Logistic 回归

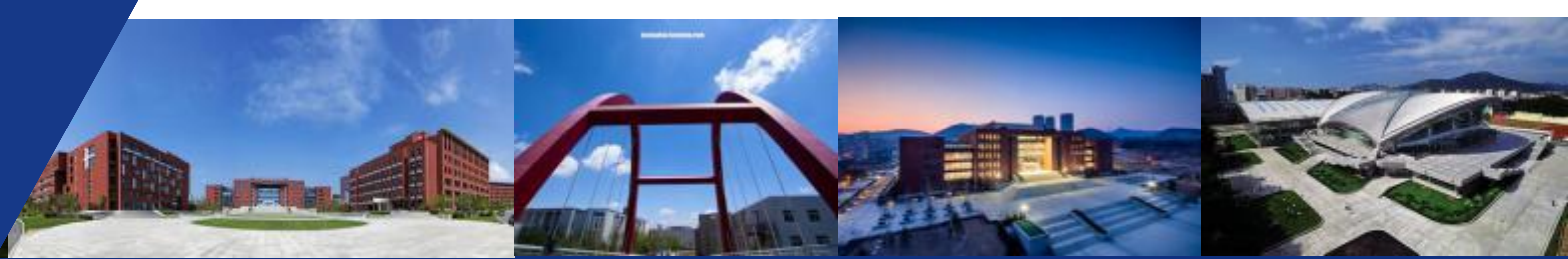




建设单位:　　大连理工大学
建 设 人:　　刘倩

参考书：周志华.机器学习. 清华大学出版社. 2016.

目录
CONTENT



01

PART ONE

线性回归（回顾）

Linear Regression (Recall)

1 线性回归（回顾）

模型

Step 1: define a set of function

$$f_{w,b}(x) = \sum_i w_i x_i + b$$

输出为实数（scalar）

代价函数（e.g.误差）

Step 2: goodness of function

参数求解

Step 3: pick the best function

1 线性回归（回顾）

模型

Step 1: define a set of function

$$f_{w,b}(x) = \sum_i w_i x_i + b$$

输出为实数（scalar）

每改变w或b就得到一个f(x)

代价函数（e.g.误差）

Step 2: goodness of function

参数求解

Step 3: pick the best function

1 线性回归（回顾）

模型

Step 1: define a set of function

$$f_{w,b}(x) = \sum_i w_i x_i + b$$

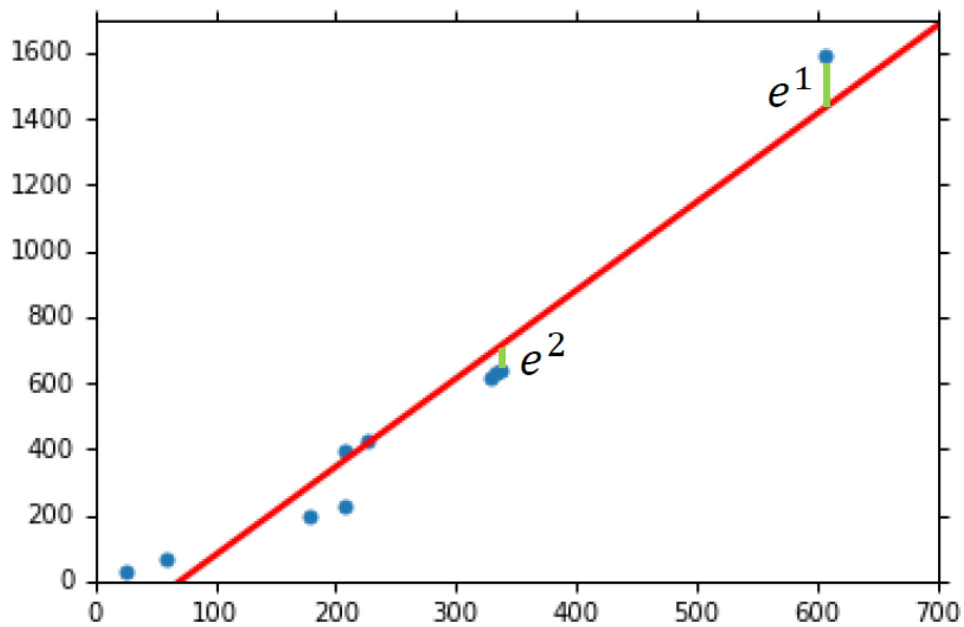
输出为实数（scalar）

每改变w或b就得到一个f(x)

代价函数（e.g.误差）

Step 2: goodness of function

Training Data



参数求解

Step 3: pick the best function

1 线性回归（回顾）

大连理工大学 Dalian University Of Technology

模型

Step 1: define a set of function

$$f_{w,b}(x) = \sum_i w_i x_i + b$$

输出为实数（scalar）

每改变w或b就得到一个f(x)

代价函数(e.g.误差)

Step 2: goodness of function

Training data: $(x^n, \hat{y}^n)$

$\hat{y}^n$: a real number

$$L(f) = \frac{1}{2}\sum_n \left(f(x^n) - \hat{y}^n\right)^2$$

定义代价函数用以衡量模型好坏

参数求解

Step 3: pick the best function

1 线性回归（回顾）

模型

Step 1: define a set of function

$$f_{w,b}(x) = \sum_i w_i x_i + b$$

输出为实数（scalar）

每改变w或b就得到一个f(x)

代价函数 (e.g.误差)

Step 2: goodness of function

Training data: $(x^n, \hat{y}^n)$

$\hat{y}^n$: a real number

$$L(f) = \frac{1}{2}\sum_n \left(f(x^n) - \hat{y}^n\right)^2$$

定义代价函数用以衡量模型好坏

参数求解

Step 3: pick the best function

min $L(f)$ 方法：梯度下降

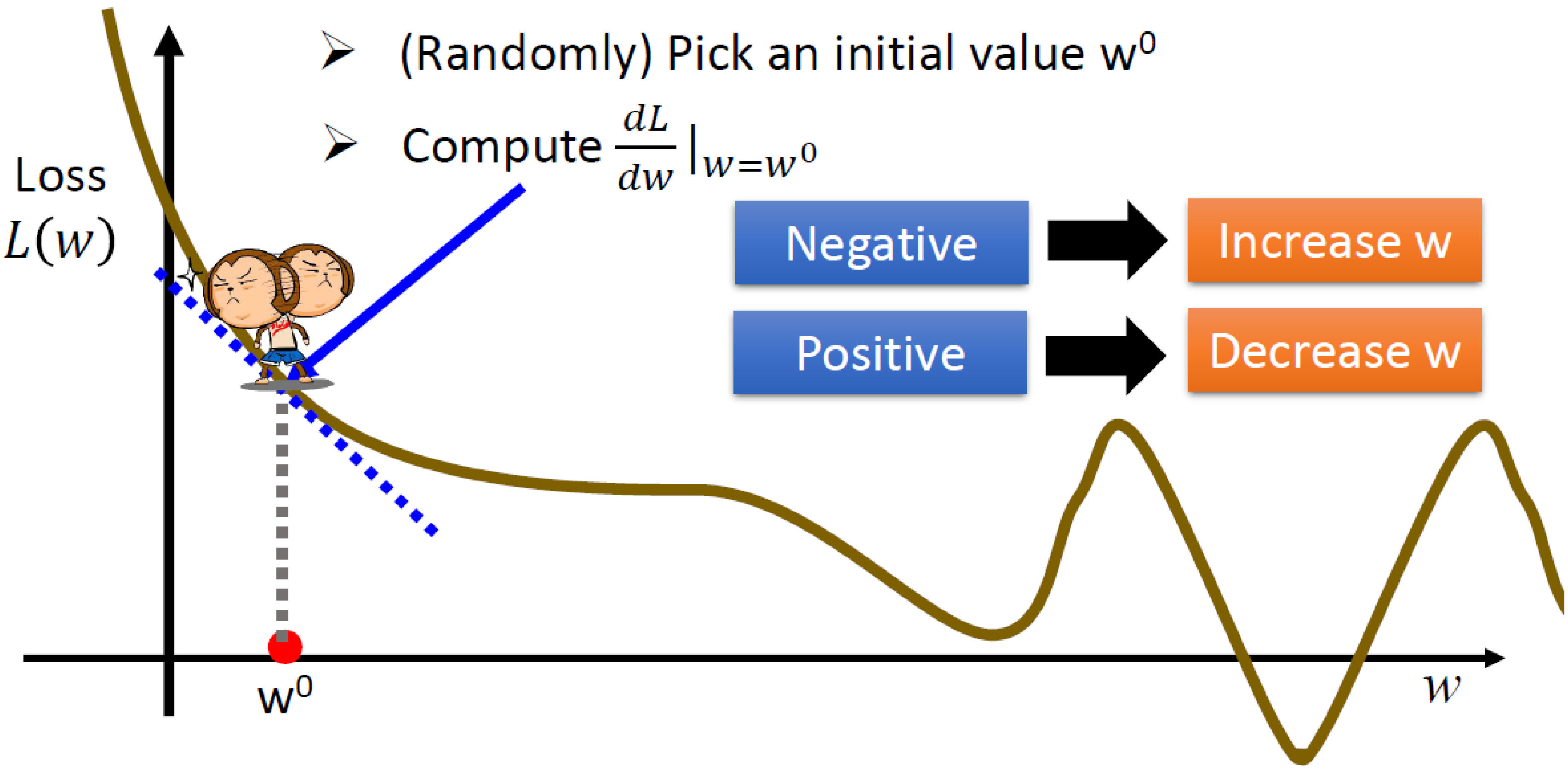
$$w_i \leftarrow w_i - \eta \sum_n -\left(\underline{\hat{y}^n - f_{w,b}(x^n)}\right) x_i^n$$

http://chico386.pixnet.net/album/photo/171572850

Step 3: Gradient Descent

$$w^* = arg\min_w L(w)$$

- Consider loss function $L(w)$ with one parameter w:



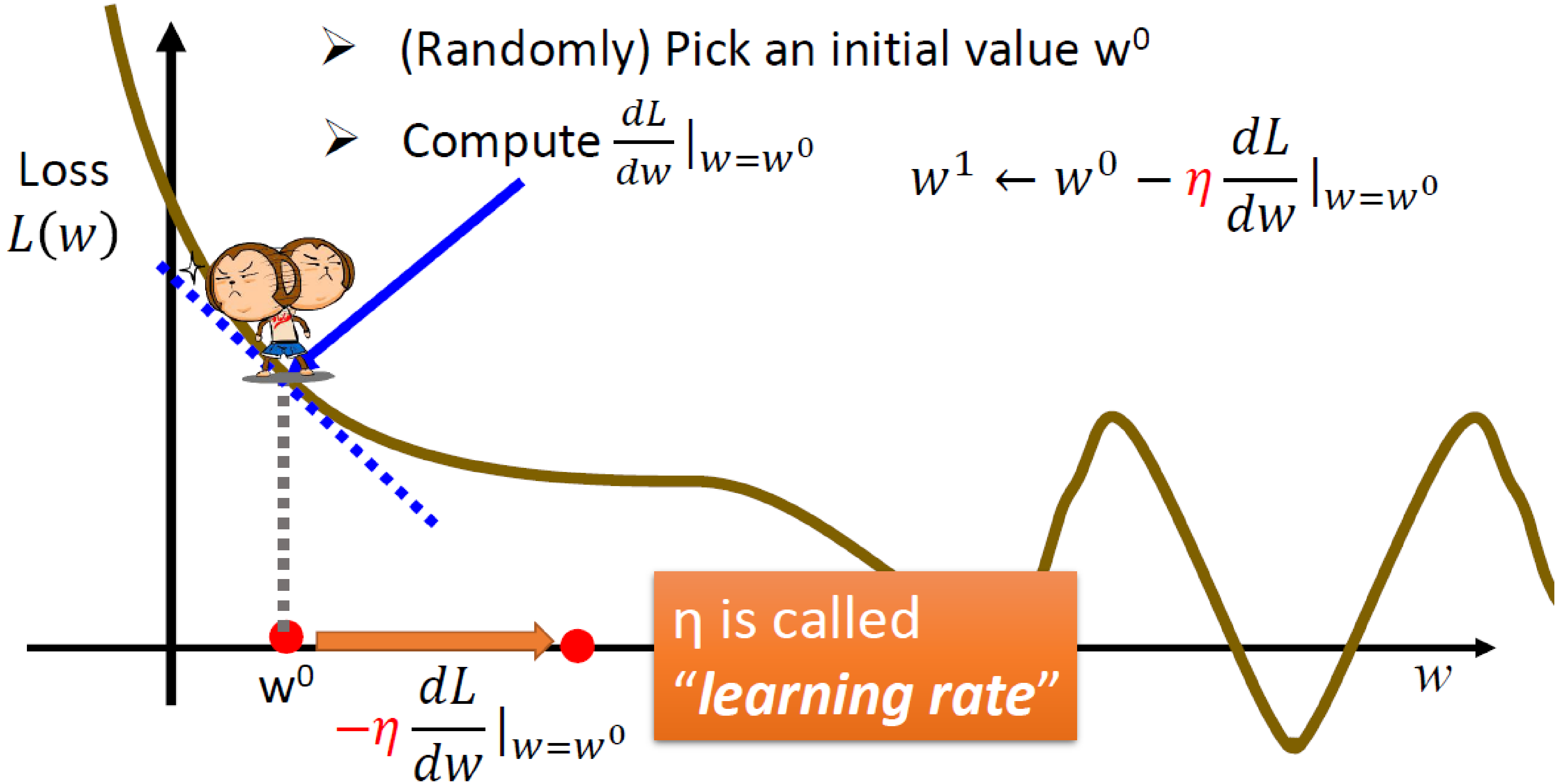
http://chico386.pixnet.net/album/photo/171572850

Step 3: Gradient Descent

$$w^* = arg\min_w L(w)$$

- Consider loss function $L(w)$ with one parameter w:
  - (Randomly) Pick an initial value w^0
  - Compute $\frac{dL}{dw}|_{w=w^0}$

$$w^1 \leftarrow w^0 - \eta \frac{dL}{dw}|_{w=w^0}$$

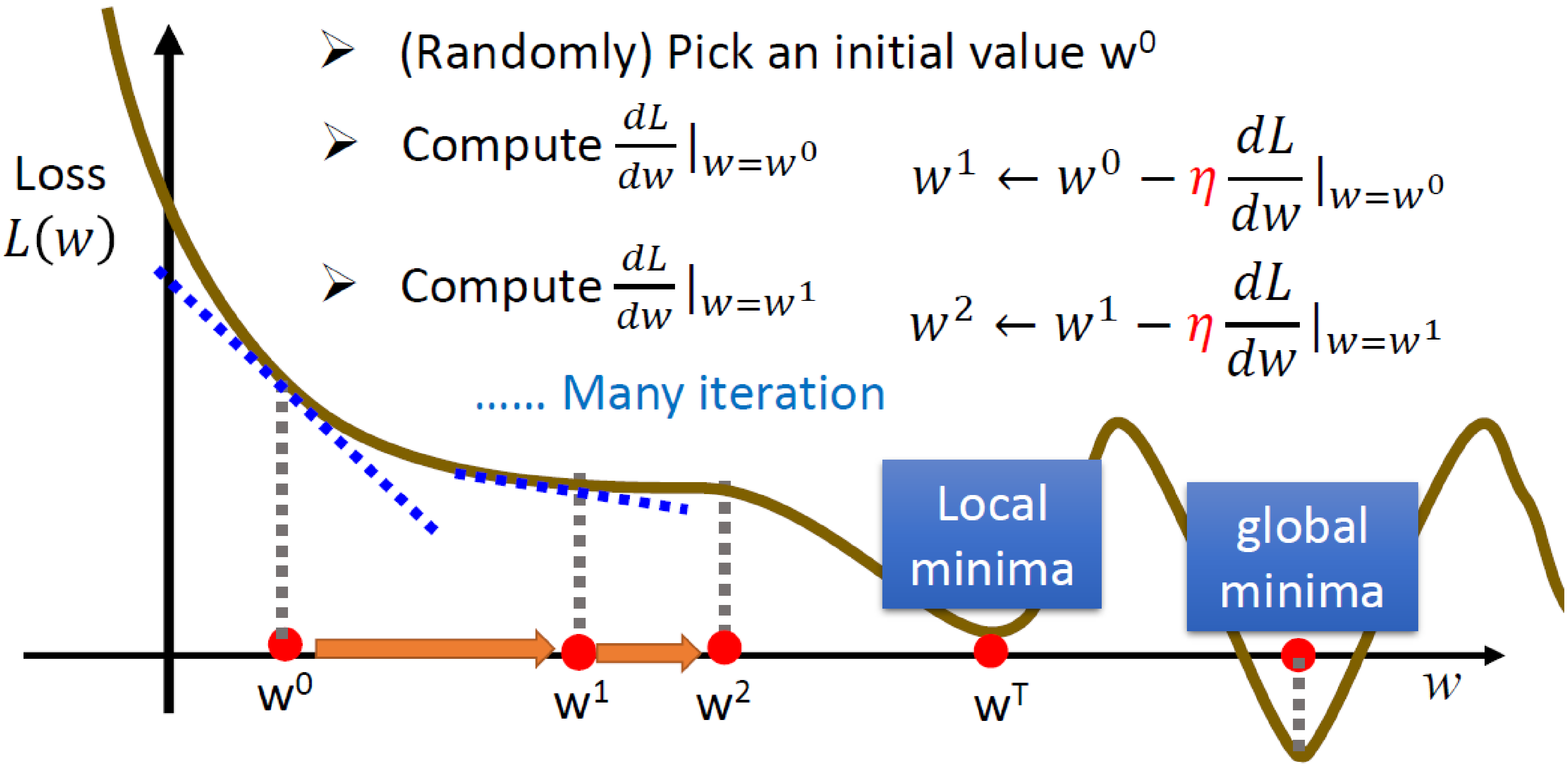


Step 3: Gradient Descent

$$w^* = arg \min_{w} L(w)$$

- Consider loss function $L(w)$ with one parameter w:
  - (Randomly) Pick an initial value w^0
  - Compute $\frac{dL}{dw}|_{w=w^0}$ $\qquad w^1 \leftarrow w^0 - \eta \frac{dL}{dw}|_{w=w^0}$
  - Compute $\frac{dL}{dw}|_{w=w^1}$ $\qquad w^2 \leftarrow w^1 - \eta \frac{dL}{dw}|_{w=w^1}$

…… Many iteration



Step 3: Gradient Descent

$$\begin{bmatrix} \frac{\partial L}{\partial w} \\ \frac{\partial L}{\partial b} \end{bmatrix}_{\text{gradient}}$$

- How about two parameters? $w^*, b^* = arg \min_{w,b} L(w, b)$

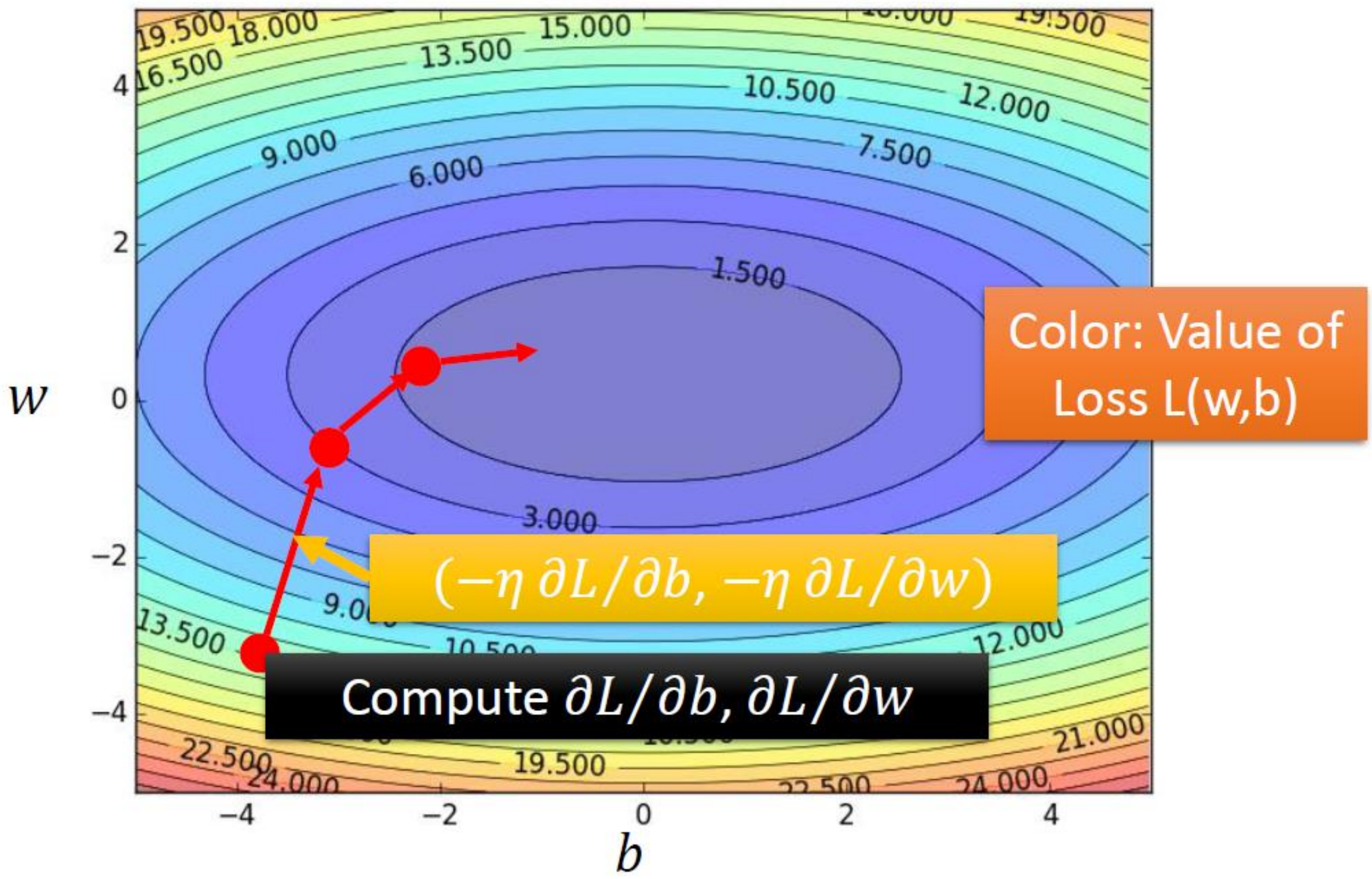
  - (Randomly) Pick an initial value w^0, b^0
  - Compute $\frac{\partial L}{\partial w}|_{w=w^0,b=b^0}, \frac{\partial L}{\partial b}|_{w=w^0,b=b^0}$

$$w^1 \leftarrow w^0 - \eta \frac{\partial L}{\partial w}|_{w=w^0,b=b^0} \qquad b^1 \leftarrow b^0 - \eta \frac{\partial L}{\partial b}|_{w=w^0,b=b^0}$$

  - Compute $\frac{\partial L}{\partial w}|_{w=w^1,b=b^1}, \frac{\partial L}{\partial b}|_{w=w^1,b=b^1}$

$$w^2 \leftarrow w^1 - \eta \frac{\partial L}{\partial w}|_{w=w^1,b=b^1} \qquad b^2 \leftarrow b^1 - \eta \frac{\partial L}{\partial b}|_{w=w^1,b=b^1}$$

Step 3: Gradient Descent



1 线性回归（回顾）

模型

Step 1: define a set of function

$$f_{w,b}(x) = \sum_i w_i x_i + b$$

输出为实数（scalar）

每改变w或b就得到一个f(x)

代价函数 (e.g.误差)

Step 2: goodness of function

Training data: $(x^n, \hat{y}^n)$

$\hat{y}^n$: a real number

$$L(f) = \frac{1}{2}\sum_n \left(f(x^n) - \hat{y}^n\right)^2$$

定义代价函数用以衡量模型好坏

参数求解

Step 3: pick the best function

$\min L(f)$ 方法：梯度下降

$$w_i \leftarrow w_i - \eta \sum_n -\left(\underline{\hat{y}^n - f_{w,b}(x^n)}\right) x_i^n$$

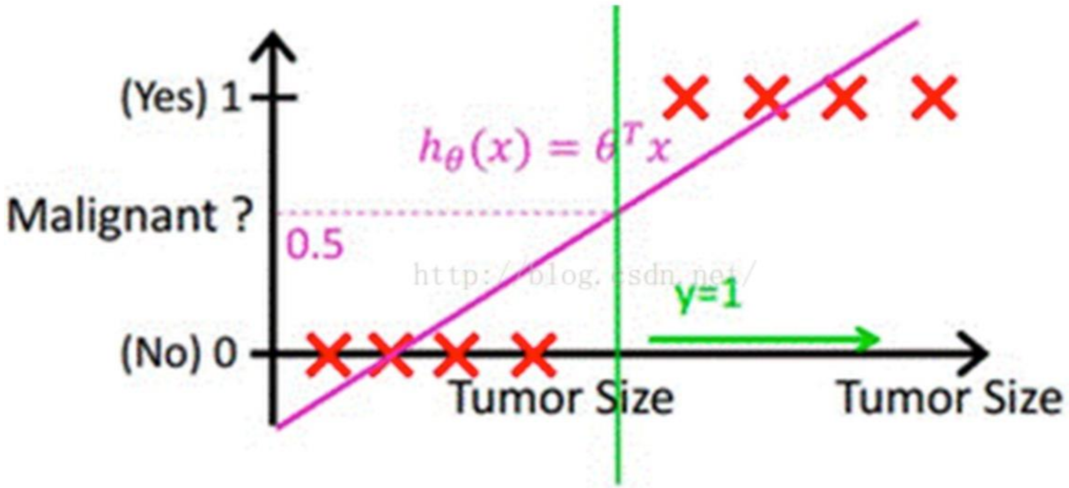
02

PART TWO

从线性回归到Logistic回归

From Linear Regression to Logistic Regression

2 从线性回归到Logistic回归



- ❑ **举例**

  1. 数据：x—肿瘤的大小（1cm，5cm等）

  y—良性或恶性（0或1）

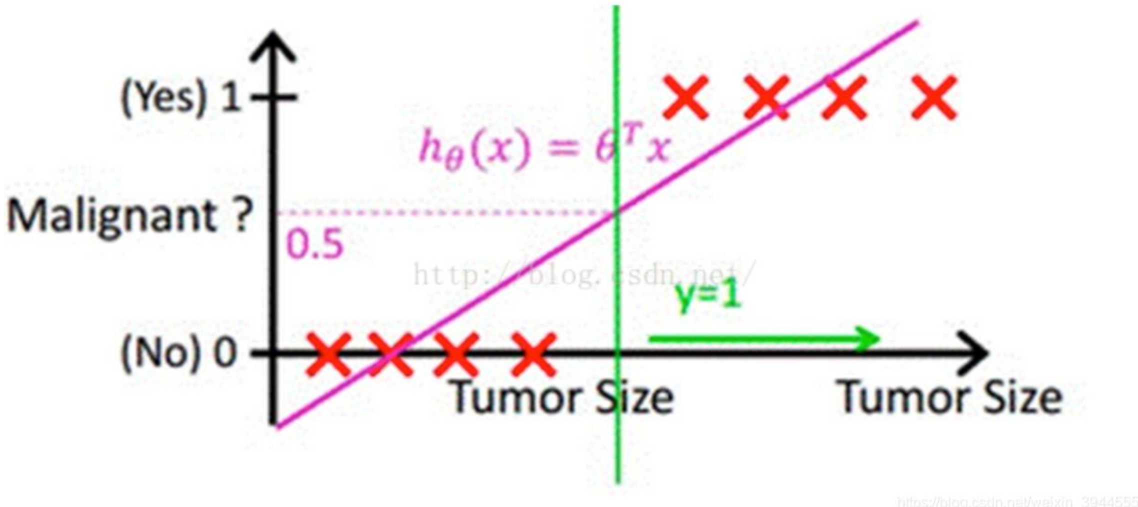
  2. 粉色线：应用线性回归得到的结果

  3. 绿色线：阈值

  y=0.5时x=f^{-1}(0.5)=Th

参考文献：https://blog.csdn.net/weixin_39445556/article/details/83930186

2 从线性回归到Logistic回归



- ❑ **举例**

1. 数据：x—肿瘤的大小（1cm，5cm等）

y—良性或恶性（0或1）

2. 粉色线：应用线性回归得到的结果

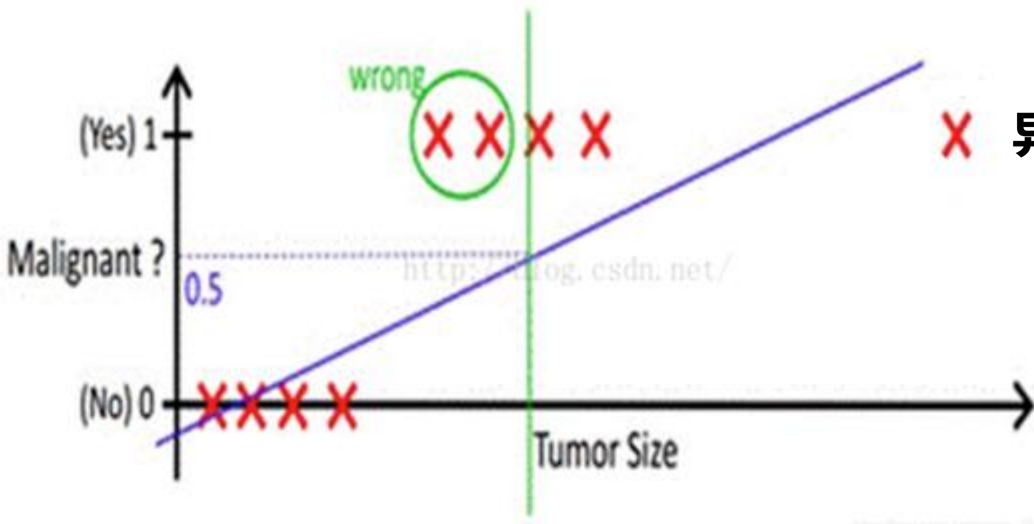
3. 绿色线：阈值

y=0.5时x=f^{-1}(0.5)=Th

y>=0.5时，为恶性
y<0.5时， 为良性

将回归问题转化为分类问题

2 从线性回归到Logistic回归



X 异常数据

- **举例**

  1. 数据：x—肿瘤的大小（1cm，5cm等）

     y—良性或恶性（0或1）

  2. 蓝色线：应用线性回归得到的结果

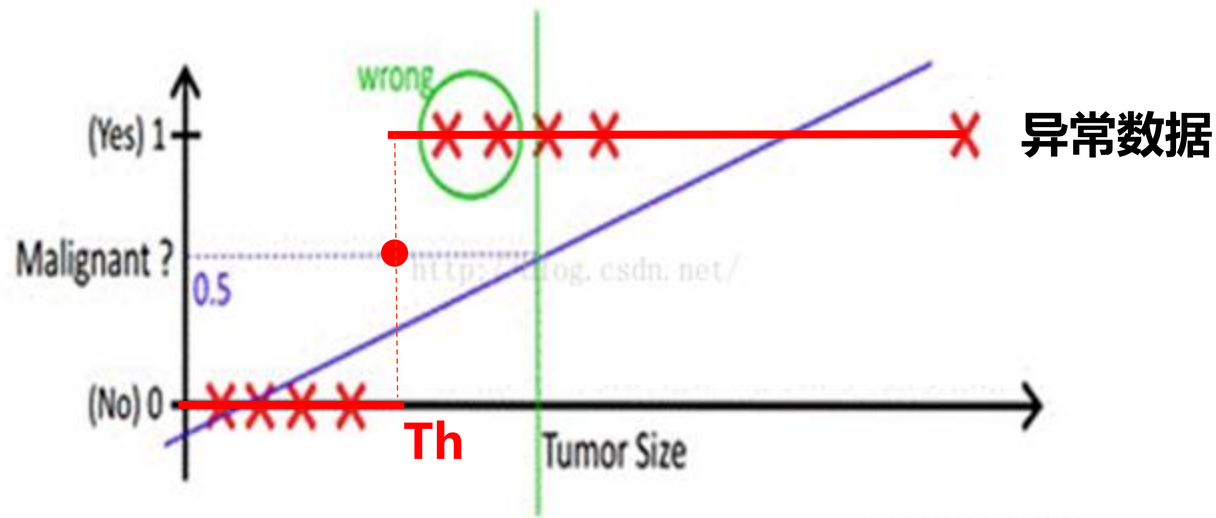
  3. 绿色线：阈值

y=0.5时x=f^{-1}(0.5)=Th

~~y>=0.5时，为恶性~~
~~y<0.5时， 为良性~~

参考文献：https://blog.csdn.net/weixin_39445556/article/details/83930186

2 从线性回归到Logistic回归



- **举例**

  1. 数据：x—肿瘤的大小（1cm，5cm等）

  y—良性或恶性（0或1）

  2. 蓝色线：应用线性回归得到的结果

  3. 绿色线：阈值

  y=0.5时$x=f^{-1}(0.5)=Th$

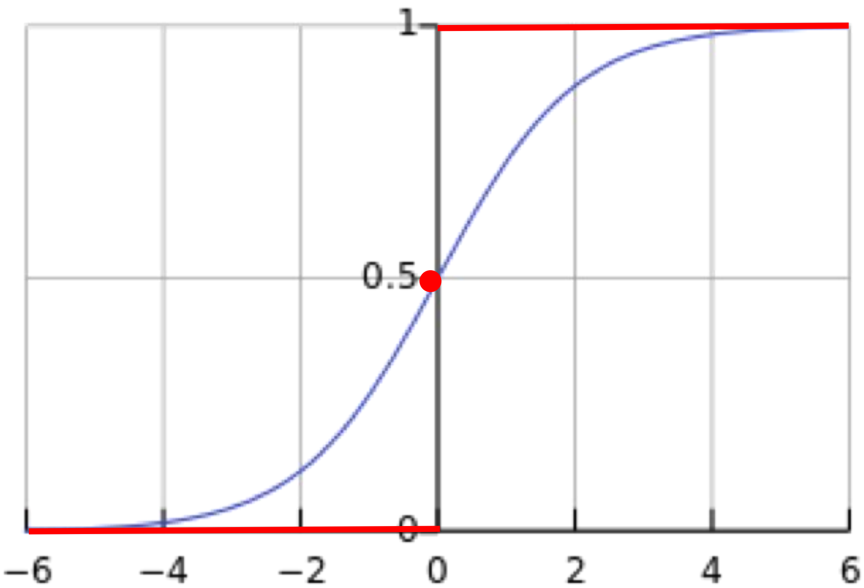
$$y=f(x)=\begin{cases}0 & x<Th(\text{阈值})\\0.5 & x=Th(\text{阈值})\\1 & x>Th(\text{阈值})\end{cases}$$

模型为阶跃函数

阶跃函数不连续，非单调可微

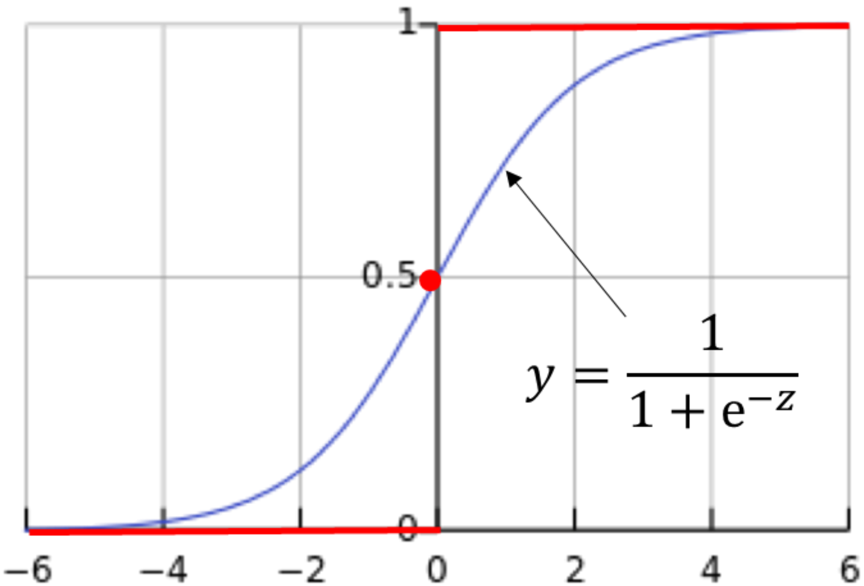
参考文献：https://blog.csdn.net/weixin_39445556/article/details/83930186

2 从线性回归到Logistic回归



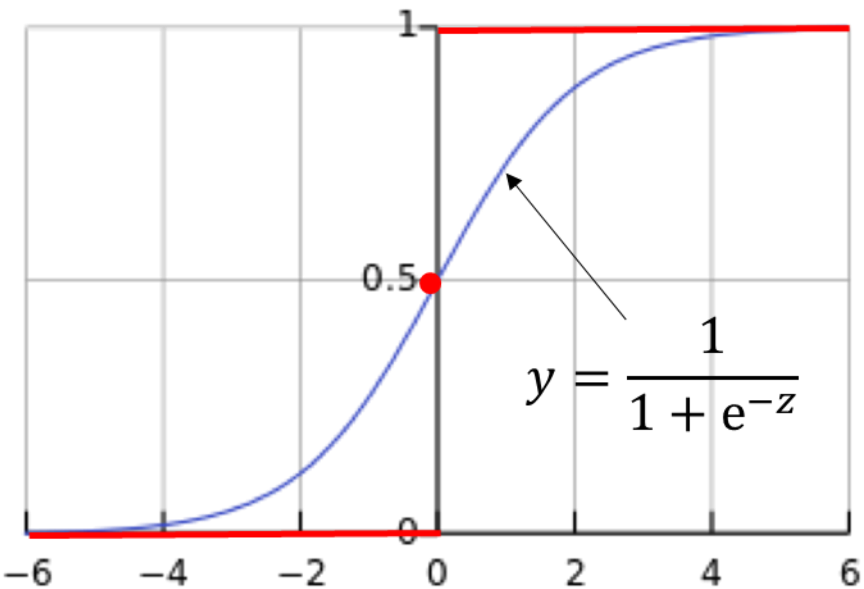
- ❑ 我们希望得到能在一定程度上近似阶跃函数的替代函数，并希望它单调可微
- ❑ 对数几率函数（Logistic function）就是这样一个常用的替代函数

2 从线性回归到Logistic回归



- ❑ 我们希望得到能在一定程度上**近似阶跃函数**的替代函数，并希望它**单调可微**
- ❑ 对数几率函数（**Logistic** function）就是这样一个常用的替代函数
- ❑ 对数几率函数是一种“Sigmoid函数”
- ❑ y在z=0附近变化陡峭

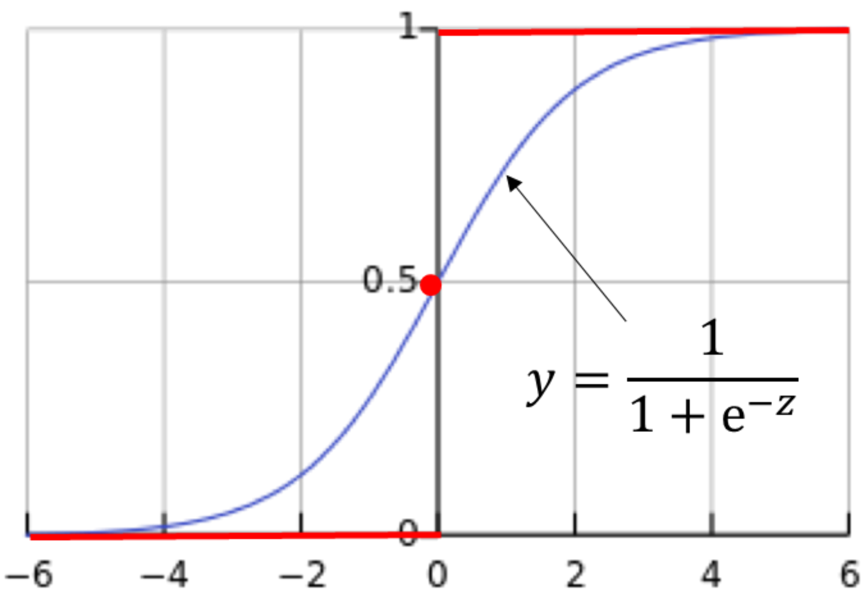
2 从线性回归到Logistic回归



- 我们希望得到能在一定程度上**近似阶跃函数**的替代函数，并希望它**单调可微**
- 对数几率函数（**Logistic** function）就是这样一个常用的替代函数
- 对数几率函数是一种“Sigmoid函数”
- y在z=0附近变化陡峭

$$y = \frac{1}{1 + e^{-(\boldsymbol{\omega}^T \mathbf{x} + b)}} \longrightarrow ln\frac{y}{1-y} = \boldsymbol{\omega}^T \mathbf{x} + b$$

2 从线性回归到Logistic回归



- 我们希望得到能在一定程度上**近似阶跃函数**的替代函数，并希望它**单调可微**
- 对数几率函数（**Logistic** function）就是这样一个常用的替代函数
- 对数几率函数是一种“Sigmoid函数”
- y在z=0附近变化陡峭

$$y=\frac{1}{1+\mathrm{e}^{-(\boldsymbol{\omega}^T\mathbf{x}+b)}} \rightarrow ln\frac{y}{1-y}=\boldsymbol{\omega}^T\mathbf{x}+b$$

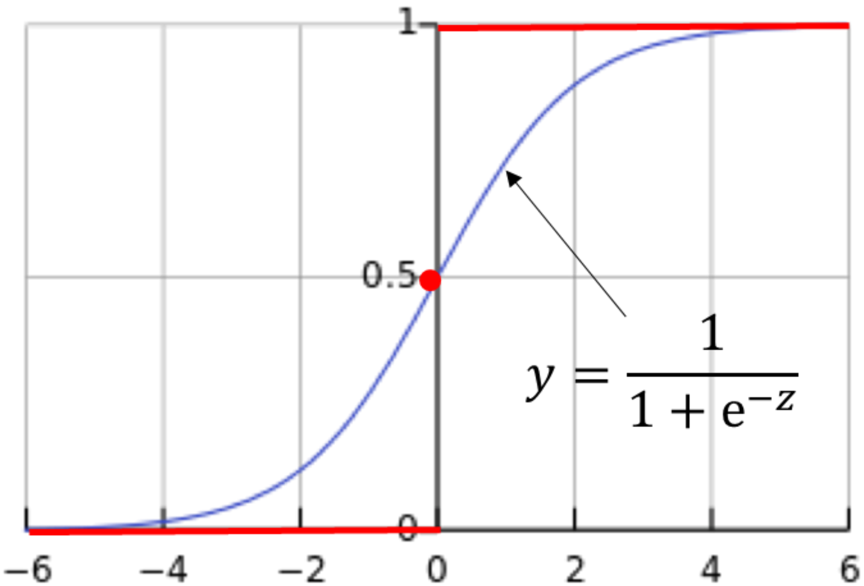
对数几率：$ln\frac{y}{1-y}$

几率：$\frac{y}{1-y}$

将y视为样本x为正例的可能性

1-y视为样本x为负例的可能性

2 从线性回归到Logistic回归



- 我们希望得到能在一定程度上**近似阶跃函数**的替代函数，并希望它**单调可微**
- 对数几率函数（**Logistic** function）就是这样一个常用的替代函数
- 对数几率函数是一种“Sigmoid函数”
- y在z=0附近变化陡峭

$$y = \frac{1}{1+e^{-(\omega^T x+b)}} \rightarrow \underbrace{\ln \underbrace{\frac{y}{1-y}}_{\text{几率}}}_{\text{对数几率}} = \omega^T x + b$$

将y视为样本x为正例的可能性

1-y视为样本x为负例的可能性

实际上是用线性回归模型的预测结果逼近真实标记的对数几率，因此称为**Logistic回归**

注意：Logistic回归虽然名字里面带“回归”，其实是一种分类学习方法

03

PART THREE

Logistic回归

Logistic Regression

3 Logistic回归

模型

Step 1: define a set of function

$\sigma(\cdot)$为Sigmoid函数

$$f_{w,b}(x) = \sigma\left(\sum_i w_i x_i + b\right)$$

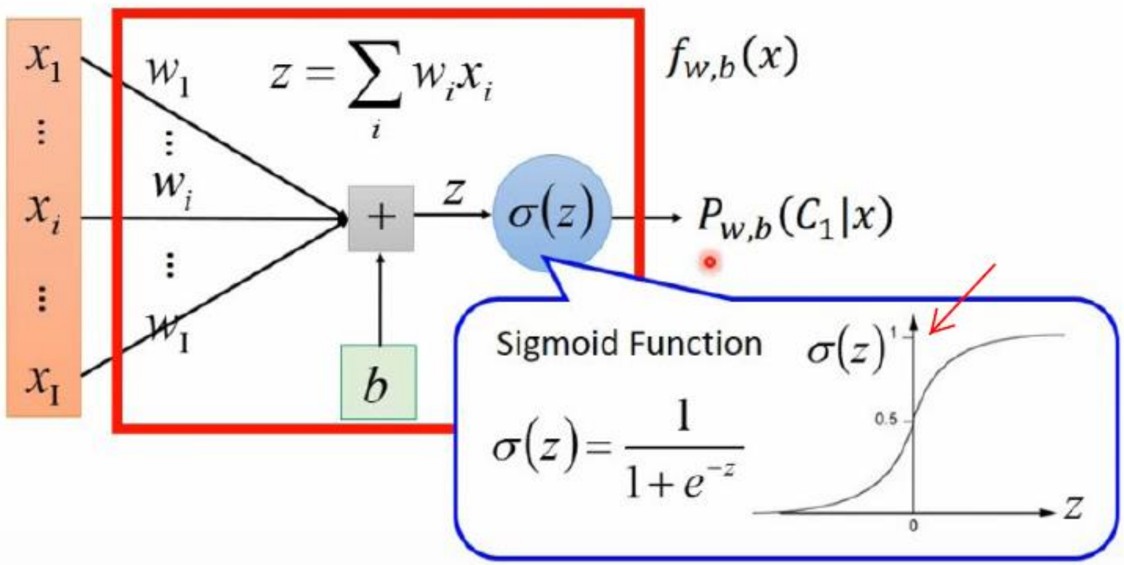
输出为[0,1]的值

代价函数 (e.g.误差)

Step 2: goodness of function

参数求解

Step 3: pick the best function



参考文献：李宏毅.机器学习课件. http://speech.ee.ntu.edu.tw/~tlkagk/courses_ML20.html

3 代价函数与参数求解

$$\ln \frac{y}{1-y} = \boldsymbol{\omega}^T \mathbf{x} + b$$

将y视为类后验概率估计 $p(y=1|\mathbf{x})$

$$\ln \frac{p(y=1|\mathbf{x})}{p(y=0|\mathbf{x})} = \boldsymbol{\omega}^T \mathbf{x} + b$$

$$p(y=1|\mathbf{x}) = \frac{1}{1+e^{-(\boldsymbol{\omega}^T\mathbf{x}+b)}}, \quad p(y=0|\mathbf{x}) = 1 - p(y=1|\mathbf{x})$$

3 代价函数与参数求解

$$\ln\frac{y}{1-y}=\boldsymbol{\omega}^T\mathbf{x}+b$$

将y视为类后验概率估计 $p(y=1|\mathbf{x})$

Logistic回归模型

$$\ln\frac{p(y=1|\mathbf{x})}{p(y=0|\mathbf{x})}=\boldsymbol{\omega}^T\mathbf{x}+b$$

$f(\mathbf{x})$

$$p(y=1|\mathbf{x})=\frac{1}{1+e^{-(\boldsymbol{\omega}^T\mathbf{x}+b)}}, \quad p(y=0|\mathbf{x})=1-p(y=1|\mathbf{x})$$

3 代价函数与参数求解

$$\ln\frac{y}{1-y} = \boldsymbol{\omega}^T\mathbf{x} + b$$

将y视为类后验概率估计 $p(y=1|\mathbf{x})$

Logistic回归模型

$$\ln\frac{p(y=1|\mathbf{x})}{p(y=0|\mathbf{x})} = \boldsymbol{\omega}^T\mathbf{x} + b$$

$f(\mathbf{x})$

$$p(y=1|\mathbf{x}) = \frac{1}{1+e^{-(\boldsymbol{\omega}^T\mathbf{x}+b)}}, \quad p(y=0|\mathbf{x}) = 1 - p(y=1|\mathbf{x})$$

训练数据 $(x^n, \hat{y}^n)$

最大似然估计 $L(\boldsymbol{\omega}, b) = \Pi\, p(\hat{y}^i | x^i; \boldsymbol{\omega}, b) = p(y=1|x^i)^{\Sigma\hat{y}^i} p(y=0|x^i)^{\Sigma(1-\hat{y}^i)}$

3 代价函数与参数求解

$$\ln\frac{y}{1-y}=\boldsymbol{\omega}^T\mathbf{x}+b$$

将y视为类后验概率估计 $p(y=1|\mathbf{x})$

Logistic回归模型

$$\ln\frac{p(y=1|\mathbf{x})}{p(y=0|\mathbf{x})}=\boldsymbol{\omega}^T\mathbf{x}+b$$

$f(\mathbf{x})$

$$p(y=1|\mathbf{x})=\frac{1}{1+e^{-(\boldsymbol{\omega}^T\mathbf{x}+b)}},\quad p(y=0|\mathbf{x})=1-p(y=1|\mathbf{x})$$

训练数据 $(x^n,\hat{y}^n)$

最大似然估计 $L(\boldsymbol{\omega},b)=\Pi\, p\left(\hat{y}^i\middle|x^i;\boldsymbol{\omega},b\right)=p\left(y=1\middle|x^i\right)^{\Sigma\hat{y}^i}p\left(y=0\middle|x^i\right)^{\Sigma(1-\hat{y}^i)}$

取对数

$$L(f)=\sum_i\hat{y}^i\ln p\left(y=1\middle|x^i\right)+\sum_i(1-\hat{y}^i)\ln p\left(y=0\middle|x^i\right)=\sum_i\hat{y}^i\ln f\left(x^i\right)+\sum_i(1-\hat{y}^i)\ln\left(1-f\left(x^i\right)\right)$$

3 代价函数与参数求解

$$\ln\frac{y}{1-y} = \boldsymbol{\omega}^T\mathbf{x} + b$$

将y视为类后验概率估计 $p(y=1|\mathbf{x})$

Logistic回归模型

$$\ln\frac{p(y=1|\mathbf{x})}{p(y=0|\mathbf{x})} = \boldsymbol{\omega}^T\mathbf{x} + b$$

$f(\mathbf{x})$

$$\boxed{p(y=1|\mathbf{x}) = \frac{1}{1+e^{-(\boldsymbol{\omega}^T\mathbf{x}+b)}}, \quad p(y=0|\mathbf{x}) = 1 - p(y=1|\mathbf{x})}$$

训练数据 $(x^n, \hat{y}^n)$

最大似然估计 $L(\boldsymbol{\omega}, b) = \Pi\, p(\hat{y}^i | x^i; \boldsymbol{\omega}, b) = p(y=1|x^i)^{\Sigma\hat{y}^i} p(y=0|x^i)^{\Sigma(1-\hat{y}^i)}$

取对数

$$L(f) = \sum_i \hat{y}^i \ln p(y=1|x^i) + \sum_i (1-\hat{y}^i)\ln p(y=0|x^i) = \sum_i \hat{y}^i \ln f(x^i) + \sum_i (1-\hat{y}^i)\ln\left(1 - f(x^i)\right)$$

损失函数 $l(f) = -\sum_n \{\hat{y}^n \ln f(x^n) + (1-\hat{y}^n)\ln(1 - f(x^n))\}$

• 选择步长，梯度更新参数 $\theta_i=\theta_{i-1}-\alpha\nabla J(\theta_{i-1})$

大连理工大学 Dalian University Of Technology

3 代价函数与参数求解

损失函数

$$l(f)=-\sum_{n}\{\hat{y}^n\ln f(x^n)+(1-\hat{y}^n)\ln\left(1-f(x^n)\right)\}$$

代价函数

$$J(\boldsymbol{\omega},b)=\min_{\mathbf{w},b} l(f)$$

最大化似然函数等价于最小化损失函数

•选择步长，梯度更新参数 $\theta_i = \theta_{i-1} - \alpha_i \nabla J(\theta_{i-1})$

大连理工大学 Dalian University Of Technology

3 代价函数与参数求解

损失函数

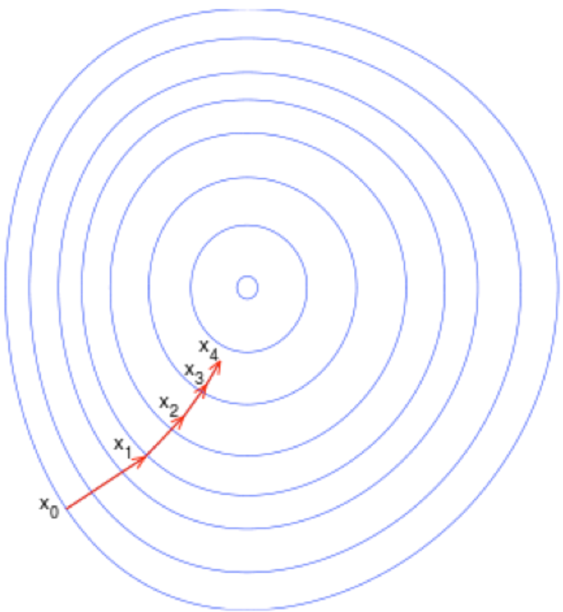
$$l(f) = -\sum_{n} \{\hat{y}^n \ln f(x^n) + (1-\hat{y}^n)\ln\big(1-f(x^n)\big)\}$$

代价函数

$$J(\boldsymbol{\omega}, b) = \min_{\mathbf{w},b} l(f)$$

最大化似然函数等价于最小化损失函数

参数求解方法举例：梯度下降(Gradient Descent)



基本步骤

1. 选择下降方向（梯度方向，$\nabla J(\boldsymbol{\omega}, b)$）
2. 选择步长，更新参数

$$w_i \leftarrow w_i - \eta \sum_{n} -\left(\underline{\hat{y}^n - f_{w,b}(x^n)}\right) x_i^n$$

3. 重复以上两步直到满足终止条件

参考文献：https://tech.meituan.com/2015/05/08/intro-to-logistic-regression.html

04

PART FOUR

总结

Conclusions

4 总结—Logistic回归 vs 线性回归

模型	Step 1: define a set of function
代价函数 (e.g.误差)	Step 2: goodness of function
参数求解	Step 3: pick the best function

4 总结—Logistic回归 vs 线性回归

大连理工大学 Dalian University Of Technology

Logistic Regression	Linear Regression
Step 1: $f_{w,b}(x) = \sigma\left(\sum_i w_i x_i + b\right)$	$f_{w,b}(x) = \sum_i w_i x_i + b$
Output: between 0 and 1	Output: any value

4 总结—Logistic回归 vs 线性回归

	Logistic Regression	***Linear Regression***
Step 1:	$f_{w,b}(x) = \sigma\left(\sum_i w_i x_i + b\right)$ Output: between 0 and 1	$f_{w,b}(x) = \sum_i w_i x_i + b$ Output: any value
Step 2:	Training data: $(x^n, \hat{y}^n)$ $\hat{y}^n$: 1 for class 1, 0 for class 2 $l(f) = \sum_n C(f(x^n), \hat{y}^n)$	Training data: $(x^n, \hat{y}^n)$ $\hat{y}^n$: a real number $L(f) = \frac{1}{2}\sum_n (f(x^n) - \hat{y}^n)^2$

Cross Entropy: $C(f(x^n), \hat{y}^n) = -\{\hat{y}^n \ln f(x^n) + (1-\hat{y}^n)\ln(1-f(x^n))\}$

4 总结—Logistic回归 vs 线性回归

大连理工大学 Dalian University Of Technology

	Logistic Regression	***Linear Regression***
Step 1:	$f_{w,b}(x) = \sigma\left(\sum_i w_i x_i + b\right)$ Output: between 0 and 1	$f_{w,b}(x) = \sum_i w_i x_i + b$ Output: any value
Step 2:	Training data: $(x^n, \hat{y}^n)$ $\boxed{\hat{y}^n\text{: 1 for class 1, 0 for class 2}}$ $l(f) = \sum_n C(f(x^n), \hat{y}^n)$	Training data: $(x^n, \hat{y}^n)$ $\hat{y}^n$: a real number $L(f) = \frac{1}{2}\sum_n (f(x^n) - \hat{y}^n)^2$

梯度下降

Step 3:

Logistic regression: $w_i \leftarrow w_i - \eta \sum_n -\left(\underline{\hat{y}^n - f_{w,b}(x^n)}\right) x_i^n$

Linear regression: $w_i \leftarrow w_i - \eta \sum_n -\left(\underline{\hat{y}^n - f_{w,b}(x^n)}\right) x_i^n$

参考文献：李宏毅.机器学习课件. http://speech.ee.ntu.edu.tw/~tlkagk/courses_ML20.html

谢谢大家！