



中国科学技术大学  
University of Science and Technology of China


# 《人工智能数学原理与算法》

## 第2章：机器学习基础

### 2.6 逻辑回归II

王翔

xiangwang@ustc.edu.cn



**01** 分类问题

**02** 逻辑回归

**03** 过拟合

# 目录



# 逻辑回归中的过拟合

## □ 更复杂的模型往往偏差更少……

使用两个特征的贷款违约预测器可以达到预期效果，但是...

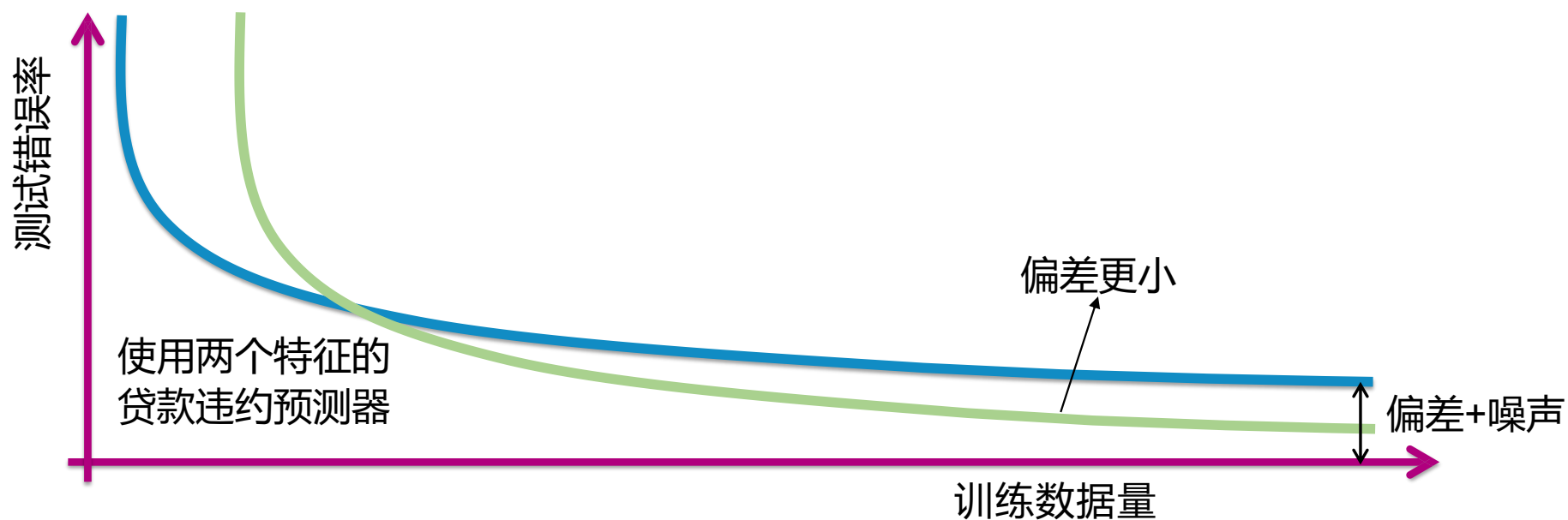
从来不存在100%的正确预测：这个人是否会违约？

更复杂的模型：考虑更多的特征（包含离散特征）？

Word	Weight
年收入	+1.5
贷款金额 ?	

更少的偏差 →  
可能更准确，需要更多数据来学习

## 逻辑回归中的过拟合

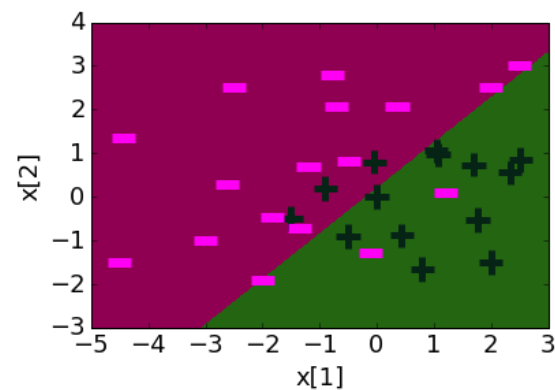
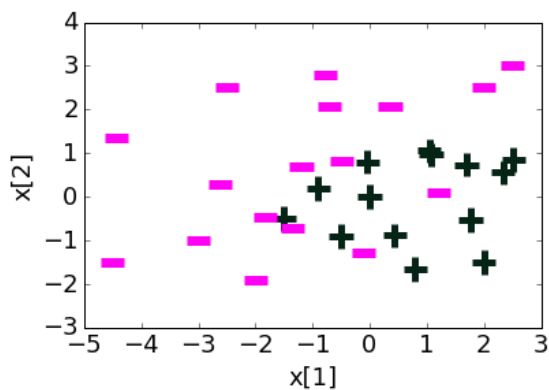


偏差较小的模型往往需要更多数据才能学得好，但如果数据充足，效果会更好

# 逻辑回归中的过拟合

## □ 学习决策边界

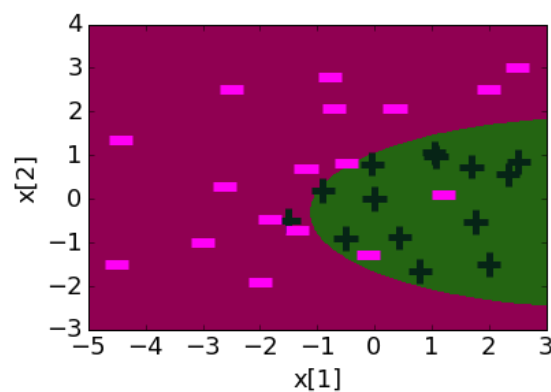
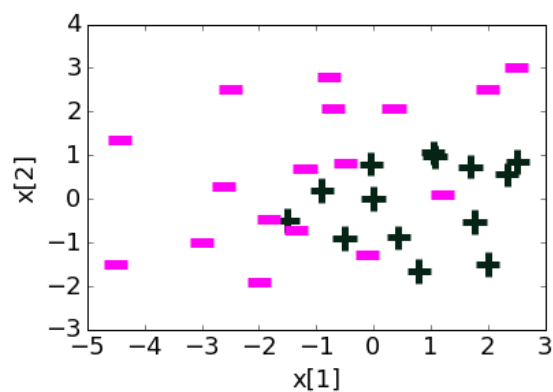
特征	值	学到的系数
$h_0(x)$	1	0.23
$h_1(x)$	$x[1]$	1.12
$h_2(x)$	$x[2]$	-1.07



# 逻辑回归中的过拟合

## □ 二次特征（二维）

特征	值	学到的系数
$h_0(x)$	1	1.68
$h_1(x)$	$x[1]$	1.39
$h_2(x)$	$x[2]$	-0.59
$h_3(x)$	$(x[1])^2$	-0.17
$h_4(x)$	$(x[2])^2$	-0.96

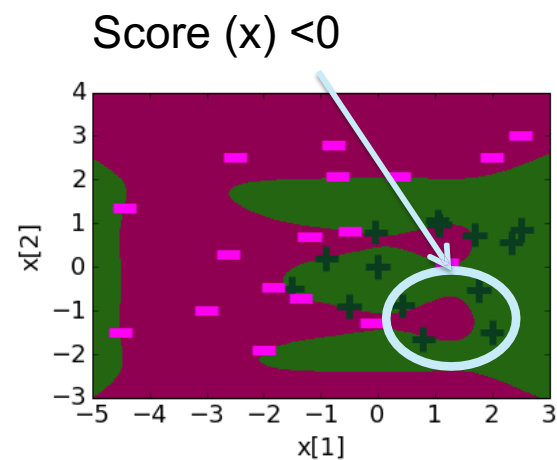
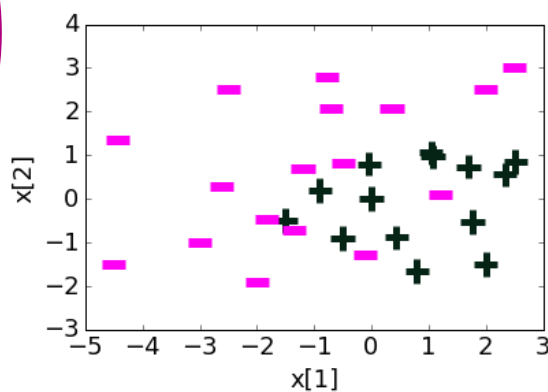


# 逻辑回归中的过拟合

## □ 6阶特征（二维）

特征	值	学到的系数
$h_0(x)$	1	21.6
$h_1(x)$	$x[1]$	5.3
$h_2(x)$	$x[2]$	-42.7
$h_3(x)$	$(x[1])^2$	-15.9
$h_4(x)$	$(x[2])^2$	-48.6
$h_5(x)$	$(x[1])^3$	-11.0
$h_6(x)$	$(x[2])^3$	67.0
$h_7(x)$	$(x[1])^4$	1.5
$h_8(x)$	$(x[2])^4$	48.0
$h_9(x)$	$(x[1])^5$	4.4
$h_{10}(x)$	$(x[2])^5$	-14.2
$h_{11}(x)$	$(x[1])^6$	0.8
$h_{12}(x)$	$(x[2])^6$	-8.6

系数值越来越大

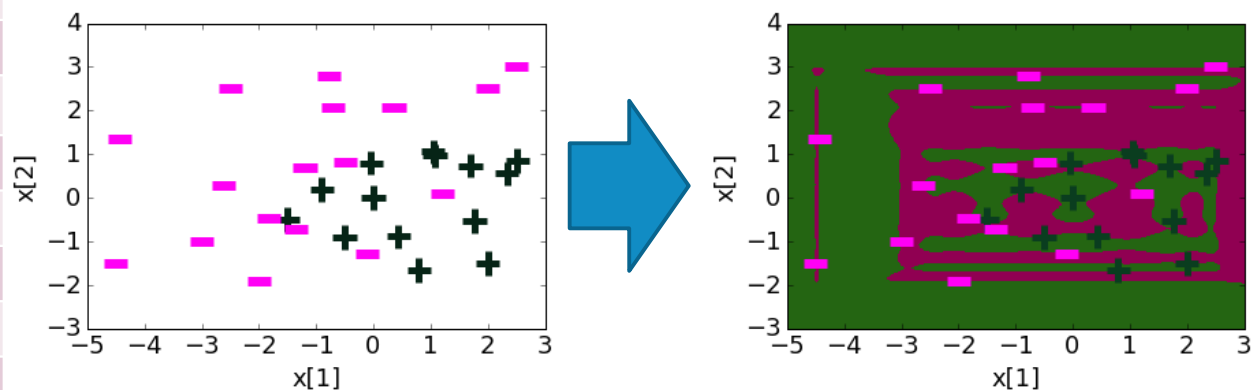


# 逻辑回归中的过拟合

## □ 20阶特征（二维）

特征	值	学到的系数
$h_0(x)$	1	8.7
$h_1(x)$	$x[1]$	5.1
$h_2(x)$	$x[2]$	78.7
...	...	...
$h_{11}(x)$	$(x[1])^6$	-7.5
$h_{12}(x)$	$(x[2])^6$	3803
$h_{13}(x)$	$(x[1])^7$	21.1
$h_{14}(x)$	$(x[2])^7$	-2406
...	...	...
$h_{37}(x)$	$(x[1])^{19}$	$-2 \times 10^{-6}$
$h_{38}(x)$	$(x[2])^{19}$	-0.15
$h_{39}(x)$	$(x[1])^{20}$	$-2 \times 10^{-8}$
$h_{40}(x)$	$(x[2])^{20}$	0.03

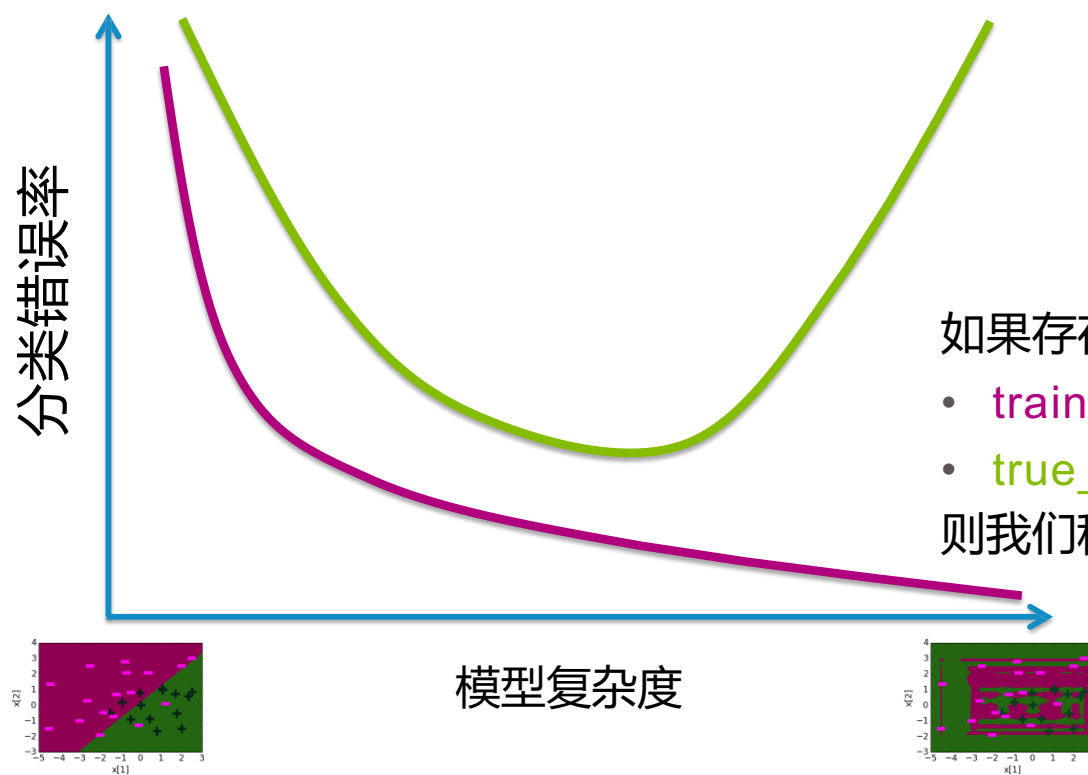
通常，过度拟合与非常大的估计系数 $\hat{w}$ 有关





# 逻辑回归中的过拟合

## □ 分类中的过拟合



如果存在 $w^*$ 满足:

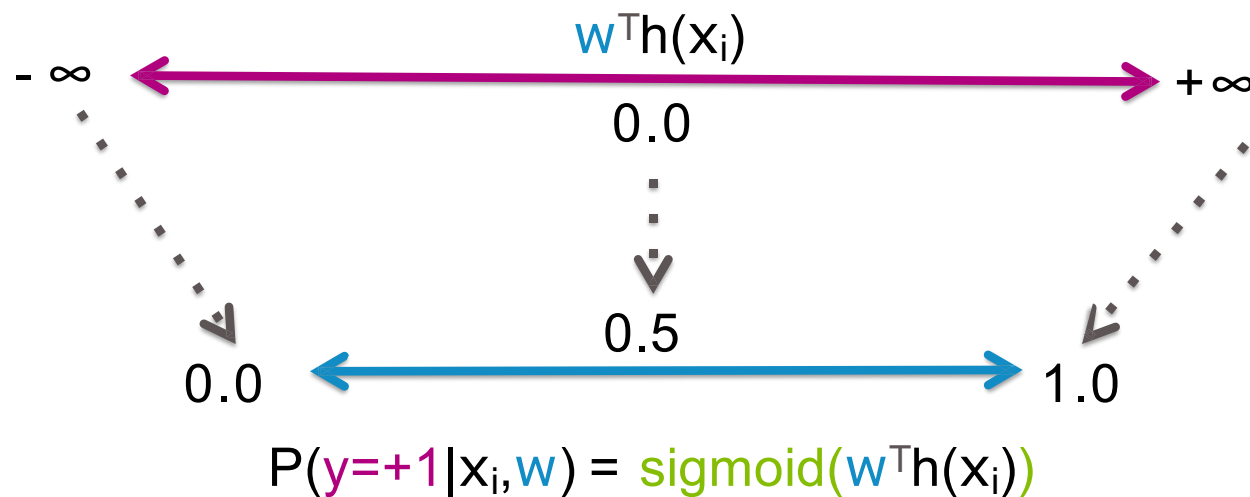
- $\text{training\_error}(w^*) > \text{training\_error}(\hat{w})$
- $\text{true\_error}(w^*) < \text{true\_error}(\hat{w})$

则我们称发生了过拟合

**分类器的过拟合：  
对预测的过度自信**

# 逻辑回归中的过拟合

## □ 逻辑回归模型



# 逻辑回归中的过拟合

## □ 逻辑回归中过度拟合的微妙后果

过拟合 → 系数过大



$\hat{w}^T(x_i)$  远大于零 (或远小于零) →  
 $\text{sigmoid}(\hat{w}^T(x_i))$  趋于 1 (或 0)



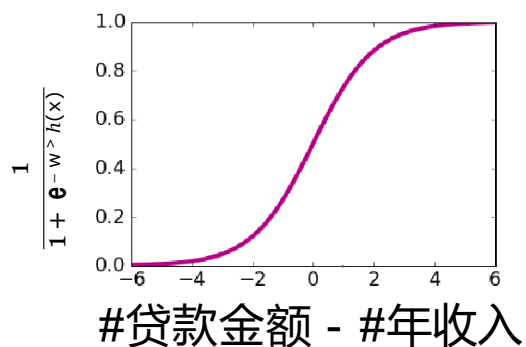
模型对预测变得过度自信

# 逻辑回归中的过拟合

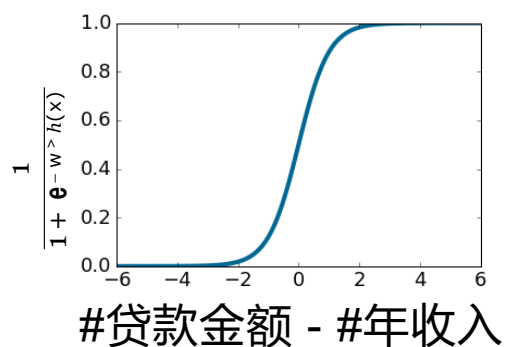
## □ 逻辑回归模型系数的影响

输入: #贷款金额=2, #年收入=1

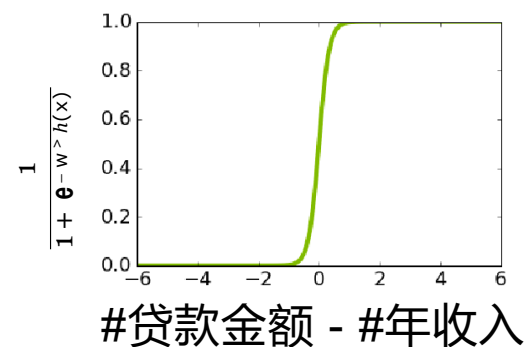
$w_0$	0
$w \# \text{贷款金额}$	+1
$w \# \text{年收入}$	-1



$w_0$	0
$w \# \text{贷款金额}$	+2
$w \# \text{年收入}$	-2



$w_0$	0
$w \# \text{贷款金额}$	+6
$w \# \text{年收入}$	-6

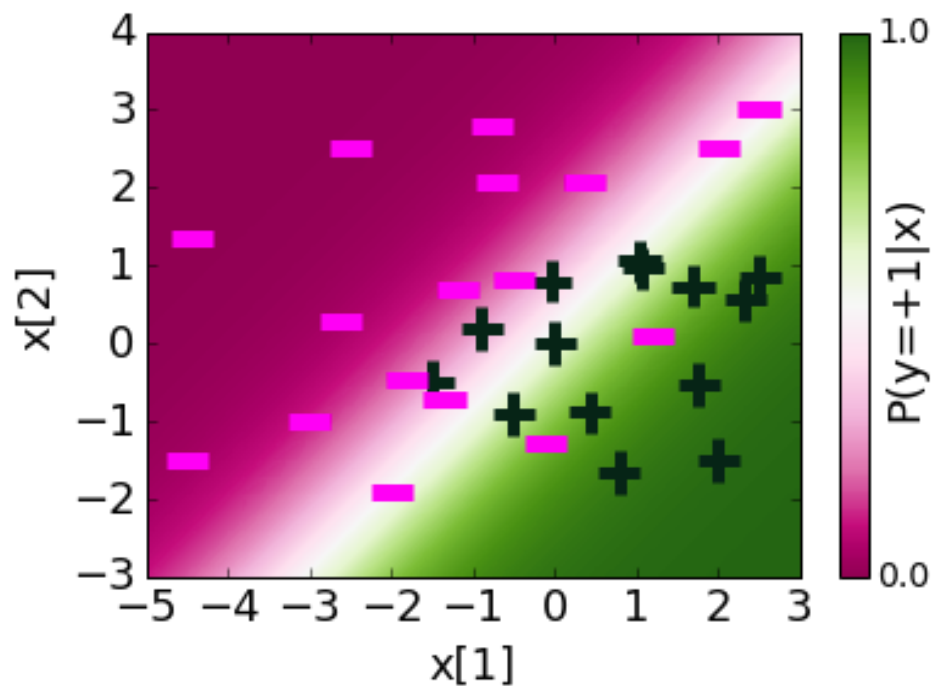


# 逻辑回归中的过拟合

## □ 学到的概率

特征	值	学到的系数
$h_0(x)$	1	0.23
$h_1(x)$	$x[1]$	1.12
$h_2(x)$	$x[2]$	-1.07

$$P(y = +1 | x, w) = \frac{1}{1 + e^{-w^T h(x)}}$$

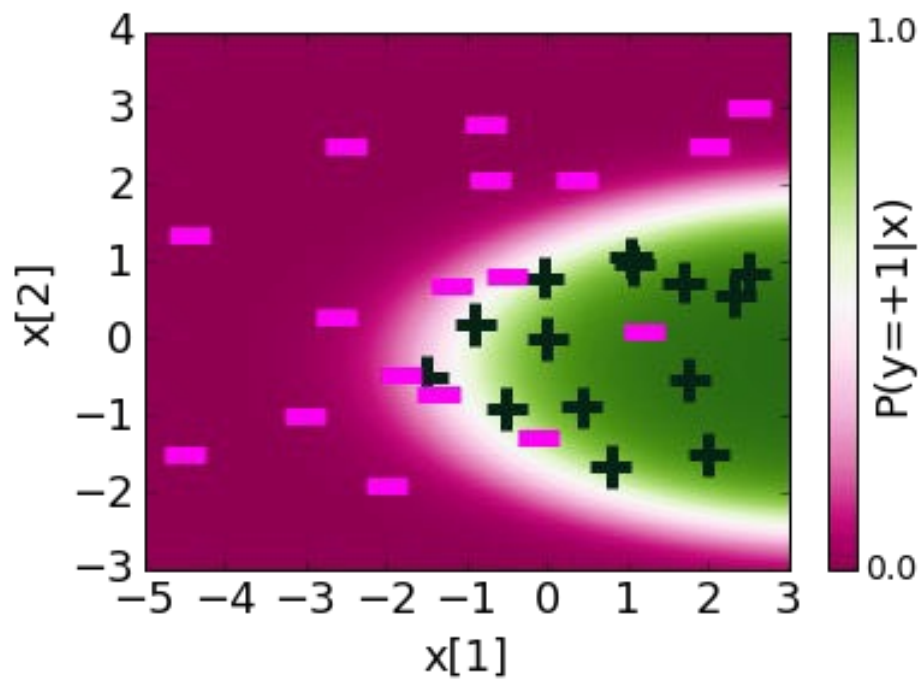


# 逻辑回归中的过拟合

## □ 学到的概率（二次特征）

特征	值	学到的系数
$h_0(x)$	1	1.68
$h_1(x)$	$x[1]$	1.39
$h_2(x)$	$x[2]$	-0.58
$h_3(x)$	$(x[1])^2$	-0.17
$h_4(x)$	$(x[2])^2$	-0.96

$$P(y = +1 | x, w) = \frac{1}{1 + e^{-w^T h(x)}}$$

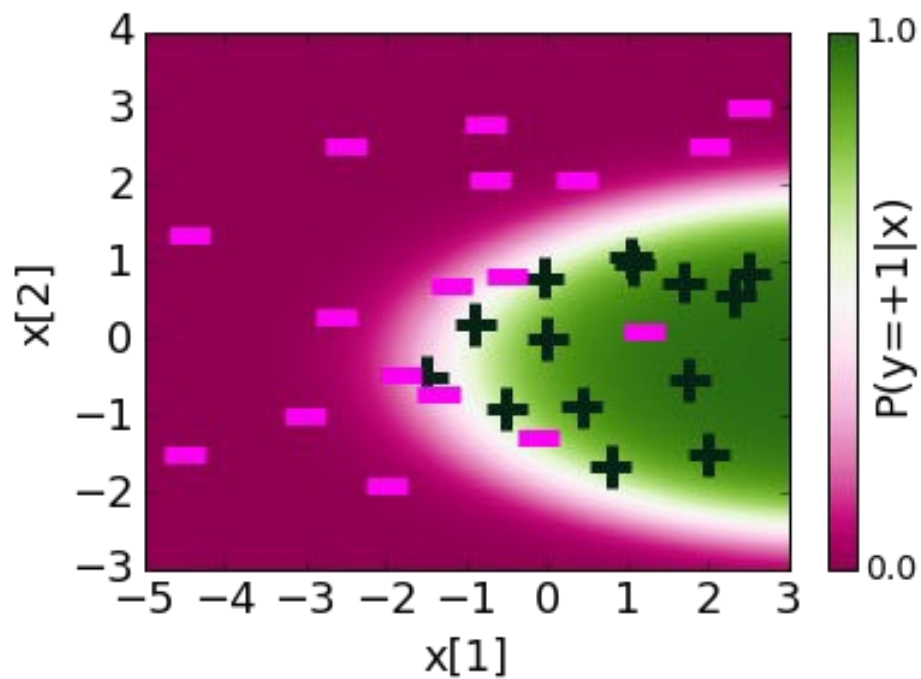


# 逻辑回归中的过拟合

## □ 学到的概率（二次特征）

特征	值	学到的系数
$h_0(x)$	1	1.68
$h_1(x)$	$x[1]$	1.39
$h_2(x)$	$x[2]$	-0.58
$h_3(x)$	$(x[1])^2$	-0.17
$h_4(x)$	$(x[2])^2$	-0.96

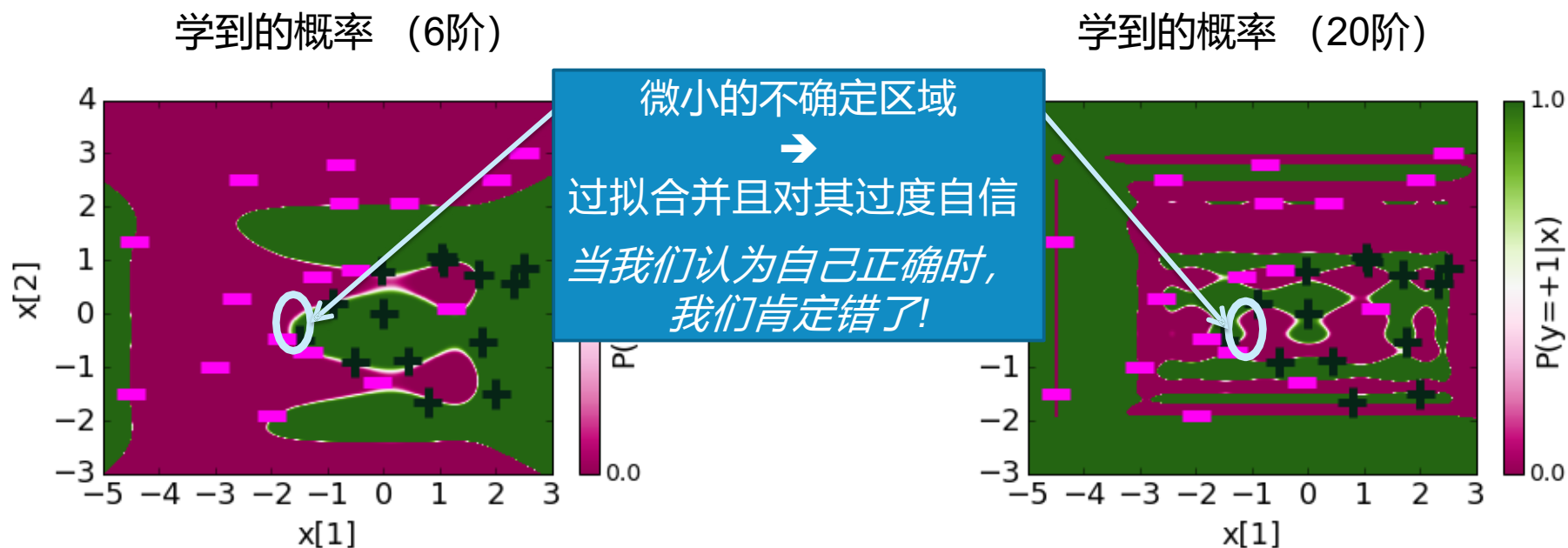
$$P(y = +1 | x, w) = \frac{1}{1 + e^{-w^T h(x)}}$$





# 逻辑回归中的过拟合

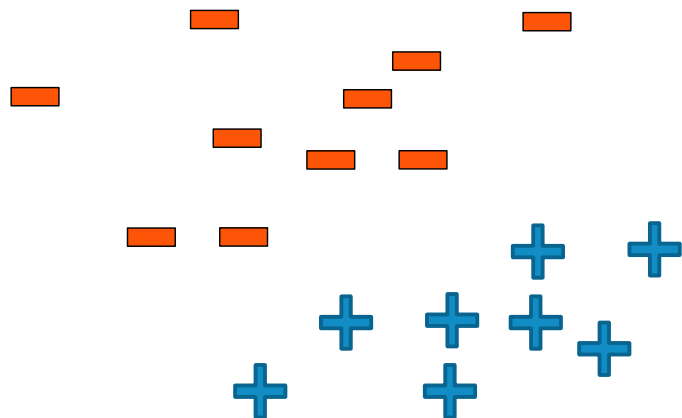
## □ 过拟合：对于预测过分自信



# 逻辑回归中的过拟合： 另一个视角

# 逻辑回归中的过拟合

## □ 线性可分数据



注 1: 如果使用  $D$  个特征, 线性可分性发生在  $D$  维空间中

注 2: 如果有足够多的特征, 数据 (几乎) 总是线性可分的

如果满足以下条件, 则数据是线性可分的:

- 存在系数  $\hat{w}$  使得:
  - 对于训练数据中所有正样本:

$$Score(x) = \hat{w}^T h(x) > 0$$

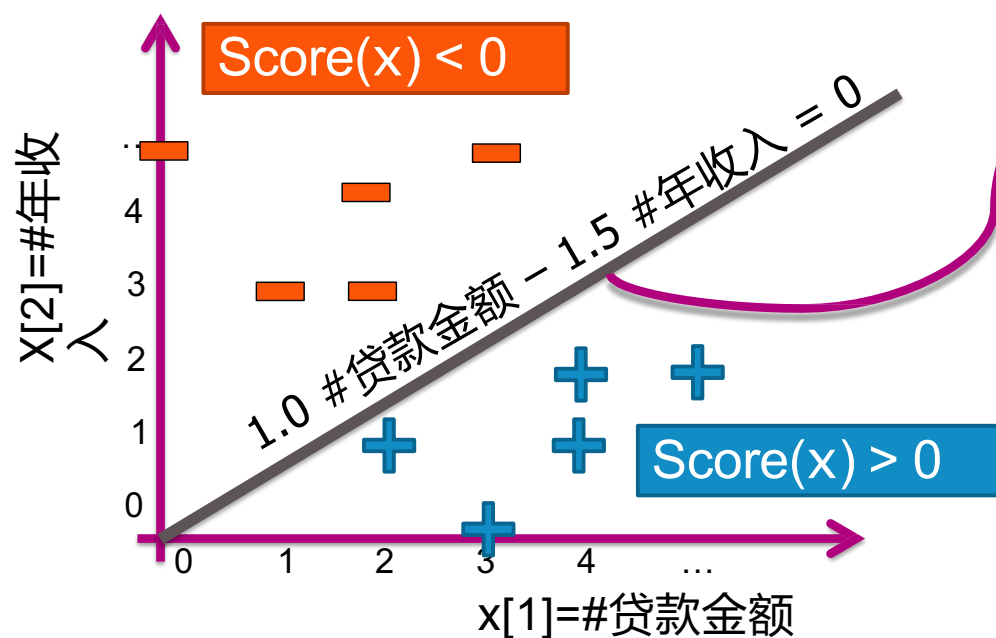
- 对于训练数据中所有负样本:

$$Score(x) = \hat{w}^T h(x) < 0$$

$$\text{training\_error}(\hat{w}) = 0$$

# 逻辑回归中的过拟合

## □ 线性可分性对系数的影响



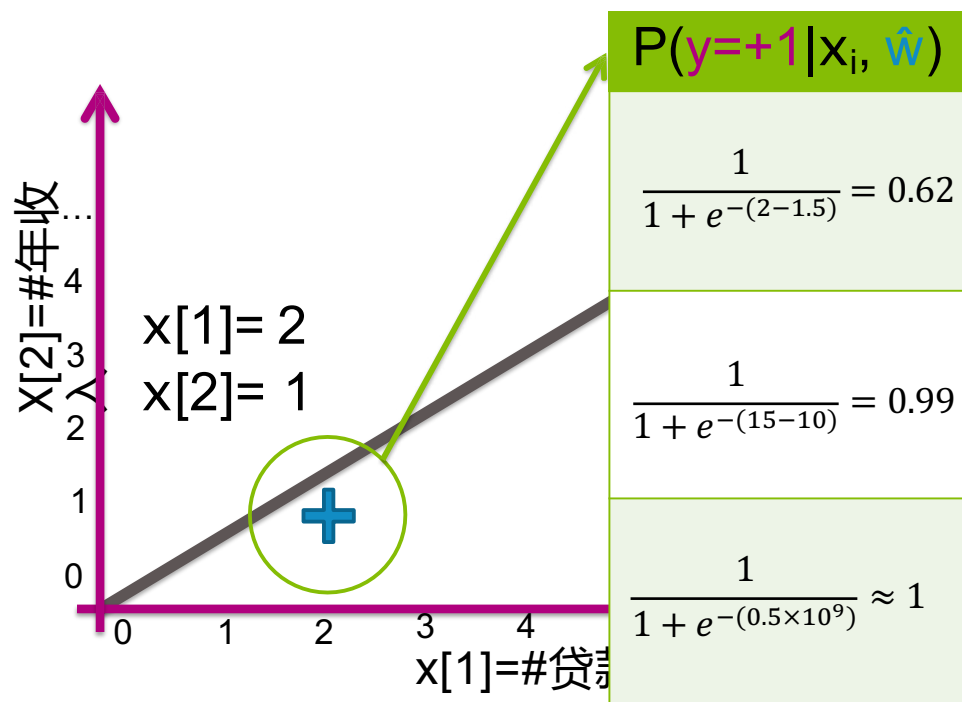
数据对  $\hat{w}_1=1.0$ ,  $\hat{w}_2=-1.5$  线性可分

数据对  $\hat{w}_1=10$ ,  $\hat{w}_2=-15$  同样线性可分

数据对  $\hat{w}_1=10^9$ ,  $\hat{w}_2=-1.5 \times 10^9$  同样线性可分

# 逻辑回归中的过拟合

最大似然估计 (MLE) 偏好特定模型→  
对于线性可分离数据，系数趋于无穷大！



数据对  $\hat{w}_1=1.0$ ,  $\hat{w}_2=-1.5$  线性可分

数据对  $\hat{w}_1=10$ ,  $\hat{w}_2=-15$  同样线性可分

数据对  $\hat{w}_1=10^9$ ,  $\hat{w}_2=-1.5 \times 10^9$  同样线性可分

# 逻辑回归中的过拟合

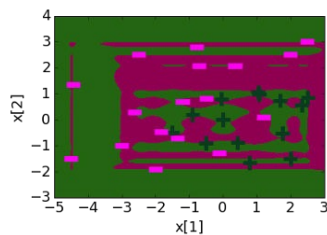
## □ 逻辑回归中过度拟合的双重后果

模型试图找到  
区分数据的决  
策边界

如果数据是  
线性可分的

过于复杂  
的边界

系数趋于无  
穷大



$$\hat{w}_1 = 10^9$$

$$\hat{w}_2 = -1.5 \times 10^9$$

**惩罚系数规模以缓解过度拟合**

## 逻辑回归中的过拟合

## □ 总损失的形式

## 我们希望平衡：

- i. 函数与数据的拟合程度
- ii. 系数的大小

总体质量 = 对于拟合的度量 - 对于系数规模的度量

我们希望平衡

↑  
(数据似然度) 大 # = 在训练数据上拟合较好

↑  
大 # = 过拟合



**惩罚系数规模以缓解过度拟合**  
L2正则的逻辑回归

# 逻辑回归中的过拟合

## □ 考虑以下优化目标

选择  $\hat{w}$  以最大化：

$$\ell(w) - \lambda ||w||_2^2$$

↖ 调整参数  $\lambda$  = 在拟合与参数规模之间平衡

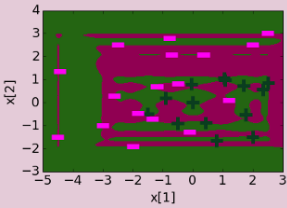
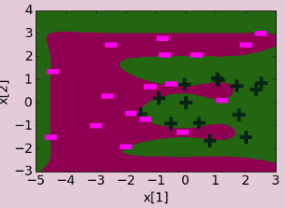
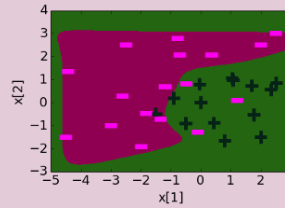
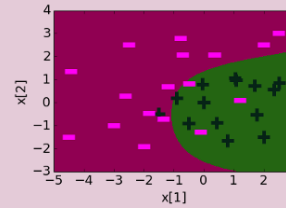
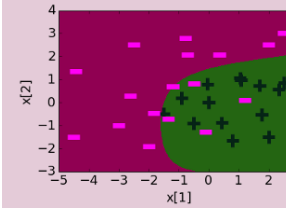
L2正则的逻辑回归

使用以下方式选择  $\lambda$ ：

- 验证集（适用于大型数据集）
- 交叉验证（适用于较小的数据集）  
（如岭/套索回归）

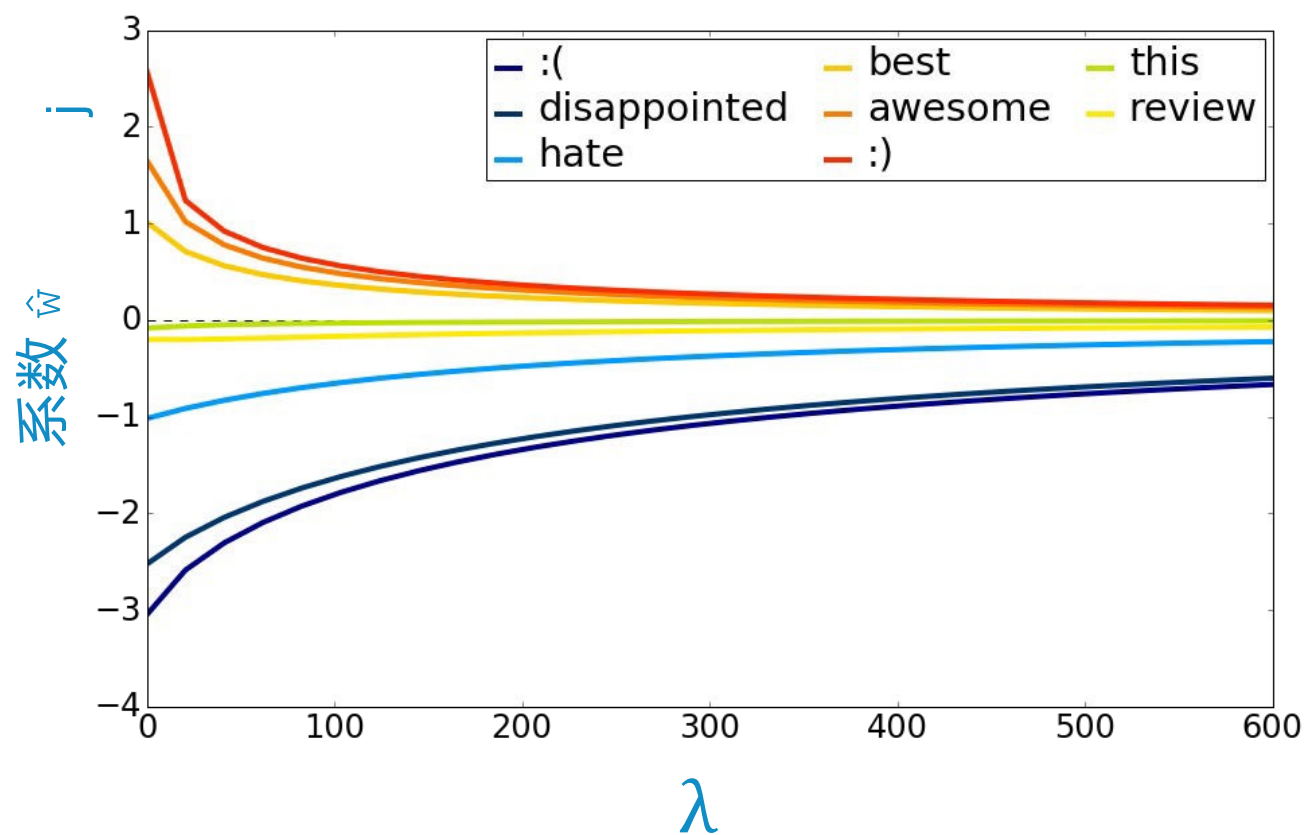
# 逻辑回归中的过拟合

## □ 20 阶特征下，正则化惩罚 $\lambda$ 的影响

正则化	$\lambda = 0$	$\lambda = 0.00001$	$\lambda = 0.001$	$\lambda = 1$	$\lambda = 10$
系数范围	-3170 ~ 3803	-8.04 ~ 12.14	-0.70 ~ 1.25	-0.13 ~ 0.57	-0.05 ~ 0.22
决策边界					

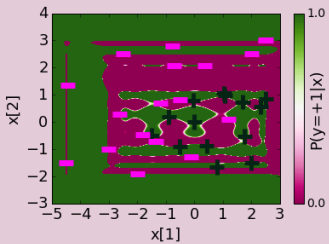
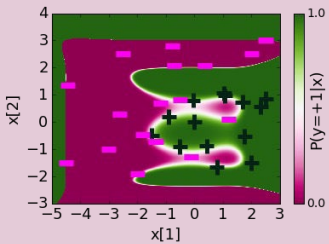
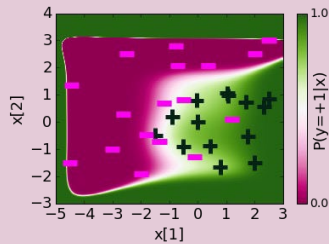
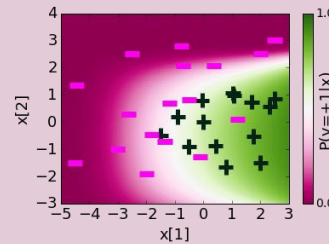
# 逻辑回归中的过拟合

## □ 系数路径 (Coefficient path)



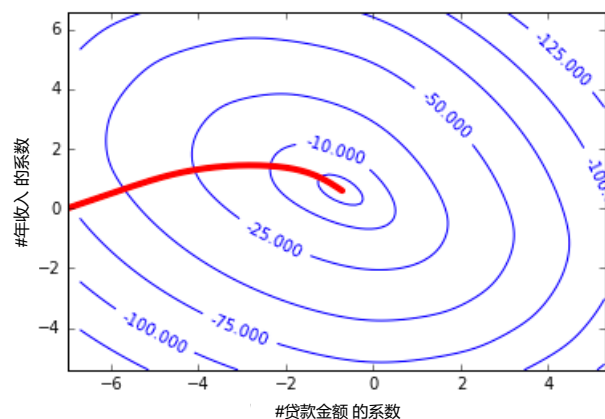
# 逻辑回归中的过拟合

□ 20 阶特征下，正则化可以降低过拟合

正则化	$\lambda = 0$	$\lambda = 0.00001$	$\lambda = 0.001$	$\lambda = 1$
系数范围	-3170 to 3803	-8.04 to 12.14	-0.70 to 1.25	-0.13 to 0.57
决策边界				

# 逻辑回归

## □ L2正则逻辑回归的梯度上升



init  $\mathbf{w}^{(1)}=0$ ,  $t=1$

**while**  $\|\nabla \ell(\mathbf{w}^{(t)})\| > \varepsilon$

**for**  $j=0,\dots,D$

$$\text{partial}[j] = \sum_{i=1}^N h_j(\mathbf{x}_i) \left( \mathbb{1}[y_i = +1] - P(y = +1 \mid \mathbf{x}_i, \mathbf{w}^{(t)}) \right)$$

$$\mathbf{w}_j^{(t+1)} \leftarrow \mathbf{w}_j^{(t)} + \eta (\text{partial}[j] - 2\lambda \mathbf{w}_j^{(t)})$$

$t \leftarrow t + 1$

**惩罚系数规模以缓解过度拟合**

L1正则的逻辑回归

# 逻辑回归中的过拟合

## □ 使用L1惩罚进行稀疏逻辑回归

选择  $\hat{w}$  以最大化：

$$\ell(w) - \lambda \|w\|_1$$

调整参数  $\lambda$  = 在拟合与参数规模之间平衡

L1正则的逻辑回归

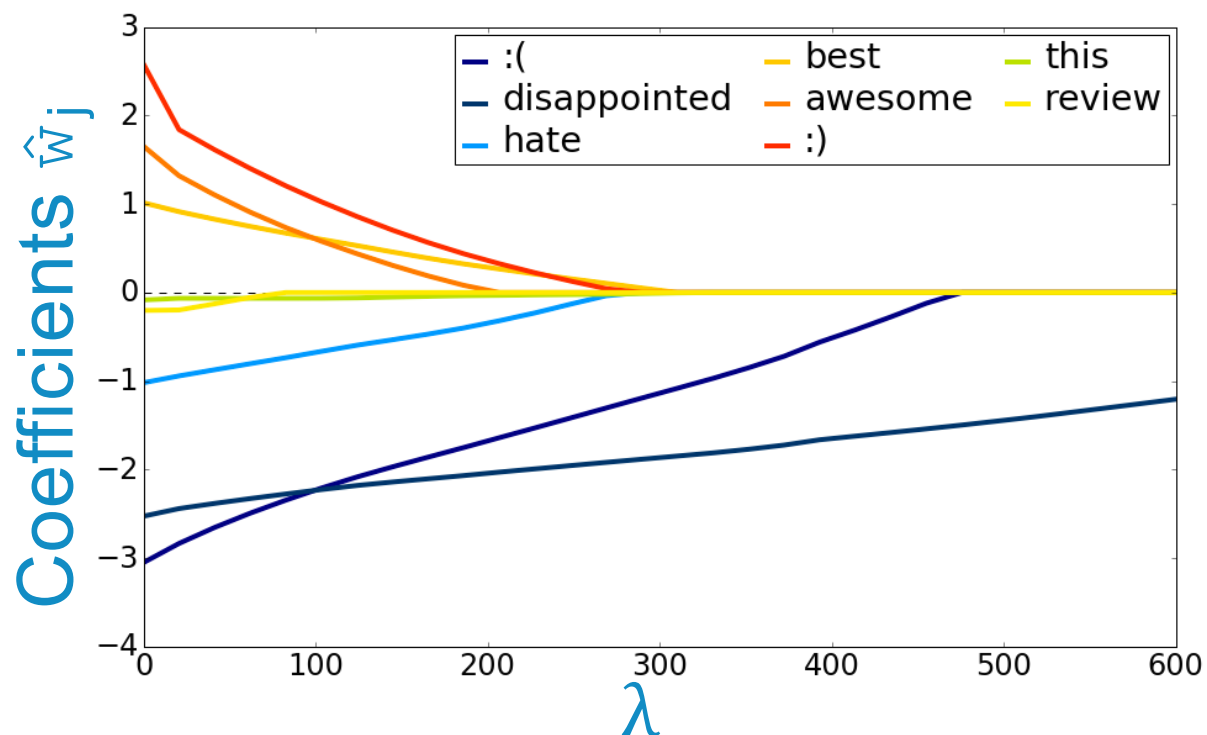
使用以下方式选择  $\lambda$ ：

- 验证集（适用于大型数据集）
- 交叉验证（适用于较小的数据集）  
（如岭/套索回归）



# 逻辑回归中的过拟合

## □ L1正则下的系数路径



# 逻辑回归中的过拟合总结

# 逻辑回归中的过拟合

- 描述分类任务中过拟合的表现与影响
  - 识别过拟合的发生时机
  - 较大的学习系数与过拟合现象
  - 分析过拟合对线性分类器决策边界及预测概率的影响
- 利用正则化缓解过拟合问题
  - 阐述L2正则化逻辑回归质量指标的构建动机
  - 采用L1正则化获得稀疏逻辑回归解
  - 分析调节参数 $\lambda$ 变化对估计系数的影响规律
  - 使用梯度上升法估计L2正则化逻辑回归系数
  - 解读系数路径图