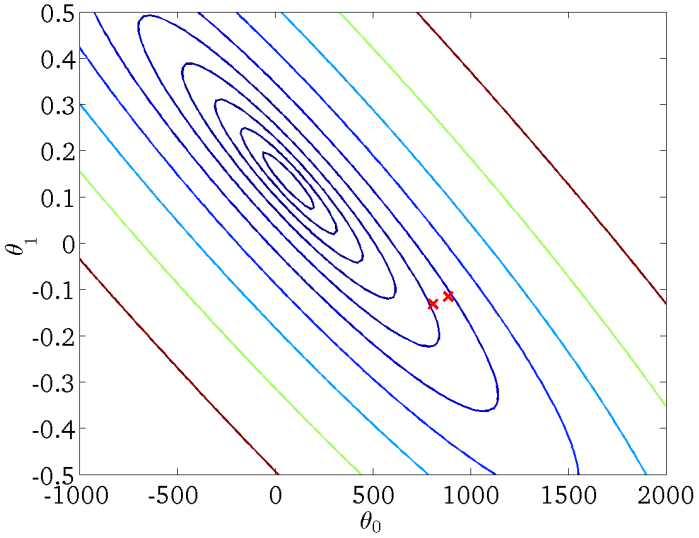
$f(x)$

(固定w和b, $f(x)$是一个关于x的函数)



$J(w,b)$

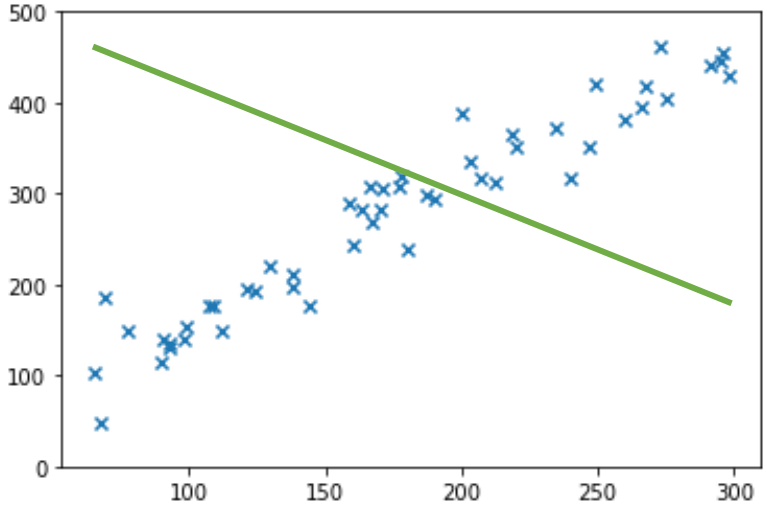
(固定x, $J(w,b)$是一个关于w,b的函数)



单变量线性回归

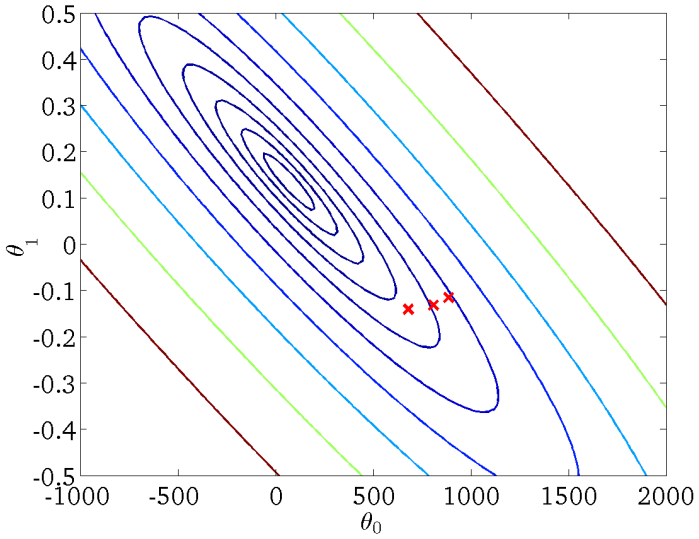
$f(x)$

(固定w和b, $f(x)$是一个关于x的函数)



$J(w,b)$

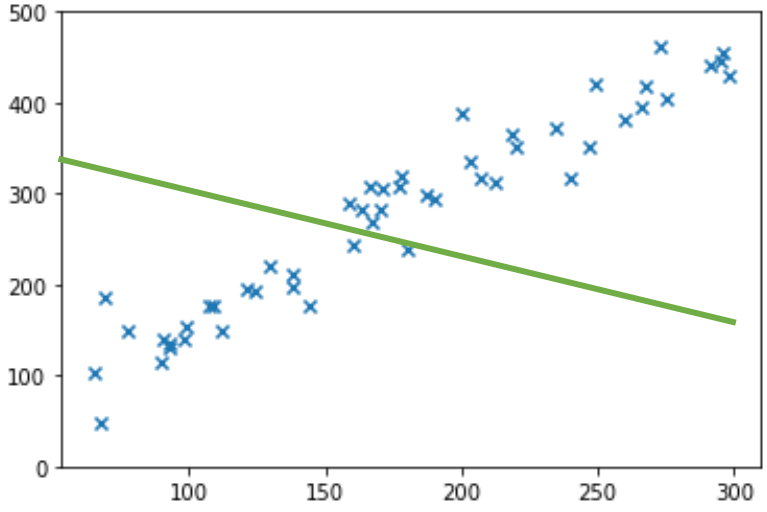
(固定x, $J(w,b)$是一个关于w,b的函数)



单变量线性回归

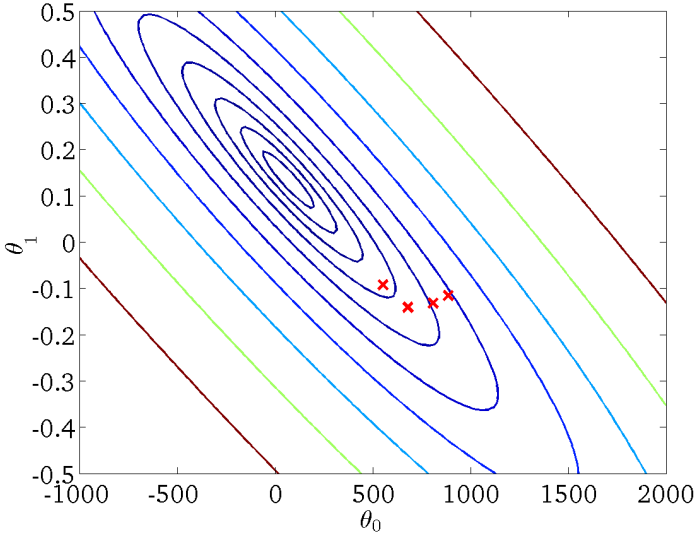
$f(x)$

(固定w和b, $f(x)$是一个关于x的函数)



$J(w,b)$

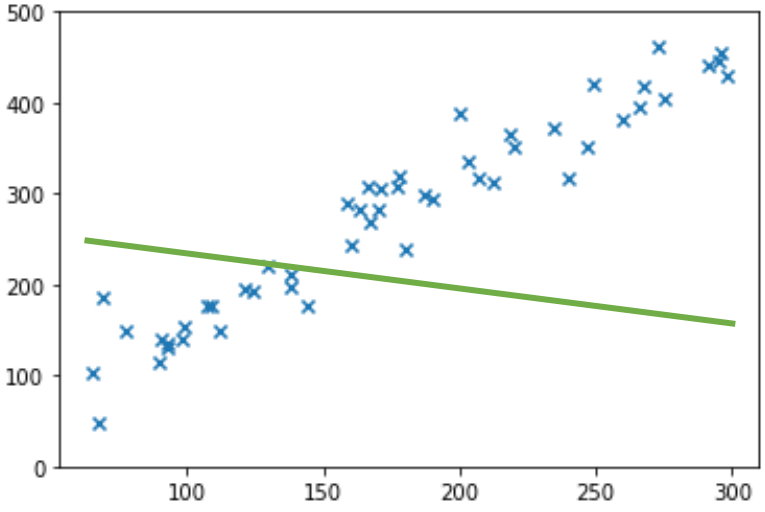
(固定x, $J(w,b)$是一个关于w,b的函数)



单变量线性回归

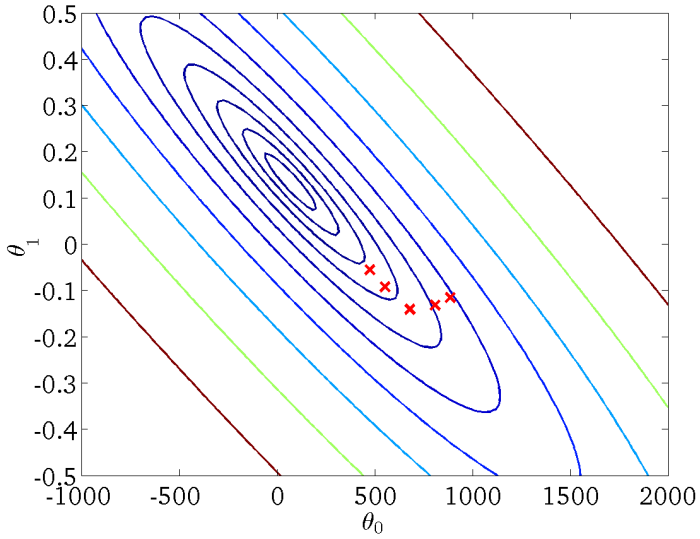
$f(x)$

(固定w和b, $f(x)$是一个关于x的函数)



$J(w,b)$

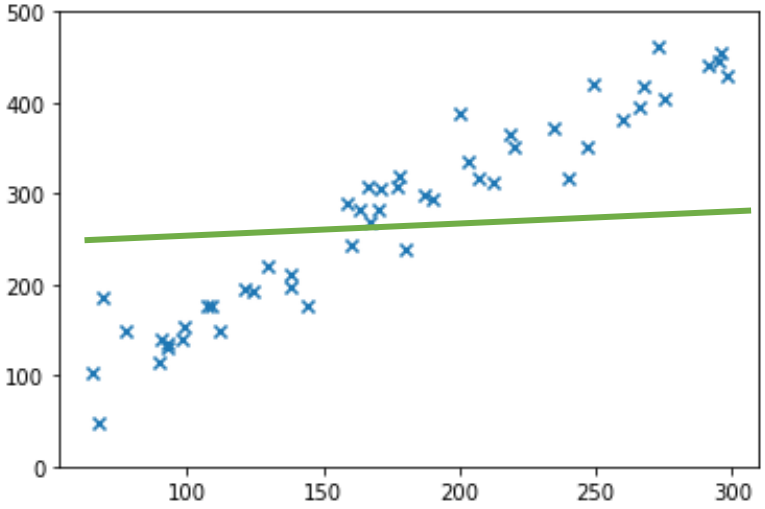
(固定x, $J(w,b)$是一个关于w,b的函数)



单变量线性回归

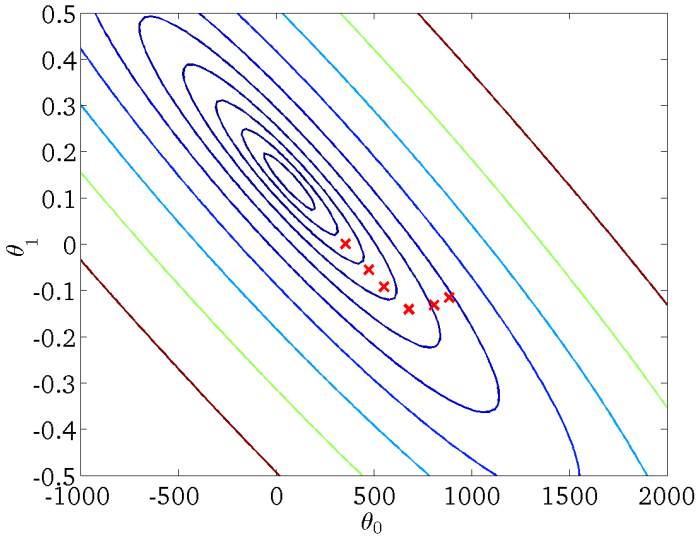
$f(x)$

(固定w和b, $f(x)$是一个关于x的函数)



$J(w,b)$

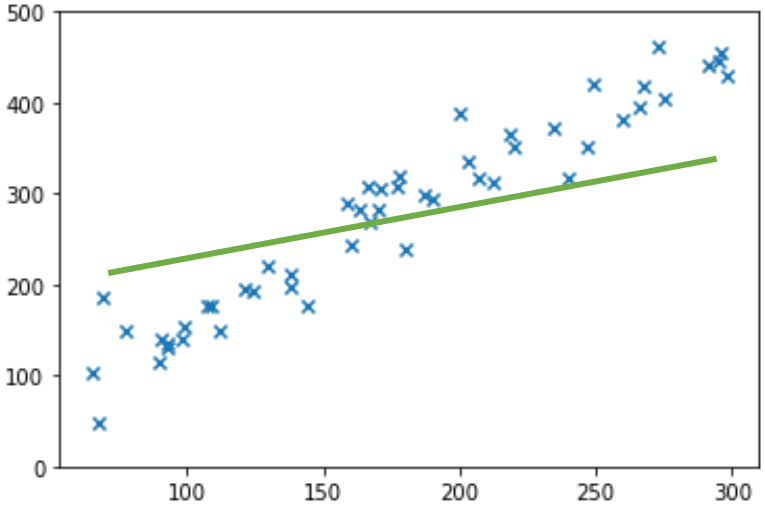
(固定x, $J(w,b)$是一个关于w,b的函数)



单变量线性回归

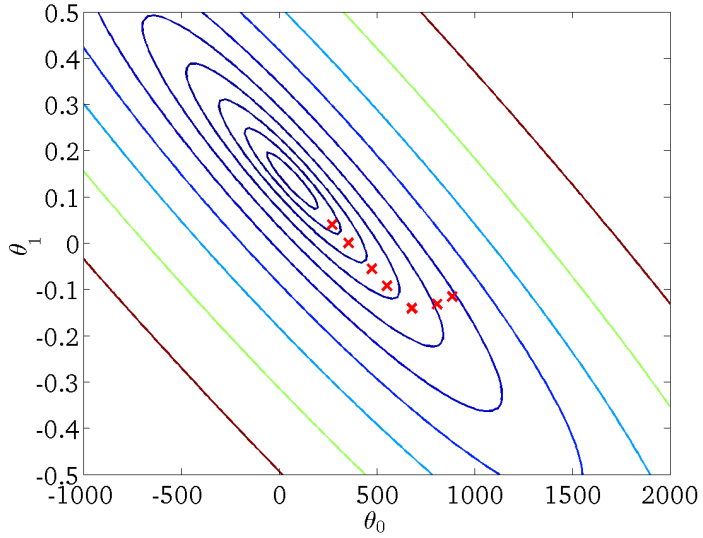
$f(x)$

(固定w和b, $f(x)$是一个关于x的函数)



$J(w,b)$

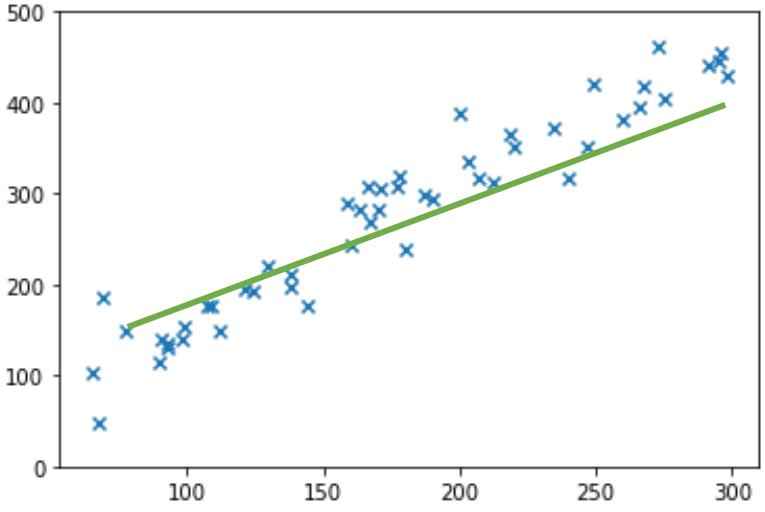
(固定x, $J(w,b)$是一个关于w,b的函数)



单变量线性回归

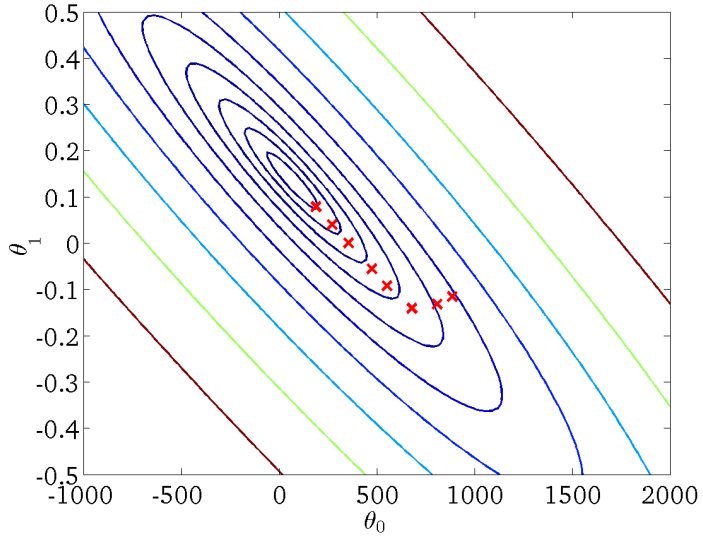
$f(x)$

(固定w和b, $f(x)$是一个关于x的函数)



$J(w,b)$

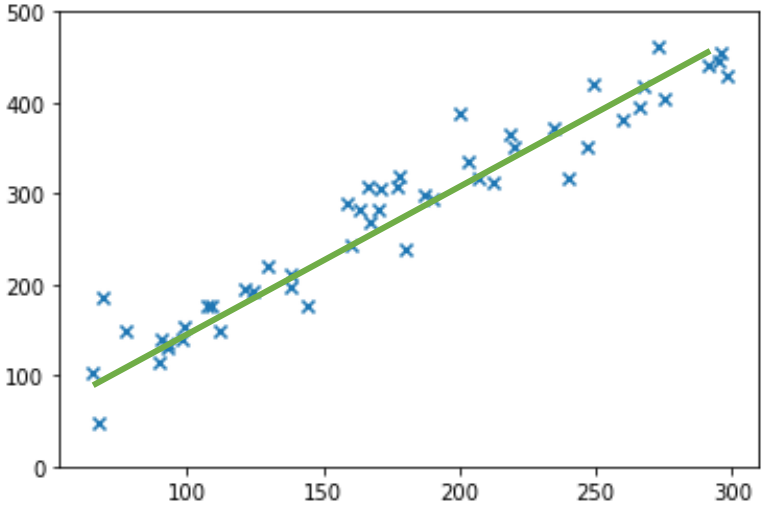
(固定x, $J(w,b)$是一个关于w,b的函数)



单变量线性回归

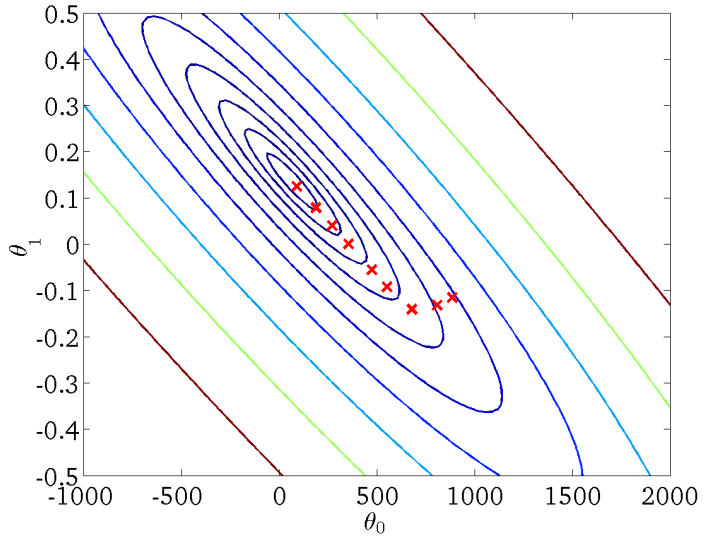
$f(x)$

(固定w和b, $f(x)$是一个关于x的函数)

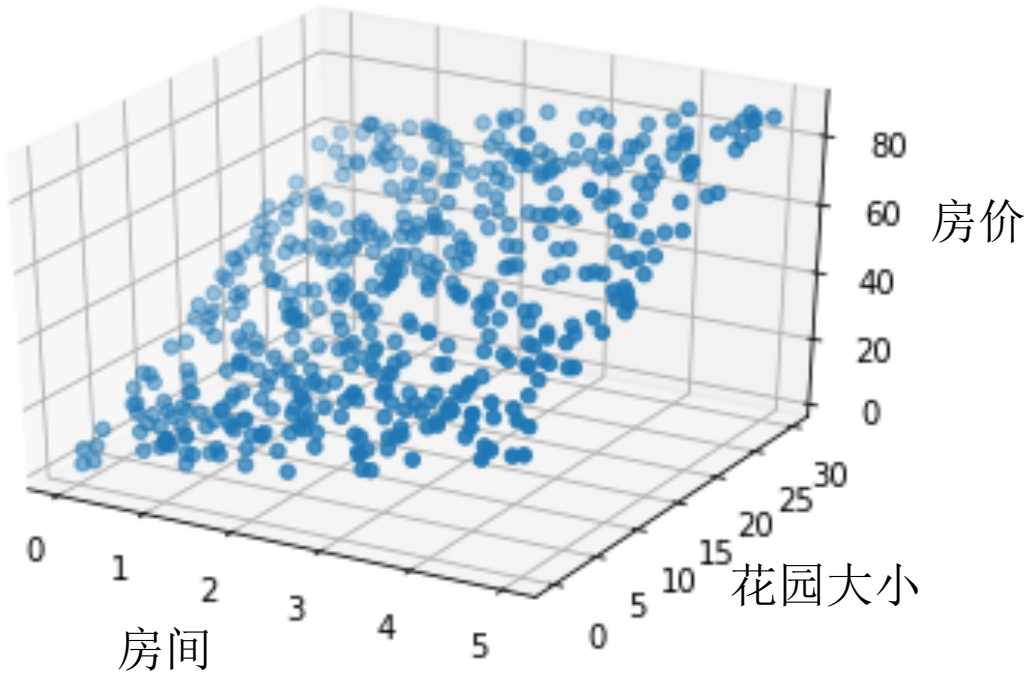


$J(w,b)$

(固定x, $J(w,b)$是一个关于w,b的函数)



房价预测（进阶）



多变量线性回归

面积（x_1）	市中心距离（ x_2 ）	价格（y）
92.6487	84.6225	210.063
276.64	54.7967	417.446
162.741	4.34897	275.893
173.522	9.96174	254.641
…	…	…

$$f(x_1, x_2) = w_1 x_1 + w_2 x_2 + b$$

多变量线性回归

- 映射

$$f(x) = w_1 x_1 + w_2 x_2 + b$$

- 参数

$$w_1, w_2, b$$

- 代价函数

$$J(w,b) = \frac{1}{m}\sum_{i=1}^{m}\left(f\left(x^{(i)}\right) - y^{(i)}\right)^2$$

- 目标

$$\min_{w_1,w_2,b} \frac{1}{m}\sum_{i=1}^{m}\left(f\left(x^{(i)}\right) - y^{(i)}\right)^2$$

多变量线性回归

- 梯度方向

$$\frac{\partial J(w_1, w_2, b)}{\partial w_1} = \frac{2}{m}\sum_{i=1}^{m}\left(w_1 x_1^{(i)} + w_2 x_2^{(i)} + b - y^{(i)}\right)x_1^{(i)}$$

$$\frac{\partial J(w_1, w_2, b)}{\partial w_2} = \frac{2}{m}\sum_{i=1}^{m}\left(w_1 x_1^{(i)} + w_2 x_2^{(i)} + b - y^{(i)}\right)x_2^{(i)}$$

$$\frac{\partial J(w_1, w_2, b)}{\partial b} = \frac{2}{m}\sum_{i=1}^{m}\left(w x^{(i)} + b - y^{(i)}\right)$$

- 更新变量w, b

$$w_1 = w_1 - \alpha\frac{\partial J(w_1, w_2, b)}{\partial w_1}$$

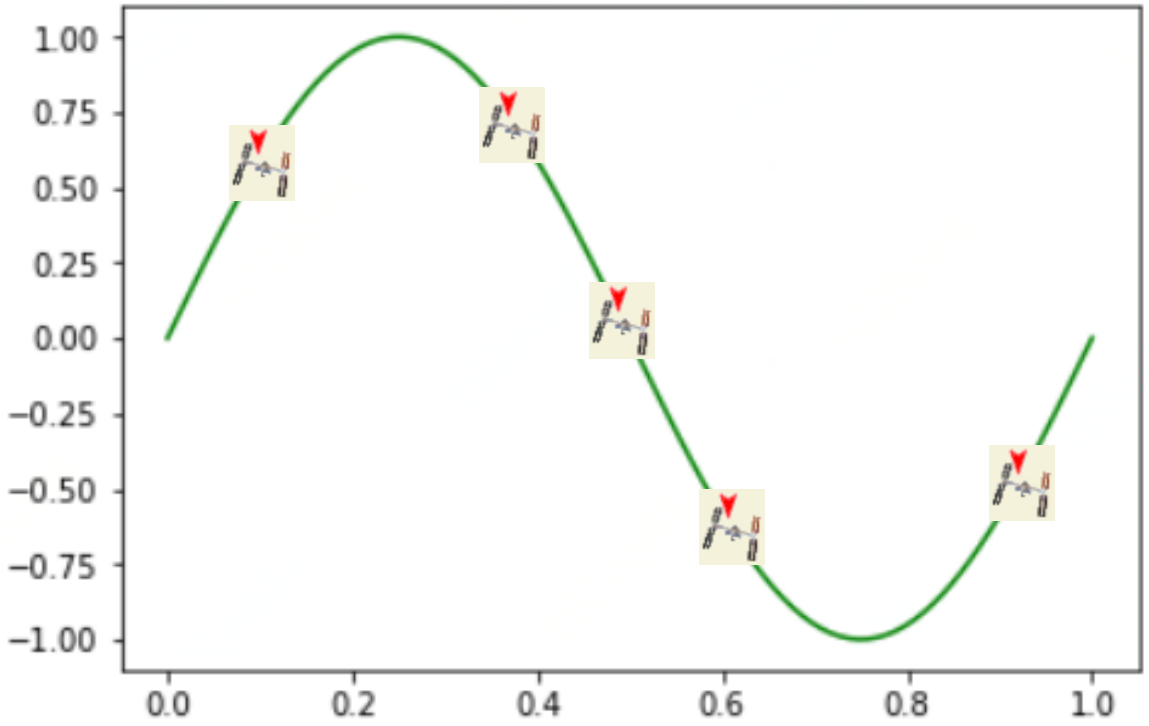
$$w_2 = w_2 - \alpha\frac{\partial J(w_1, w_2, b)}{\partial w_2}$$

$$b = b - \alpha\frac{\partial J(w_1, w_2, b)}{\partial b}$$

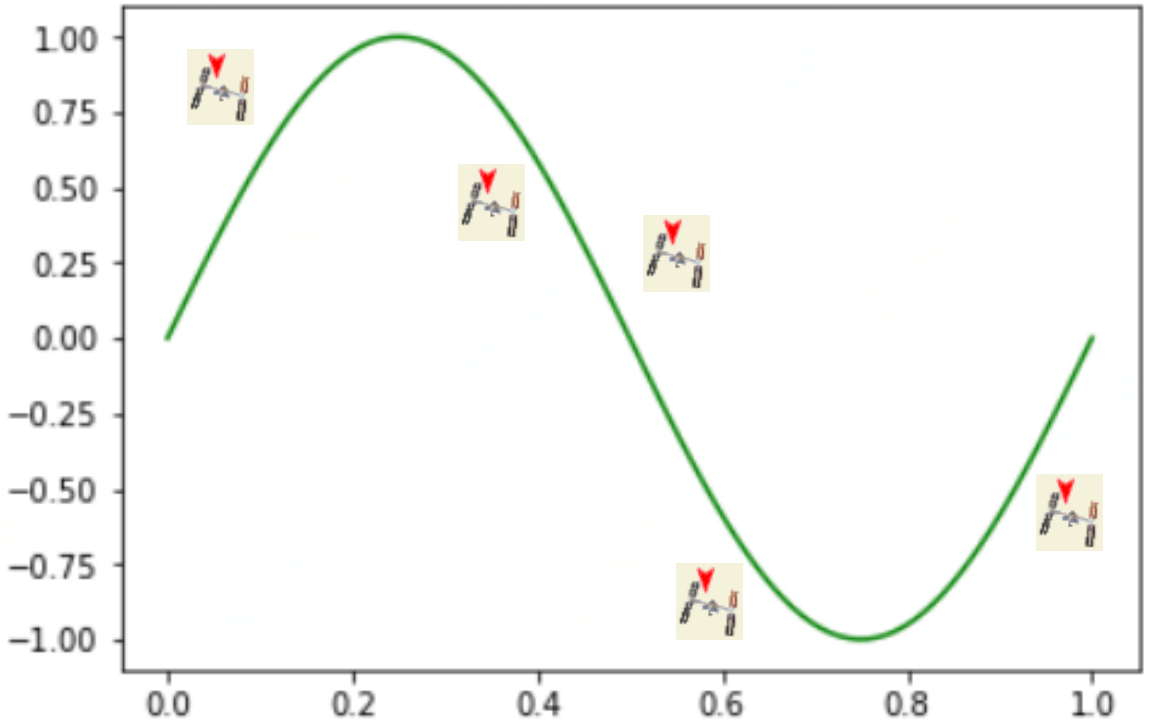
轨道预测（课后作业1）


NORTH AMERICA
EUROPE
ASIA
Atlantic
Ocean
AFRICA
SOUTH AMERICA
Indian Ocean
AUSTRAL
Leaflet | Esri

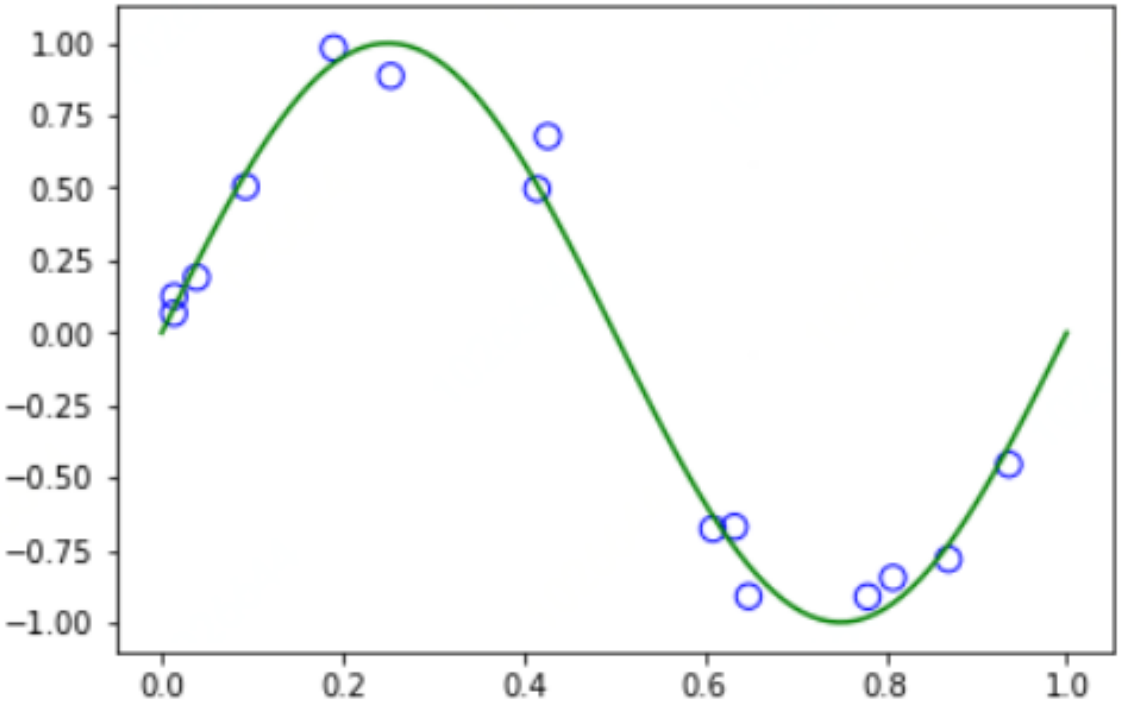
问题简化



噪声干扰

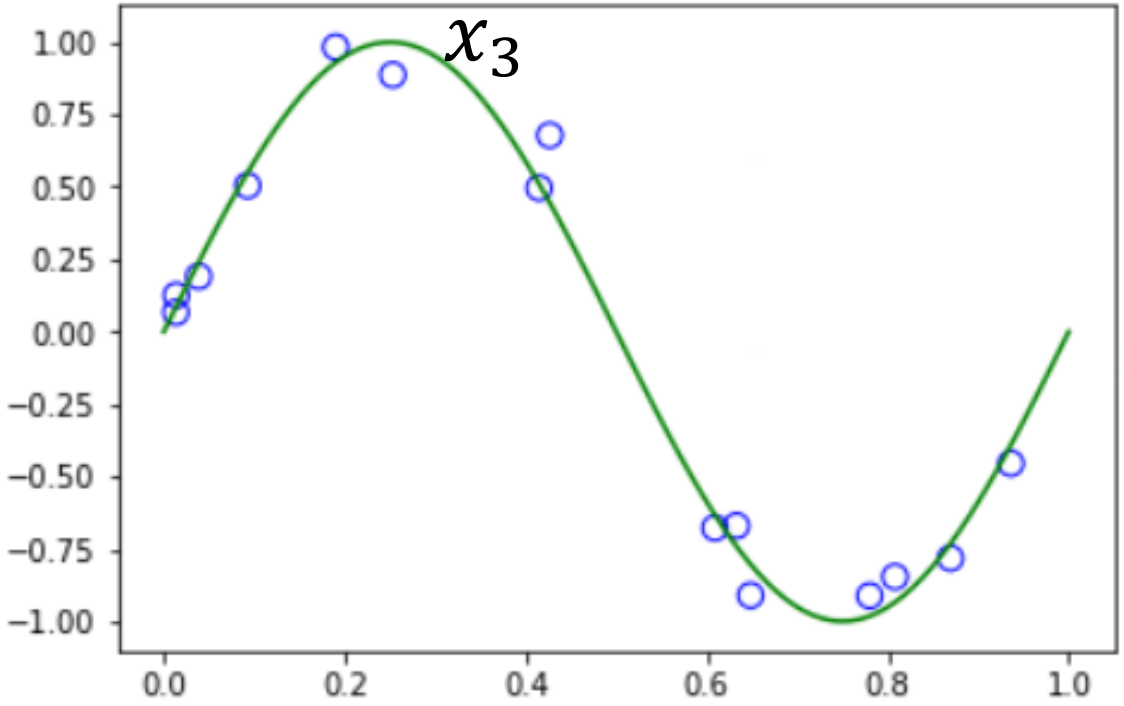


回归问题



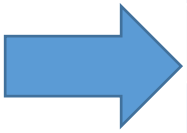
特征

Input	Output
$x^{(1)} = 0.03$	$y^{(1)} = 0.19$
$x^{(2)} = 0.78$	$y^{(2)} = -0.91$
$x^{(3)} = 0.25$	$y^{(3)} = 0.88$
…	…
$x^{(15)} = 0.64$	$y^{(15)} = -0.91$



特征扩充

Input	Output
$x^{(1)} = 0.03$	$y^{(1)} = 0.19$
$x^{(2)} = 0.78$	$y^{(2)} = -0.91$
$x^{(3)} = 0.25$	$y^{(3)} = 0.88$
…	…
$x^{(15)} = 0.64$	$y^{(15)} = -0.91$



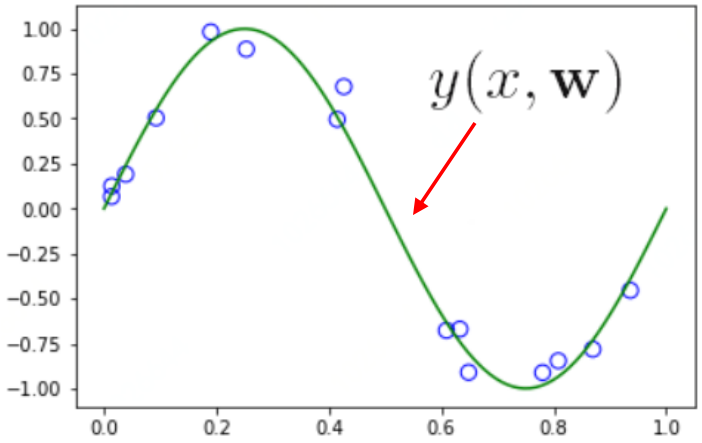
Input			Output
$x_1^{(1)} = 0.03$	$x_2^{(1)} = 0.001$	$x_3^{(1)} = 0.00005$	$y^{(1)} = 0.1907$
$x_1^{(2)} = 0.78$	$x_2^{(2)} = 0.608$	$x_3^{(2)} = 0.47455$	$y^{(2)} = -0.9118$
$x_1^{(3)} = 0.25$	$x_2^{(3)} = 0.0625$	$x_3^{(3)} = 0.01562$	$y^{(3)} = 0.8866$
…	…	…	…
$x_1^{(15)} = 0.64$	$x_2^{(15)} = 0.409$	$x_3^{(15)} = 0.26214$	$y^{(15)} = -0.9104$

x

x_1 x_2 x_3

多项式拟合

Input			Output
$x_1^{(1)} = 0.03$	$x_2^{(1)} = 0.001$	$x_3^{(1)} = 0.00005$	$y^{(1)} = 0.1907$
$x_1^{(2)} = 0.78$	$x_2^{(2)} = 0.608$	$x_3^{(2)} = 0.47455$	$y^{(2)} = -0.9118$
$x_1^{(3)} = 0.25$	$x_2^{(3)} = 0.0625$	$x_3^{(3)} = 0.01562$	$y^{(3)} = 0.8866$
...	...	...	...
$x_1^{(15)} = 0.64$	$x_2^{(15)} = 0.409$	$x_3^{(15)} = 0.26214$	$y^{(15)} = -0.9104$



$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$

作业2

- 使用sklearn进行多变量房价预测