



# 机器学习

苏州大学计算机科学与技术学院

自然语言处理实验室

主讲：周夏冰

邮箱：[zhouxiabing@suda.edu.cn](mailto:zhouxiabing@suda.edu.cn)



# 自我介绍

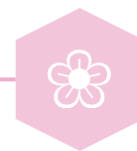
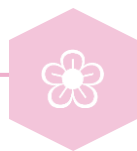
- 姓名：周夏冰
- 本科毕业于西安交通大学，博士毕业于北京大学
- 专业：计算机科学与技术
- 工作经历：中国科学院自动化研究所，苏州大学计科NLP组
- 研究方向：自然语言处理，机器学习



# 机器学习



- 机器学习是从人工智能中产生的一个重要的分支



# 课程初识

## ❖ 身边的智能



# 机器学习



- 机器学习是从人工智能中产生的一个重要的分支

1956-1960s: Logic reasoning

一切都是逻辑

符号主义学习：基于逻辑的归纳系统，一阶谓词逻辑

1970s-1980s: Knowledge engineering

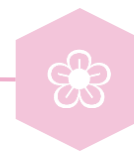
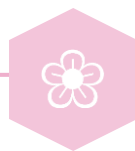
知识就是力量

统计学习：支持向量机

1990s-now: Machine learning

让它们自己学

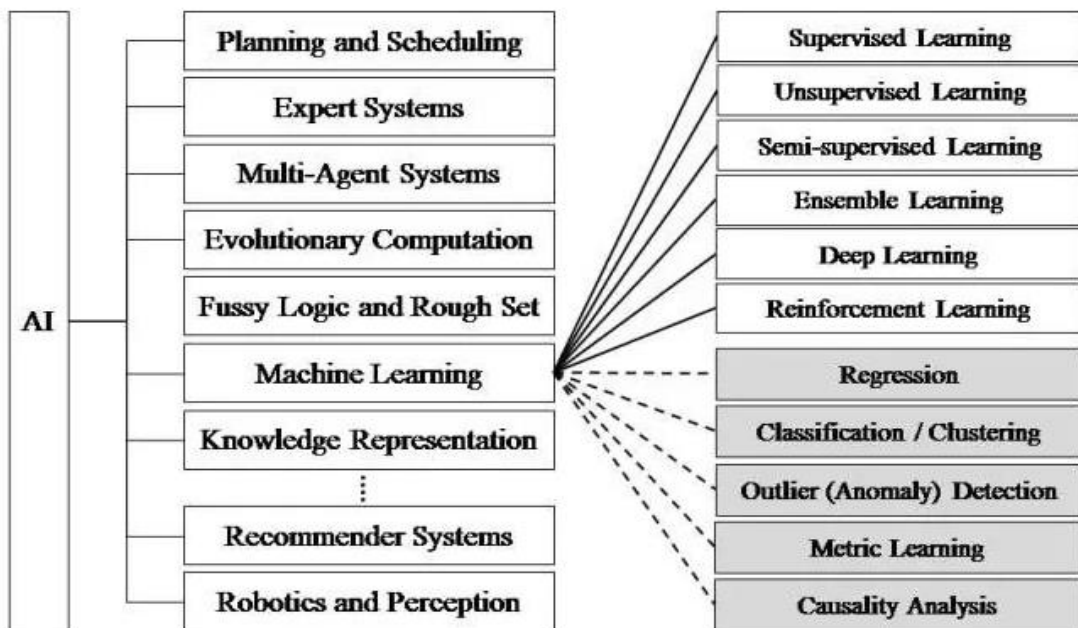
连接主义：神经网络





# 课程初识

- 机器学习(ML): 一种实现人工智能的方法



- ML致力于研究如何通过计算的手段, 利用经验来改善系统自身的性能。
- 在计算机系统中, “经验”通常以“数据”形式存在。
- ML主要研究内容: 是关于在计算机上从数据中产生“模型”(model)的算法, 即“学习算法”(learning algorithm)。
- 通过学习算法, 把经验数据提供给它, 它能基于这些数据产生模型。在面对新的情况时, 模型会给我们提供的判断。

# 机器学习技术的应用

- 循证医学(evidence-based medicine)

- 针对特定的临床问题，先要对相关研究报告进行详细评估

在一项关于婴儿和儿童残疾的研究中，美国Tufts医学中心筛选了 33,000 篇摘要

尽管专家效率很高，每篇摘要只需 30s，但该工作仍然花费了 250 小时

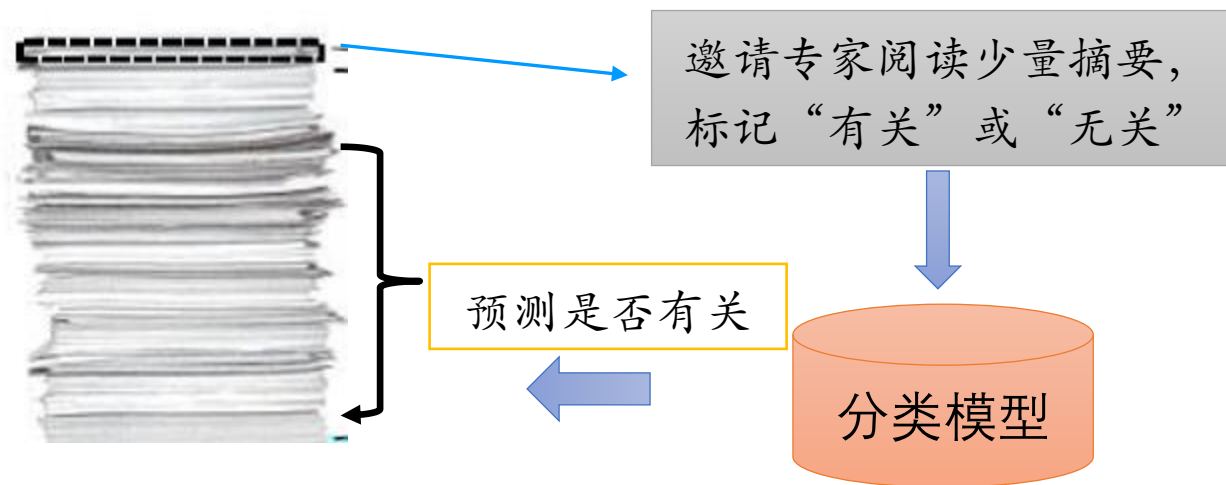


每项新研究均需要重复过程，非常麻烦



# 机器学习技术的应用

## • 循证医学(evidence-based medicine)



- 人类专家只需阅读 50 篇摘要，系统的自动筛选精度就达到93%
- 人类专家阅读1,000篇摘要，则系统的自动筛选敏感度达到95%

人类专家以前需要阅读33,000篇摘要才能获得此效果





# 机器学习技术的应用

## • 语音识别

语音输入

类别: 中文普通话

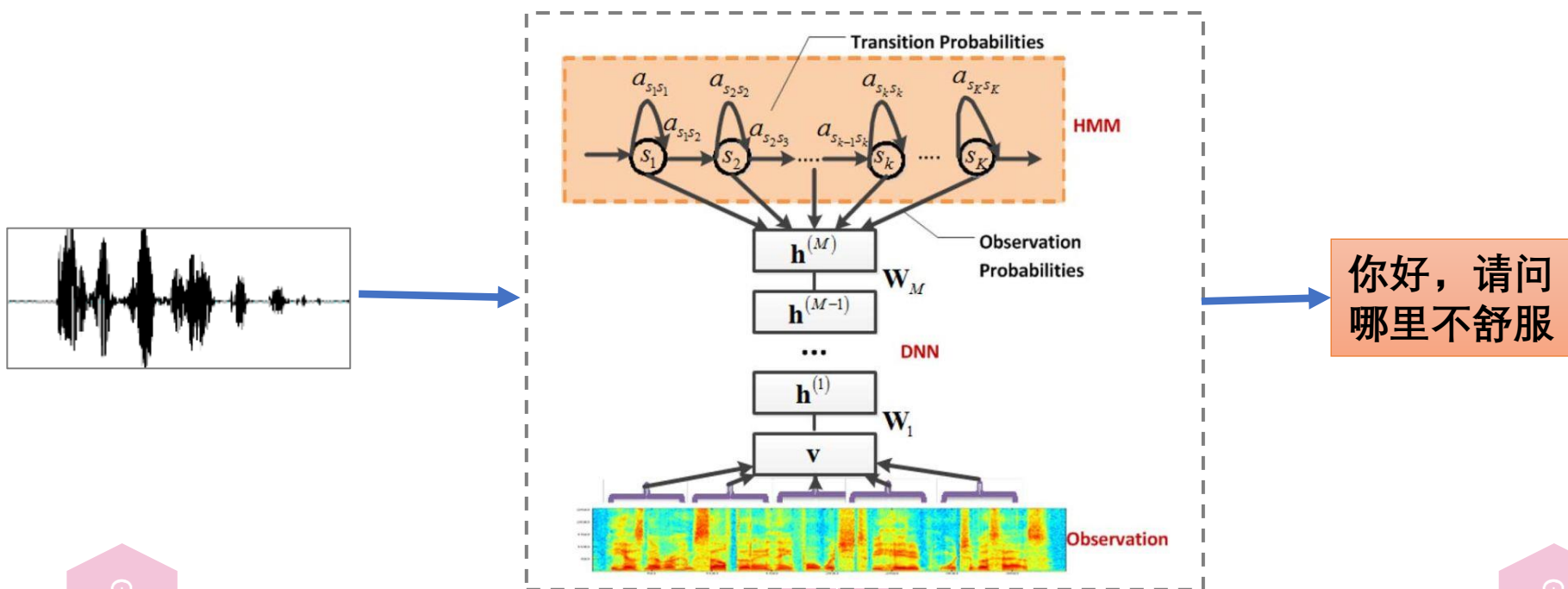


00: 17/ 01: 00

结束识别

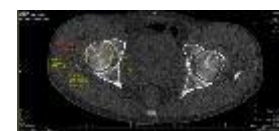
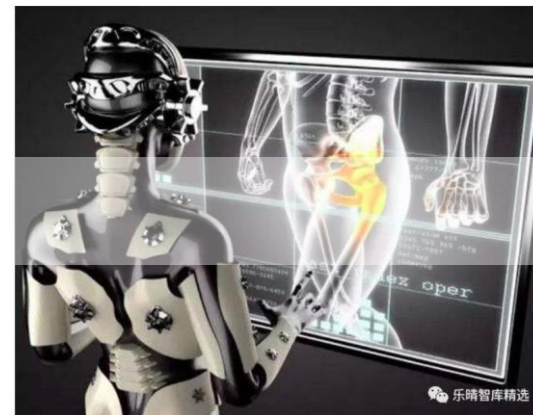
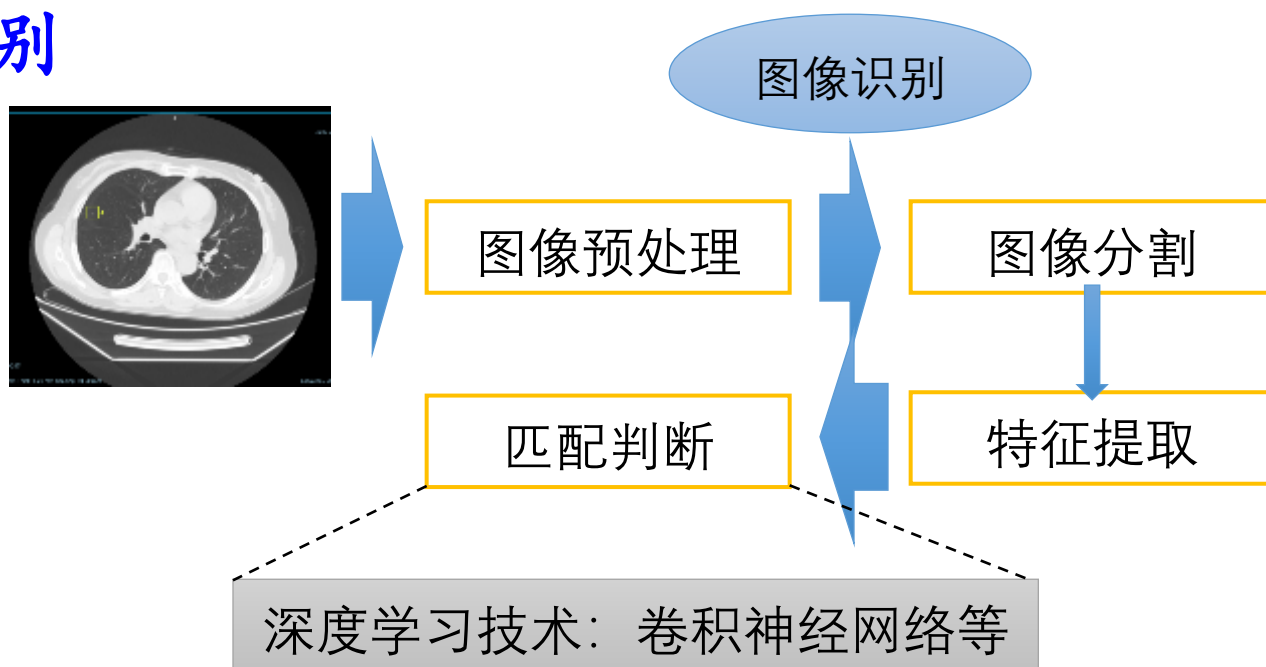
识别结果

你好人工智能今天天气好好, 天空晴朗

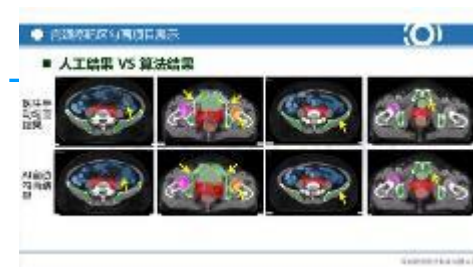


# 机器学习技术的应用

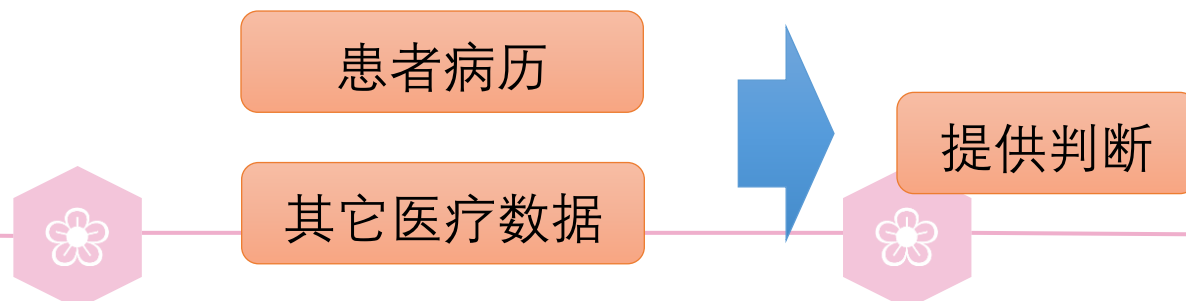
## • 图像识别



病灶标注



靶向勾画



# 机器学习技术的应用

## • 图像识别



团队	内容	AI准确率	人工准确率
斯坦福研究团队	皮肤癌诊断	至少91%	至少91%
谷歌	糖尿病视网膜病变诊断	与眼科医生一致	
上交、浙大联合科研团队	核磁共振影响的直肠癌识别	23秒完成300张影响的病灶勾画，准确率达95.22%	5分钟完成149张影像图勾画，准确率93%
香港中文大学团队	肺癌、乳腺癌影响识别	肺癌识别率达91%；乳腺癌识别率99%	
科大讯飞	肺结节智能诊断	召回率达94.1%	
卫宁健康	骨龄检测	掌指骨、腕骨、尺骨和桡骨定位准确率达98%，骨龄平均绝对差仅0.4岁	
北大第一医院	前列腺癌诊断	超过90%	





# 机器学习技术的应用

- 自然语言处理(Natural Language Processing)

- 研究能实现人与计算机之间用**自然语言**进行有效通信的各种理论和方法

词法、句法分析

命名实体识别

情感分析

知识图谱

问答、对话

机器翻译

.....





# 机器学习技术的应用

- 自然语言处理(Natural Language Processing)

词法、句法分析

患者3月前因“直肠癌”于我院全麻行直肠癌根治术，手术过程顺利

汉语分词

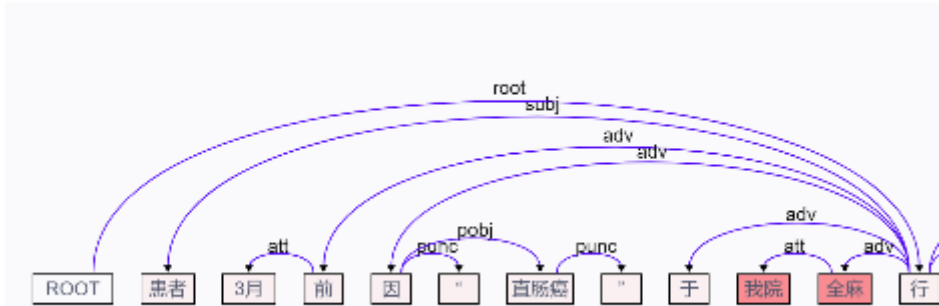


患者 3月 前 因 “ 直肠癌 ” 于 我院 全麻 行 直肠癌 根治术 ， 手术 过程 顺利

词性分析

名词	时间名词	方位词	介词	标点	名词	标点	介词	名词	副词	动词	名词	名词	标点	名词
患者	3月	前	因	“	直肠癌	”	于	我院	全麻	行	直肠癌	根治术	,	手术
名词	谓词形容词													
过程	顺利													

句法分析





# 机器学习技术的应用

## • 自然语言处理(Natural Language Processing)

很难!

词法、句法分析

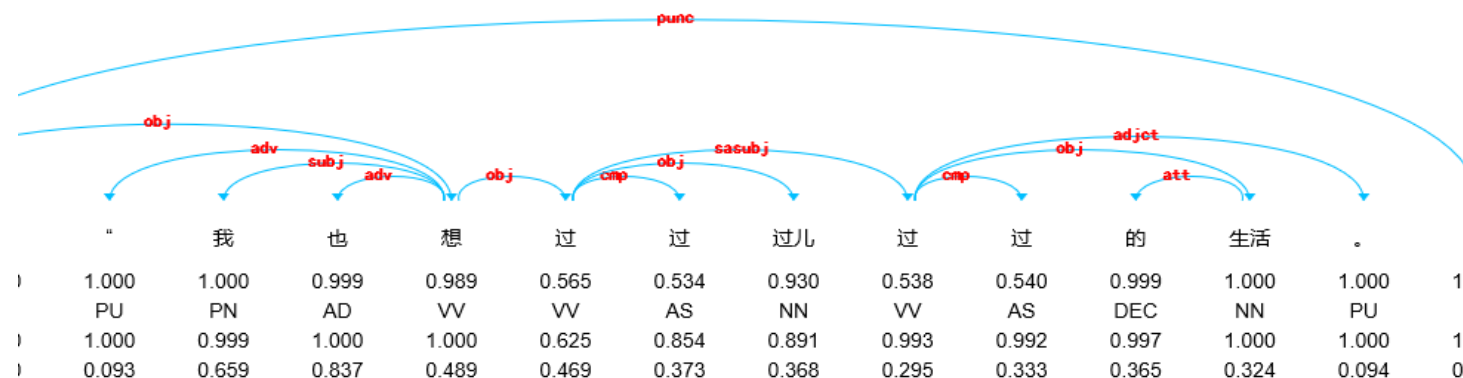
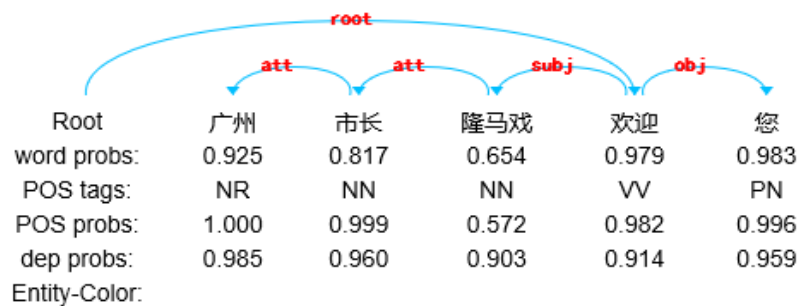



Diagram illustrating a dependency parse tree for the sentence "广州市长隆马戏欢迎您" (Welcome to the Guangzhou City Mayor's Circus.). The tree shows dependencies between words and their grammatical functions (e.g., subject, object, modifier).

	Root	广州	市长	隆马戏	欢迎	您
word probs:		0.925	0.817	0.654	0.979	0.983
POS tags:		NR	NN	NN	VV	PN
POS probs:		1.000	0.999	0.572	0.982	0.996
dep probs:		0.985	0.960	0.903	0.914	0.959
Entity-Color:						

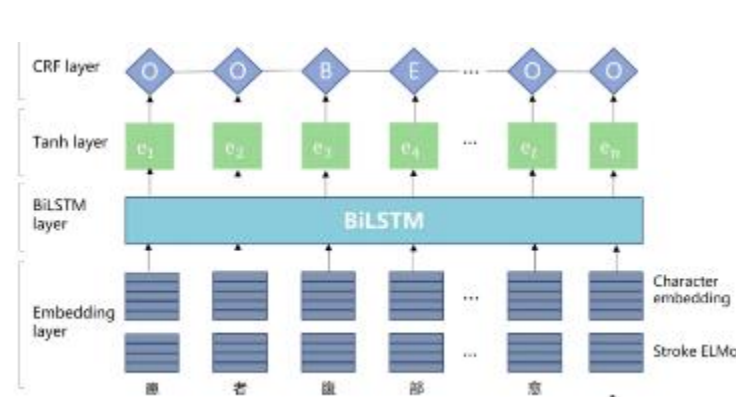
# 机器学习技术的应用

## • 自然语言处理(Natural Language Processing)

### 命名实体识别

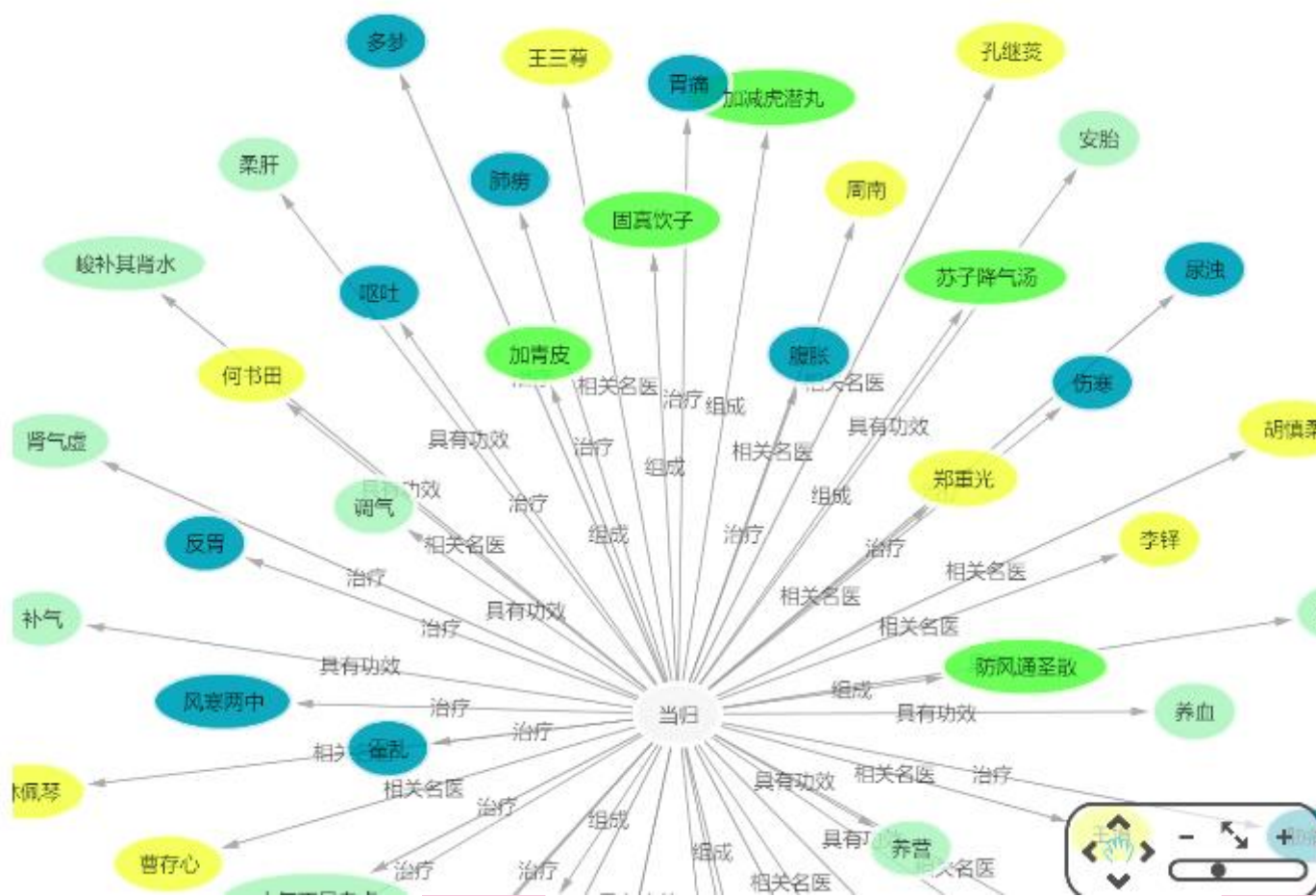
患	者	3	月	余	前	于	我	院	诊	断	为	"	直	肠	癌	"
O	O	O	O	O	O	O	O	O	O	O	O	O	B-DIS	M-DIS	E-DIS	O

- 疾病和诊断（**直肠癌、直肠腺癌（中低度分化），浸润溃疡型**）
- 影像检查（**腹部B超、CT**）
- 实验室检验（**AFP、白细胞计数**）
- 手术（**胃癌根治术、根治性全胃切除术（食管空腔ROUX-Y吻合、D2+、R0）**）
- 药物（**奥沙利铂、卡培他滨片**）
- 解剖部位（**腹腔、肠壁二站（2个）淋巴**



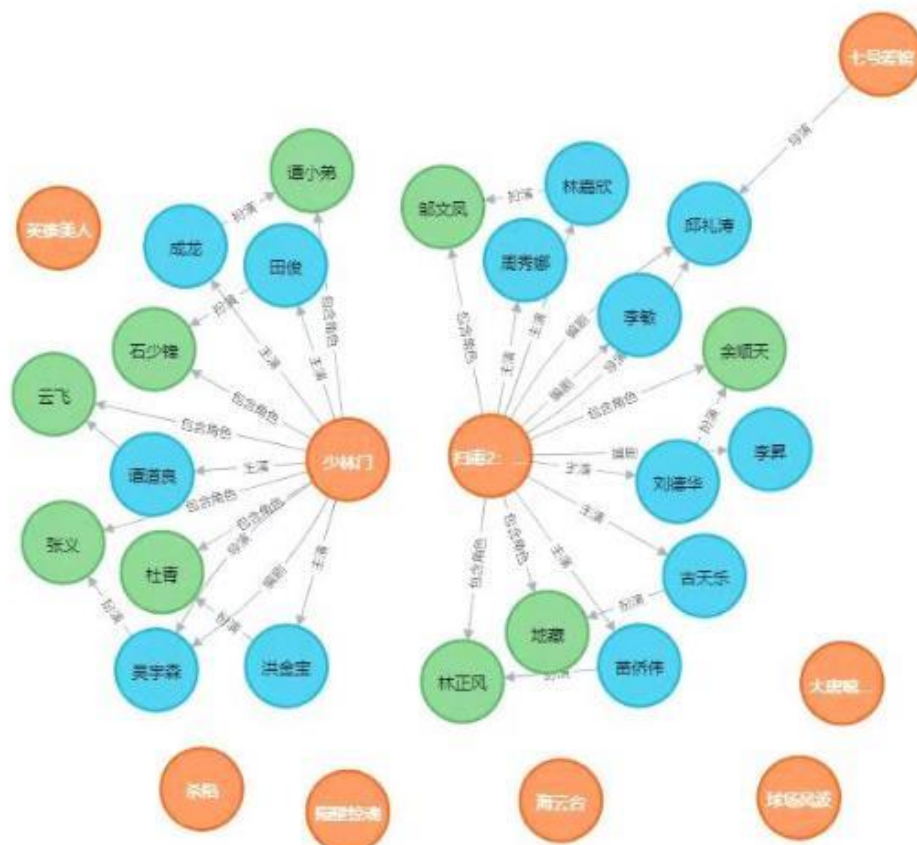
- 自然语言处理(Natural Language Processing)

## 知识图谱



- 自然语言处理(Natural Language Processing)

# 知识图谱



# 机器学习技术的应用

## • 自然语言处理(Natural Language Processing)

### 问答、对话

“Q”: “当归有什么功效?”

“A”: “调气, 补气, 柔肝, 养血”

“Q”: “重实是指什么?”

“A”: “所谓重实, 如大热病人, 邪气甚热, 而脉象又盛满, 内外俱实, 便叫重实”

A: 今天下午幼儿园老师说我家汉字扁桃体肿大, 他自己说嗓子不舒服, 已经3天了, 量了体温不发烧, 吃了双黄连口服液, 请问还要吃其它什么药么?

B: 可以吃点罗红霉素

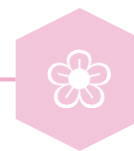
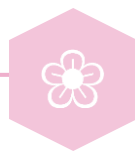
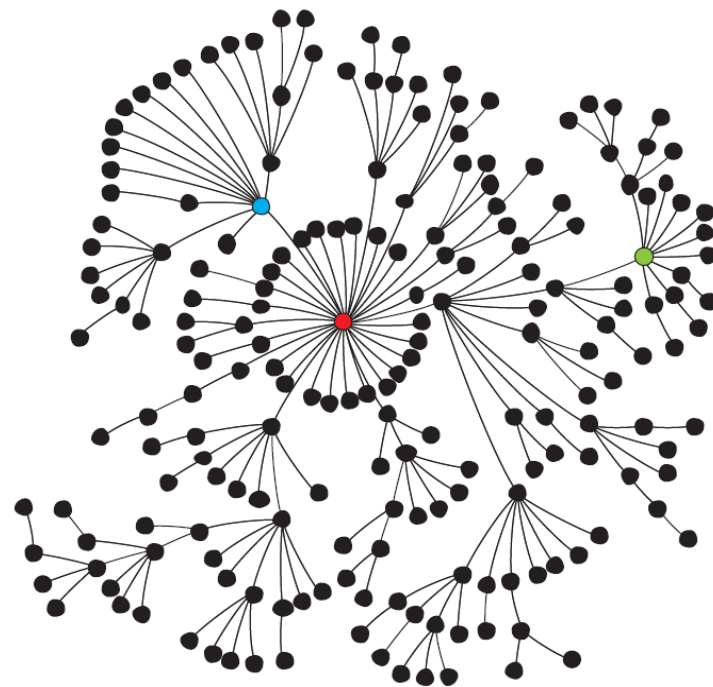
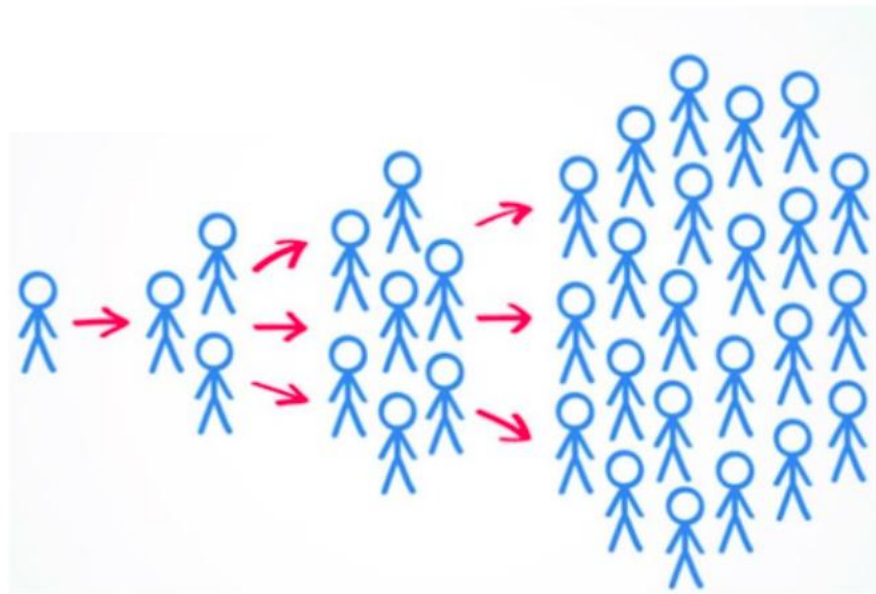
A: 罗红霉素怎么吃啊?

B: 儿童一次按体重2.5~5mg/kg, 一日2次

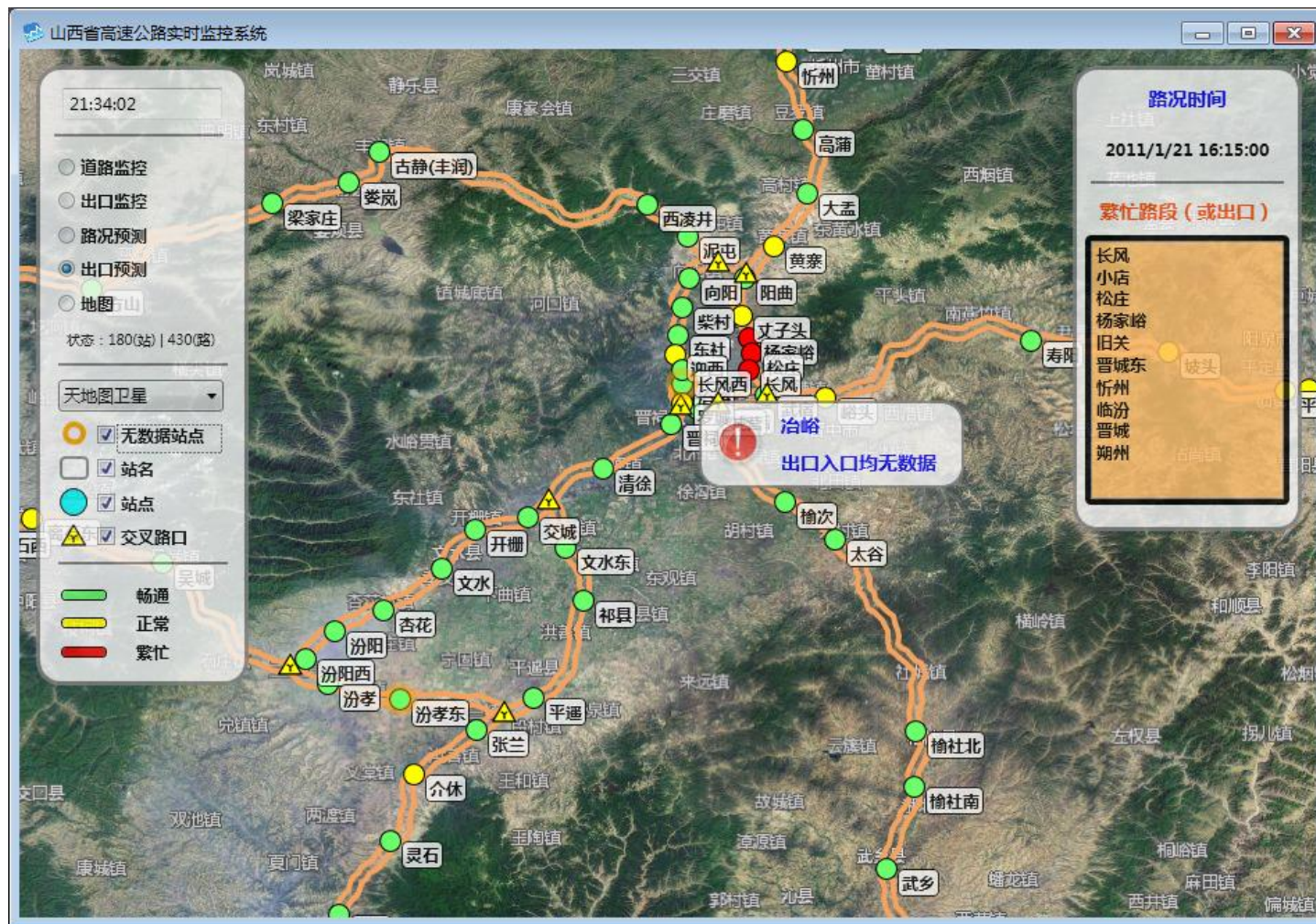
A: 好的, 谢谢



# 网络影响最大化



# 智能交通





# 情感分析

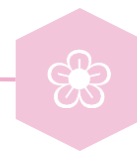
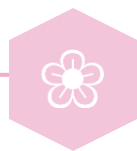
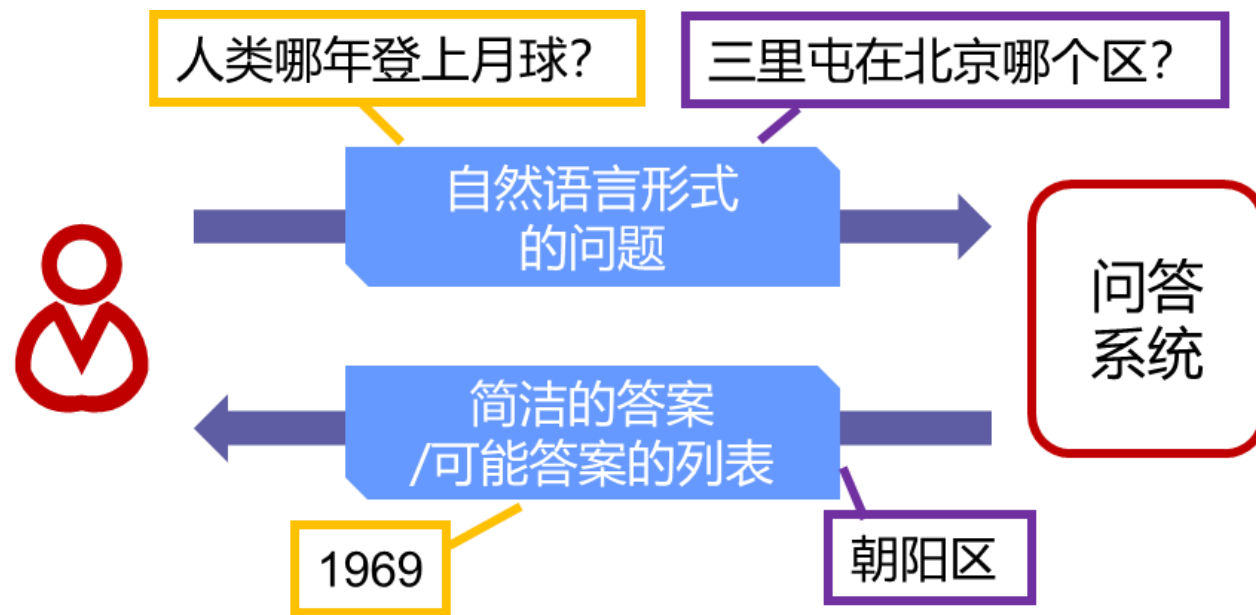


情感	微博消息示例
正面情感	1、九寨沟山美水美人更美。 2、九寨沟的海别有景致，五彩斑斓，流连忘返。只有亲临其境，才能感受这种美妙！
中性情感	1、今天要去九寨沟。 2、云南、鼓浪屿、西藏、海南、黄山、九寨沟。选一地方五一走起来
负面情感	1、看天气预报，九寨沟那边好冷哦！这可就麻烦了，要带什么衣服呢？ 2、有人说九寨沟不适合我.....是因为我看起来不能够修身养性么.....

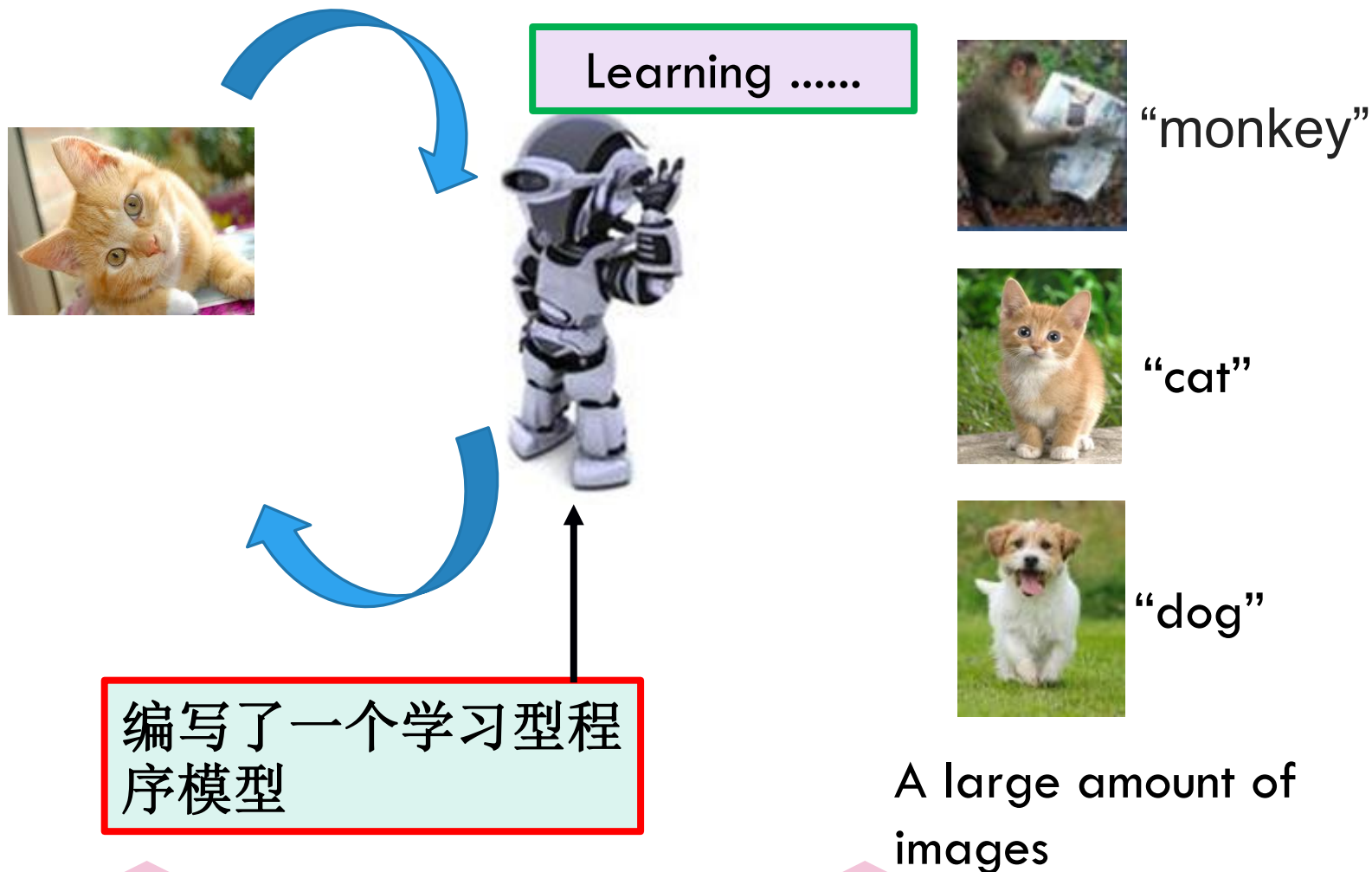
情绪	微博
喜	周末见到了高中同学，她一点都没变，还是那么美丽。就连说话的语调都一如当年。
怒	请不要等我变了以后，才说怀恋从前的我好吗？
哀	真该下定决心减肥了，我自己都闲胖了，更别说别人了！
惊	大早上刚进餐厅就看见桌上竖俩鸡蛋.....这年头，没点技术活你都不好意思占座！
恐	周一领导约我谈话，是我的春天来了还是冬天到了.....



# 问答系统



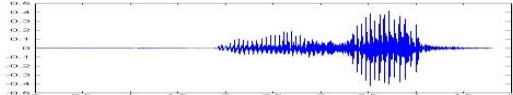
# 机器学习——初识





# 了解机器学习


- **Speech Recognition**

$$f(\text{  }) = \text{"How are you"}$$

- **Image Recognition**

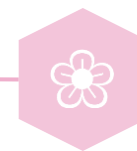
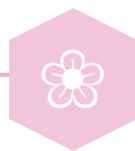
$$f(\text{  }) = \text{"Cat"}$$

- **Playing Go**

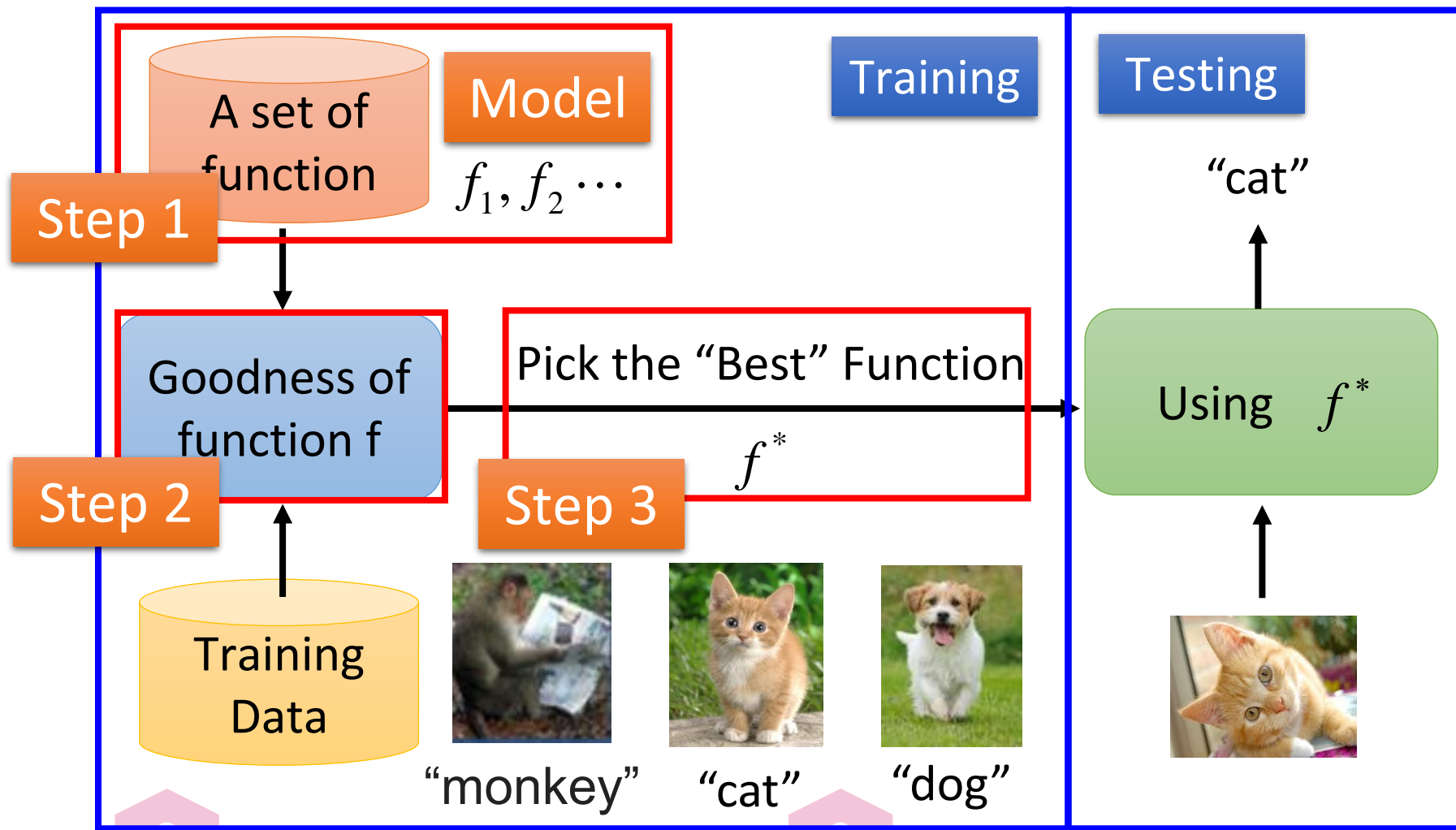
$$f(\text{  }) = \text{"5-5"}_{\text{(next move)}}$$

- **Dialogue System**

$$f(\text{"Hi"}_{\text{(what the user said)}}) = \text{"Hello"}_{\text{(system response)}}$$

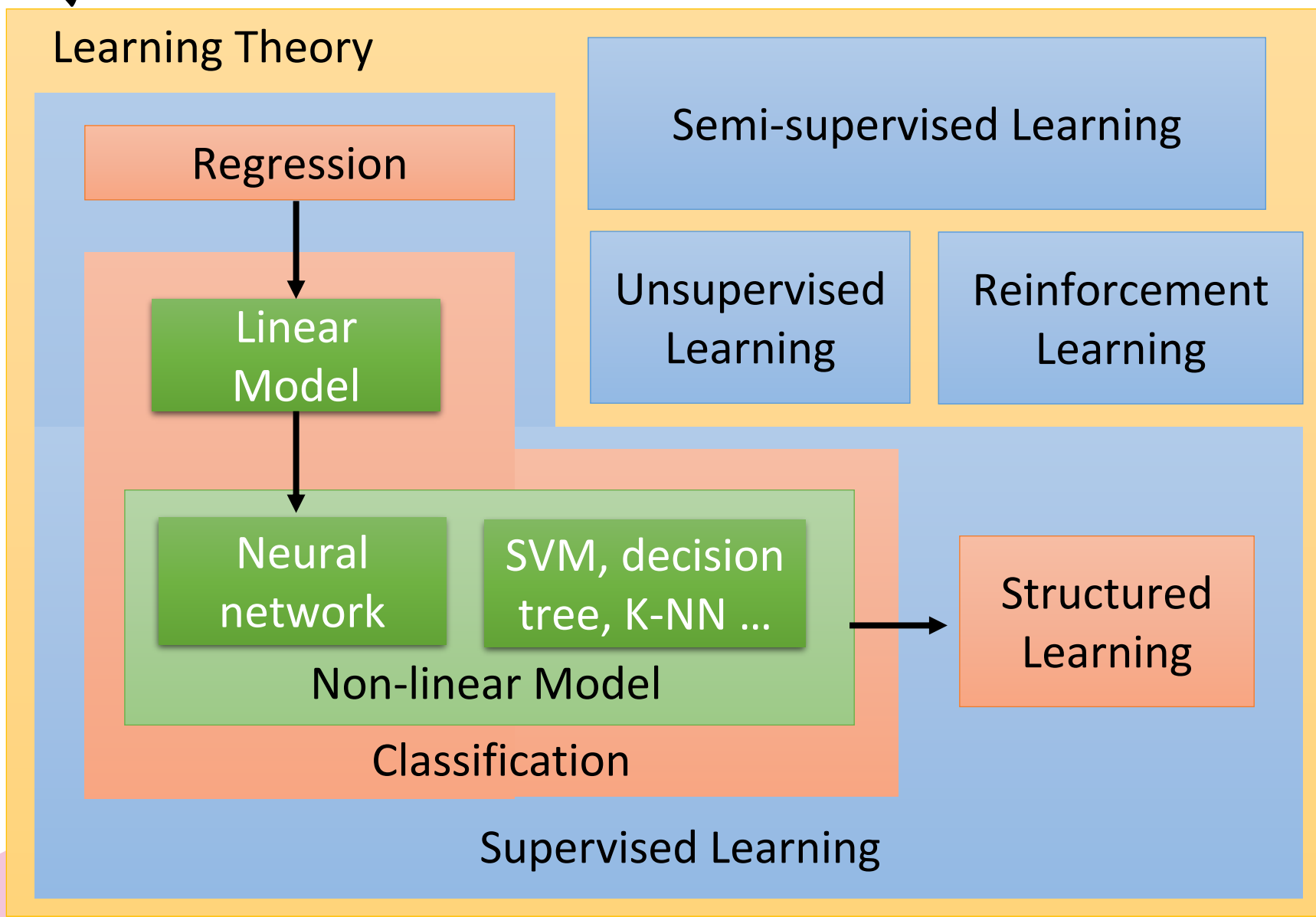


# 机器学习——初识

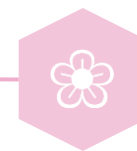
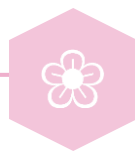
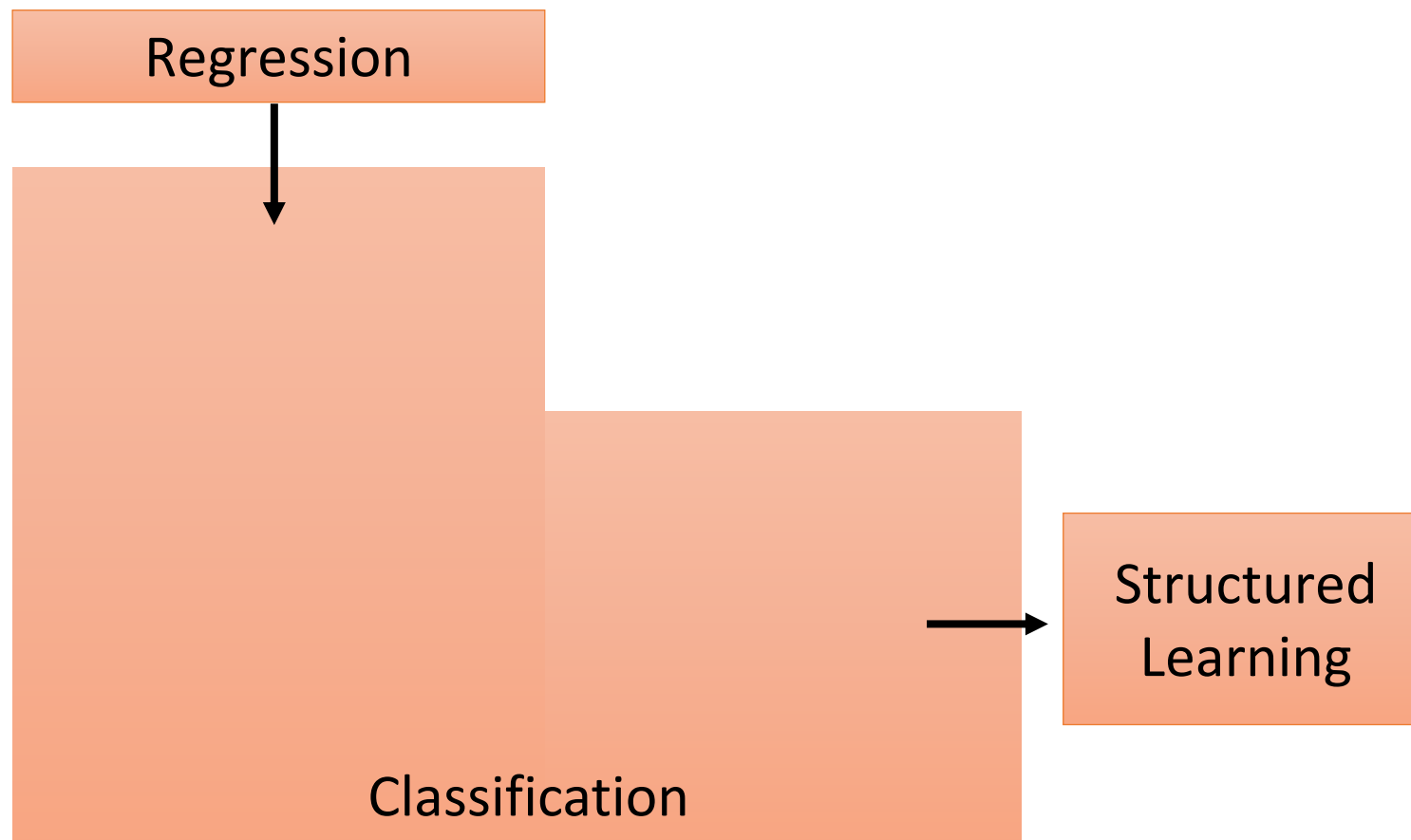


# 机器学习

scenario task method



# 了解机器学习



# 了解机器学习

## • 回归

Regression

The output of the target function  $f$  is “scalar”.

预测  
PM2.5

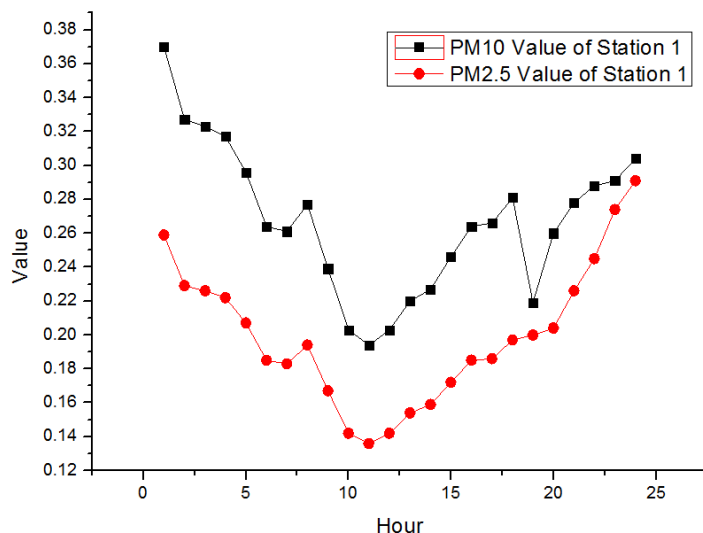
今天上午 PM2.5  
昨天上午 PM2.5  
.....



明天上午 PM2.5  
(scalar)

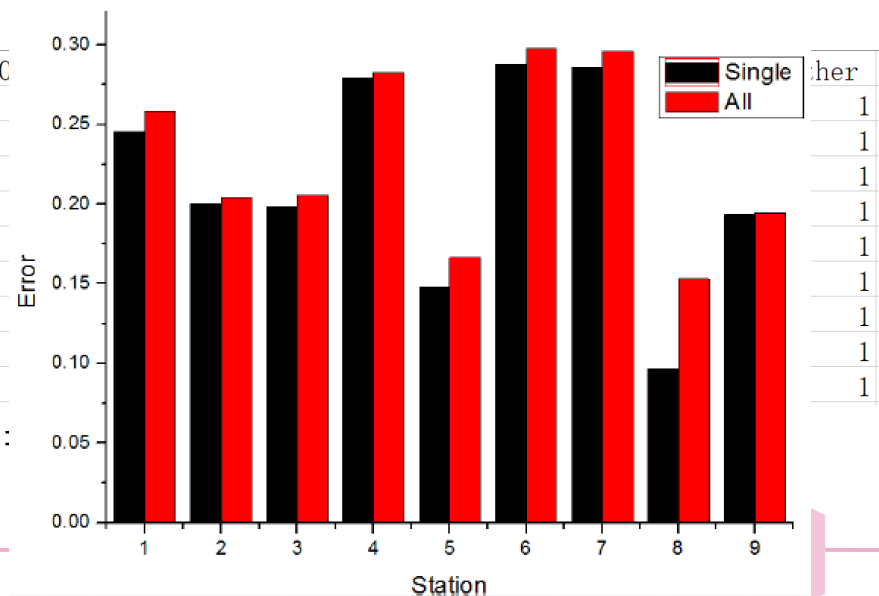
Train

station index
100
100
100
100
100
100
100
100
100
100
100



_value	NC
59	
65	
73	
73	
73	
73	
73	
73	
73	
73	

PM2.5 :

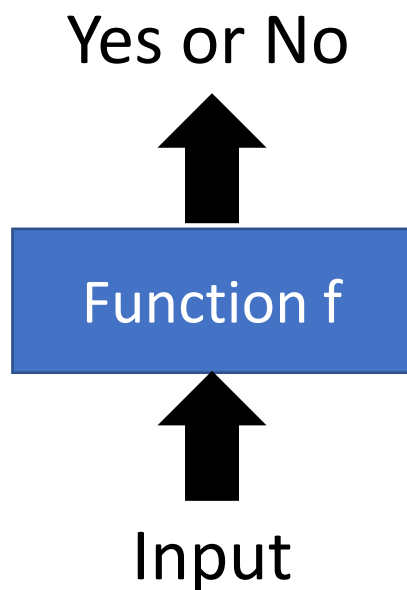




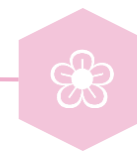
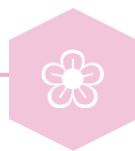
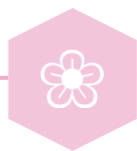
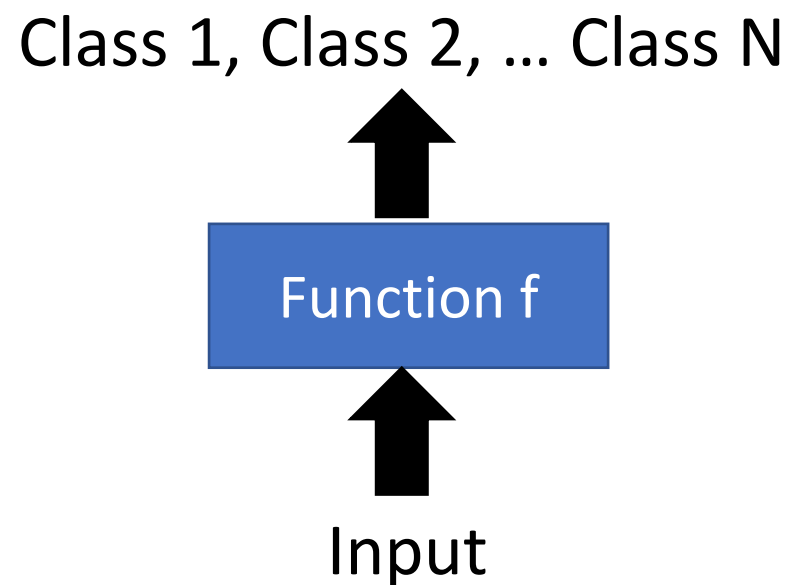
# 了解机器学习

- 分类

- **二分类**



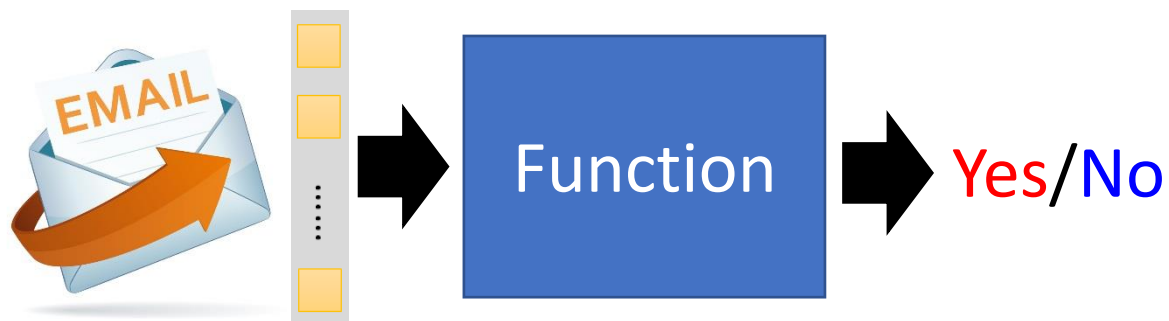
- **多分类**



# 了解机器学习

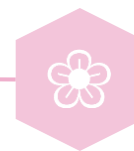
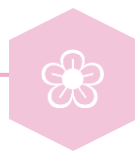
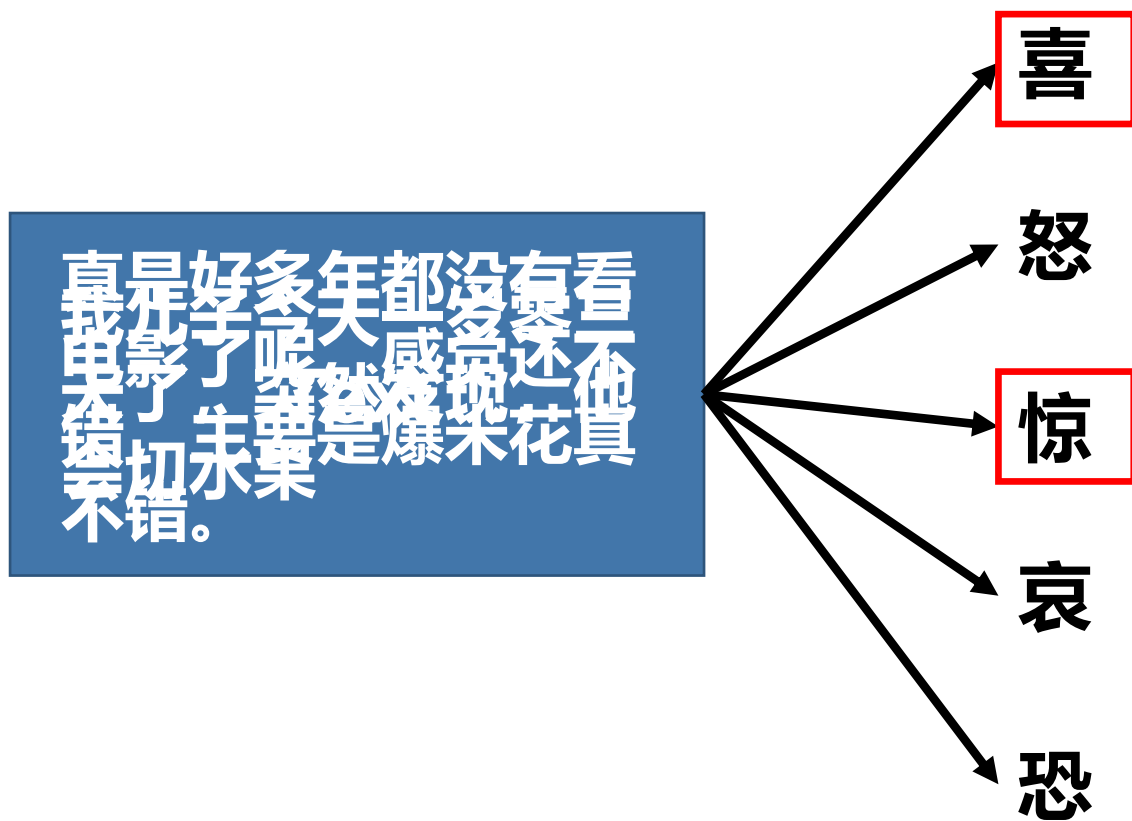
- 二分类

Spam  
filtering



# 了解机器学习

- 多分类



# Structured Learning

## - Beyond Classification

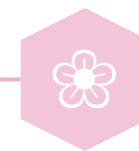
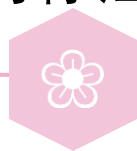


Speech Recognition

输入：At Microsoft Research, we have an insatiable curiosity and the desire to create new technology that will help define the computing experience.

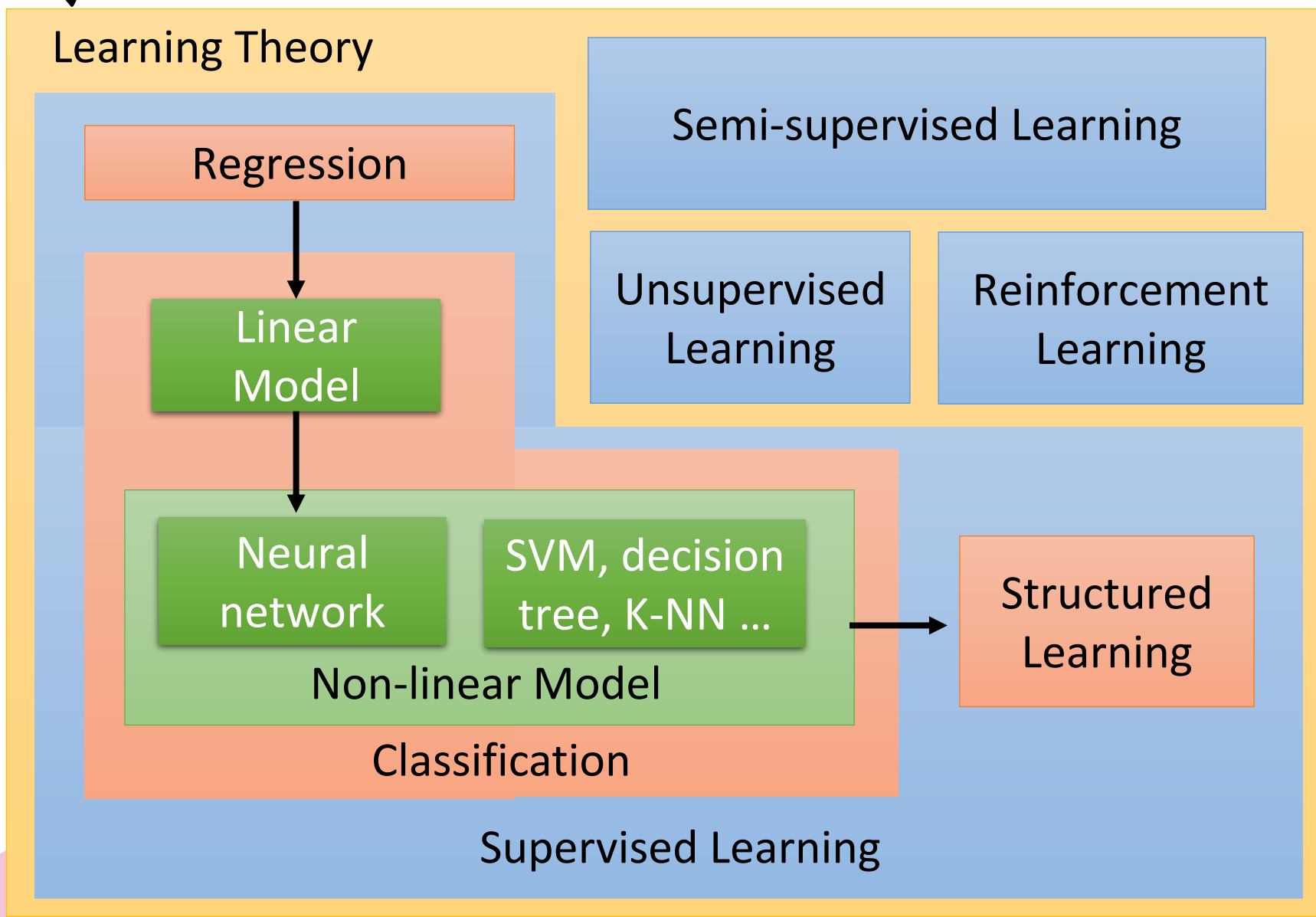
输出：At/**O** Microsoft/**B** Research/**E**, we/**O** have/**O** an/**O** insatiable/**B** curiosity/**E** and/**O** the/**O** desire/**BE** to/**O** create/**O** new/**B** technology/**E** that/**O** will/**O** help/**O** define/**O** the/**O** computing/**B** experience/**E**.

序列标注



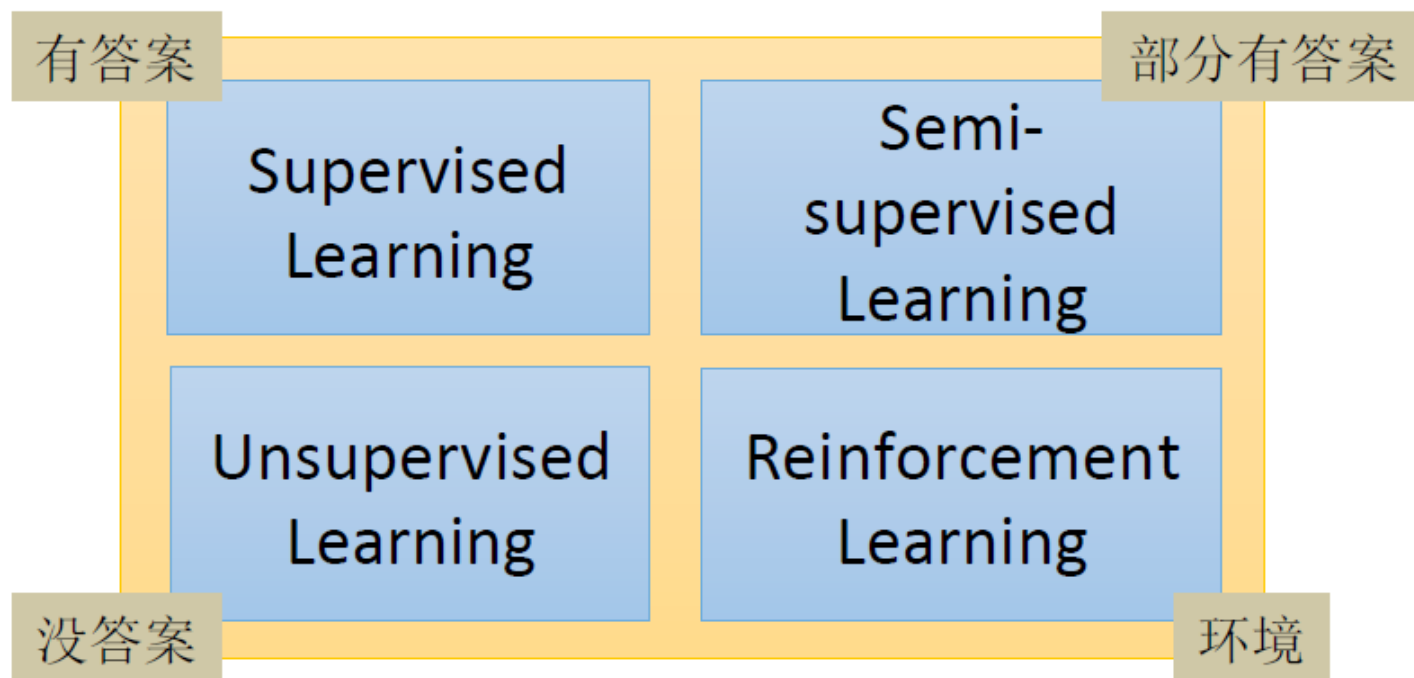
# 机器学习

scenario task method





# 机器学习的基本模式





# Semi-supervised Learning 半监督学习

## ➤ 训练集中的部分示例没有标记

Labelled  
data



cat



dog

Unlabeled  
data



(Images of cats and dogs)

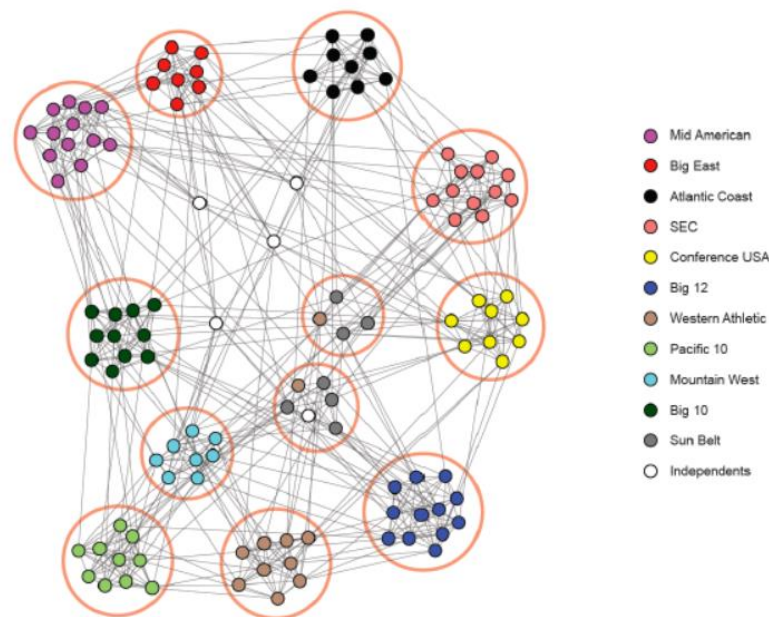


# Unsupervised Learning



## ➤ 训练集不含任何标记

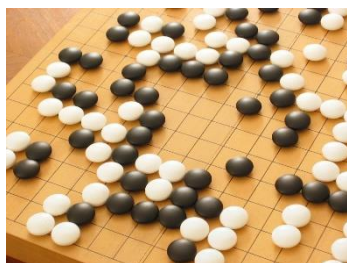
- **Machine Reading:** 机器从大量的文本中学习出单词的意思
- **Community detection:** 从大规模社交网络提取社团结构



# Supervised v.s. Reinforcement

## ➤ 与环境的交互中学习

- Supervised:



Next move:  
"5-5"



Next move:  
"3-3"

- Reinforcement Learning

First move ➡ ..... many moves ..... ➡ Win!

Alpha Go is supervised learning + reinforcement learning.







# 关于课程

- 参考书籍：

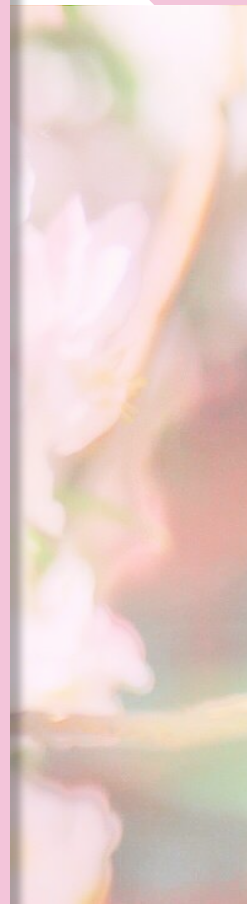
- 李航《机器学习方法》，周志华《机器学习》，邱锡鹏《神经网络与深度学习》

- 课程考核

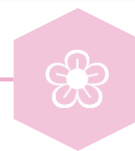
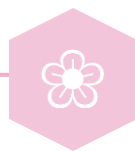
- 考勤5%
- 平时作业（实验）25%
- 期中考试20%
- 期末考试50%

◆ 学习通  
课程邀请码：  
91575025





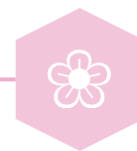
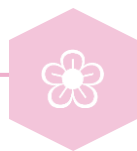
# 基本概念





# 基本概念

- $X = (x_1, x_2, \dots, x_i, \dots, x_N)^T$  实例 (Instance)
- $Y = (y_1, y_2, \dots, y_i, \dots, y_N)^T$  样本 (Sample)
- $x_i = (x_i^1, x_i^2, \dots, x_i^n)$  特征向量 (feature vector)
- $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$  数据集 (dataset)





# 基本概念

$$y = wx + b$$
$$y = w^T x + b$$

视频收益

特征数据

## • 评估一个视频的收益



【盐】一百多年不曾停歇的盐井，仍在续写名为“味道”的故事  
352.3万 2021-7-14



小时候睡的旧木床，蜕变成古色古香的流觞曲水桌  
306.4万 2021-6-5



深深扎根在中华土地上，历久弥新的茶文化  
278.1万 2021-5-27



【绢花】捣鼓了一身看花的衣裳，和我家老小孩一起拥抱春天！  
312.1万 2021-4-6



【春节年货】挂灯笼、贴对联、备好年货过大年啦  
244.2万 2021-2-12



【萝卜的一生】带着年味的萝卜，落下了“一生系列”的帷幕  
395.9万 2021-1-9



愿一串串的红柿子给大家带来新年的柿柿顺心  
278.8万 2020-12-9



芋头的一生，辣椒...是送的  
375.2万 2020-11-8



【丰收季】一年忙忙碌碌，为的就是谷物满仓的踏实感  
309.4万 2020-10-10



【西瓜和葡萄的一生】摘了西瓜，酿了葡萄酒，迎着月色一起喝一杯  
425.9万 2020-9-14



玫.....瑰花...的一生？  
314.1万 2020-8-28



【秋葵和牵牛】用秋葵“湊”了一堆菜，吃饱喝足乐颠颠  
271.8万 2020-8-13



【黄瓜的一生】黄瓜正当季，那就变着法来吃个黄瓜吧  
400.4万 2020-7-20



【番茄】泥土地里肆意地生长！小苗苗结出大果实——番茄的一生  
334.6万 2020-7-9



【祝大家端午安康】我小时候吃过的传统笋壳粽子，你那有吗？  
321万 2020-6-25

29.0万 12.0万 2.4万 8658

$$y = xw + b$$

$$y = [y_1, y_2, \dots, y_N]^T$$

N个样本

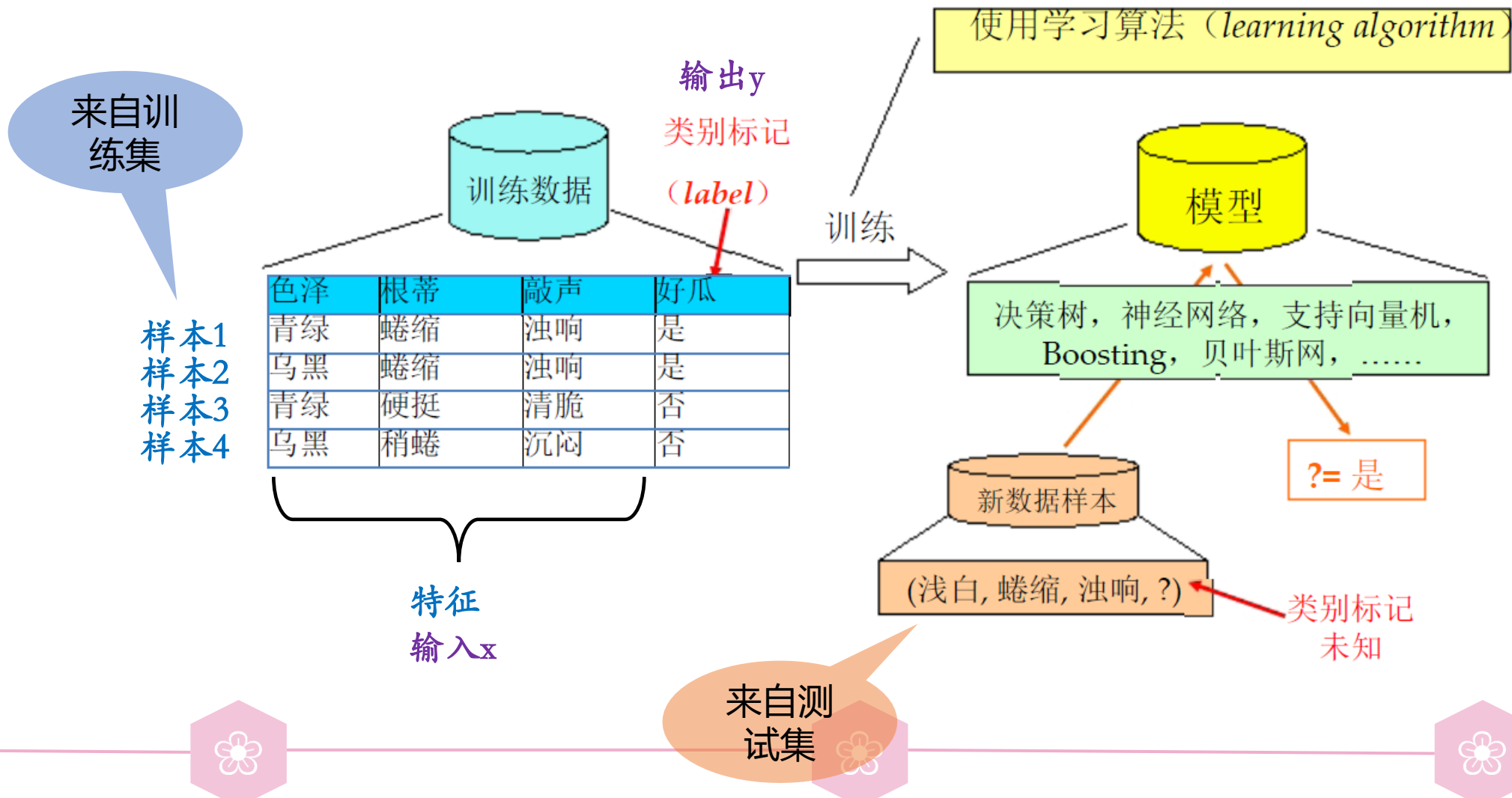
d个特征: 点赞, 转发, 投币, 收藏, 观看次数等

$$x = \begin{bmatrix} x_{11}, x_{12}, \dots, x_{1d} \\ x_{21}, x_{22}, \dots, x_{2d} \\ \dots \\ x_{N1}, x_{N2}, \dots, x_{Nd} \end{bmatrix}$$

参数:  $w = [w_1, w_2, \dots, w_d]^T$   
 $b$ 标量



# 机器学习基本概念

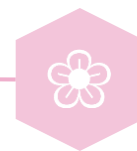
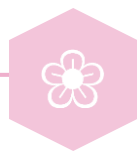






# 基本概念

- 方法 = 模型 + 策略 + 算法
- 模型：学习什么样的模型？
- 策略：按照什么准则学习或选择最优的模型？
- 算法：学习模型的具体计算方法





# 机器学习三要素

## • 模型

- 例如：如果决策函数需要是线性函数，那么所有线性函数构成的函数集合就是**模型的假设空间**
- 决策函数的集合  $F = \{f \mid Y = f(X)\}$
- 条件概率的集合  $F = \{P \mid P = f(Y \mid X)\}$

$$y = xw + b \quad F = \{y_1 = xw_1 + b_1, y_2 = xw_2 + b_2, \dots\}$$





# 机器学习三要素

## • 策略

- 损失函数：一次预测的好坏。
- 风险函数：平均意义下模型预测的好坏

- 风险函数（期望损失）  $R_{\text{exp}}(f) = E_p[L(Y, f(X))] = \int_{x \times y} L(y, f(x)) P(x, y) dx dy$

- 经验风险（经验损失）

$$R_{\text{exp}}(f) = \frac{1}{N} \sum_{i=1}^N L(y_i, f(x_i))$$

病态问题

大数定律中，当样本容量N趋于无穷大，  
经验风险趋于期望风险





# 机器学习三要素

## • 策略

### • 常用损失函数

– 0-1 损失函数 
$$L(Y, f(X)) = \begin{cases} 1, & Y \neq f(X) \\ 0 & Y = f(X) \end{cases}$$

– 对数损失函数 或对数似然损失函数

$$L(Y, f(X)) = -\log P(Y | X)$$

– 平方损失函数 
$$L(Y, f(X)) = (Y - f(X))^2$$

– 绝对损失函数 
$$L(Y, f(X)) = |Y - f(X)|$$

常用分类问题

常用回归问题

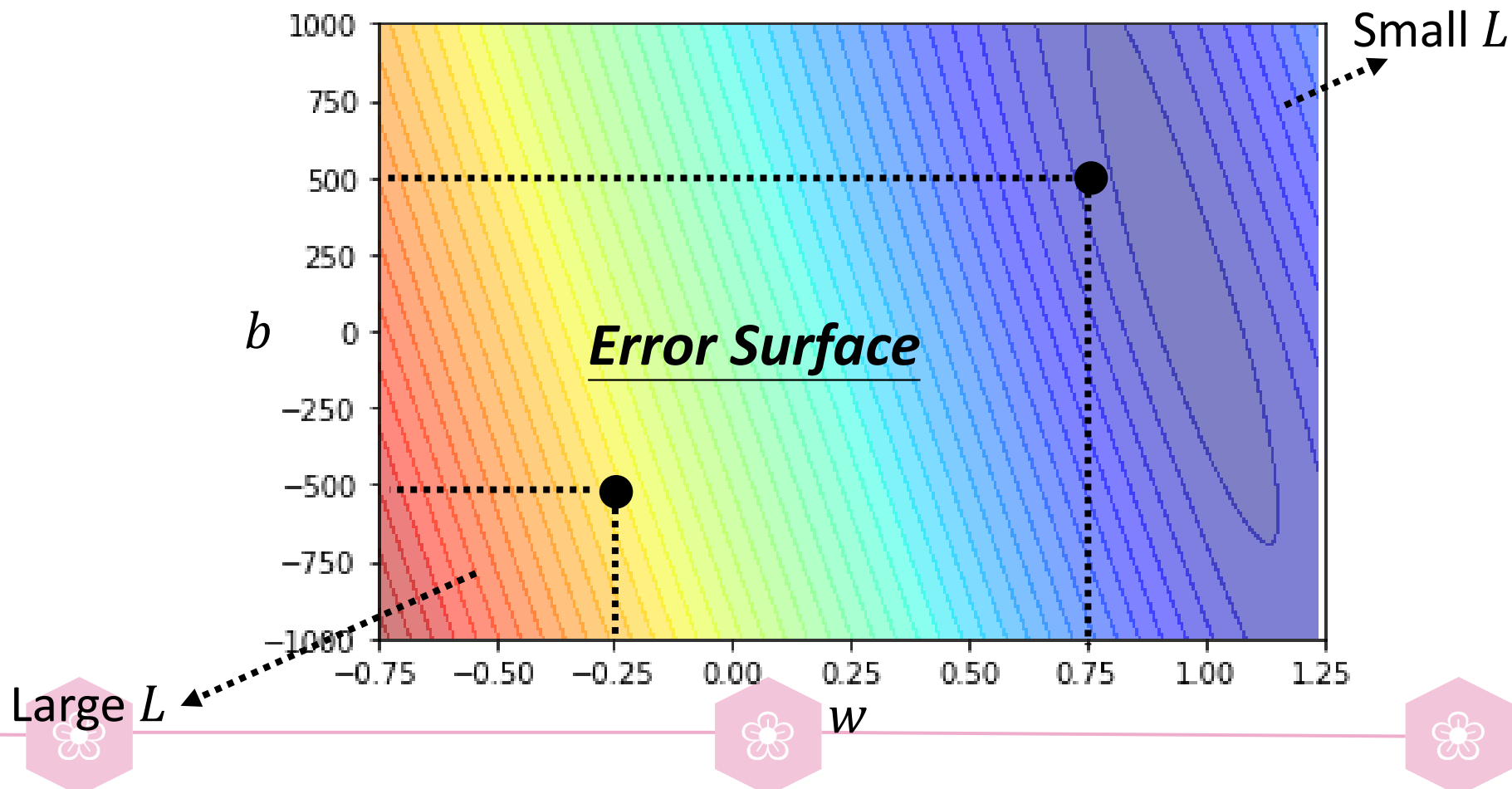


# Loss

$e = |y - \hat{y}|$       $L$  is mean absolute error (MAE)

$e = (y - \hat{y})^2$       $L$  is mean square error (MSE)

- $L(w, b) = \sum_{i=1}^N (\hat{y}_i - (b + w \cdot x_i))^2$





# 机器学习三要素

## • 算法

- 学习模型的具体计算方法
- 求解最优化问题
  - 显式的解析解
  - 数值计算的方法

$$\sum_{j=1}^n X_{ij}\beta_j = y_i, (i = 1, 2, 3 \dots m)$$

其中 $m$ 代表有 $m$ 个等式， $n$ 代表有 $n$ 个未知数 $\beta$ ， $m > n$ ；将其进行向量化后为：

$$X\beta = y$$

$$X = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2n} \\ \vdots & \vdots & & \vdots \\ X_{m1} & X_{m2} & \dots & X_{mn} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

显然该方程组一般而言没有解，所以为了选取最合适的 $\beta$ 让该等式“尽量成立”，引入残差平方和函数 $S$

$$S(\beta) = \|X\beta - y\|^2$$

(在统计学中，残差平方和函数可以看成 $n$ 倍的均方误差MSE)

当 $\beta = \hat{\beta}$ 时， $S(\beta)$ 取最小值，记作：

$$\hat{\beta} = \operatorname{argmin}(S(\beta))$$

通过对 $S(\beta)$ 进行微分<sup>[2]</sup>求最值，可以得到：

$$X^T X \hat{\beta} = X^T y$$

如果矩阵 $X^T X$ 非奇异则 $\beta$ 有唯一解<sup>[3]</sup>：

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

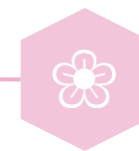
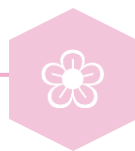


# optimization

- $L(w, b) = \sum_{i=1}^N (\hat{y}_i - (b + w \cdot x_i))^2$

$$f^* = \arg \min_f L(f)$$

$$w^*, b^* = \arg \min_{w, b} L(w, b) = \arg \min_{w, b} \sum_{i=1}^N (\hat{y}_i - (b + w \cdot x_i))^2$$



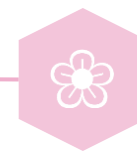
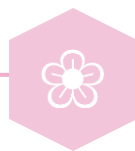
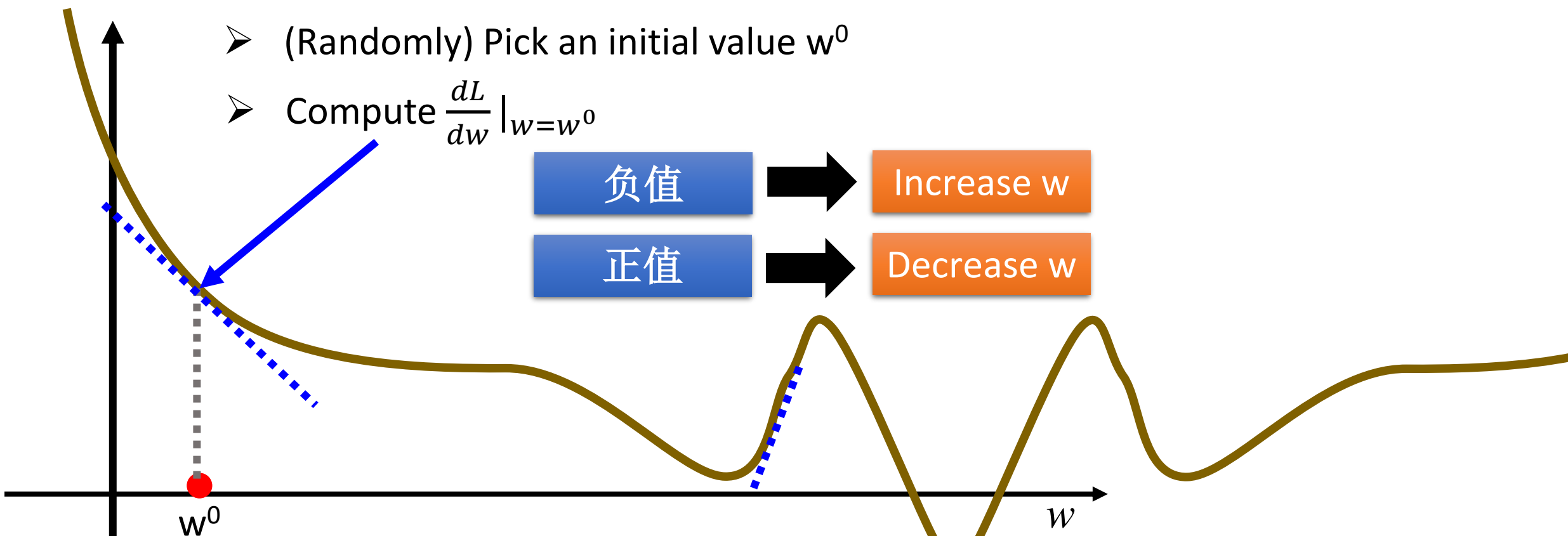
# 梯度下降

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

- 考虑一个参数 $w$ 的 loss function  $L(w)$  :

➤ (Randomly) Pick an initial value  $w^0$

➤ Compute  $\frac{dL}{dw} \big|_{w=w^0}$



# 梯度下降

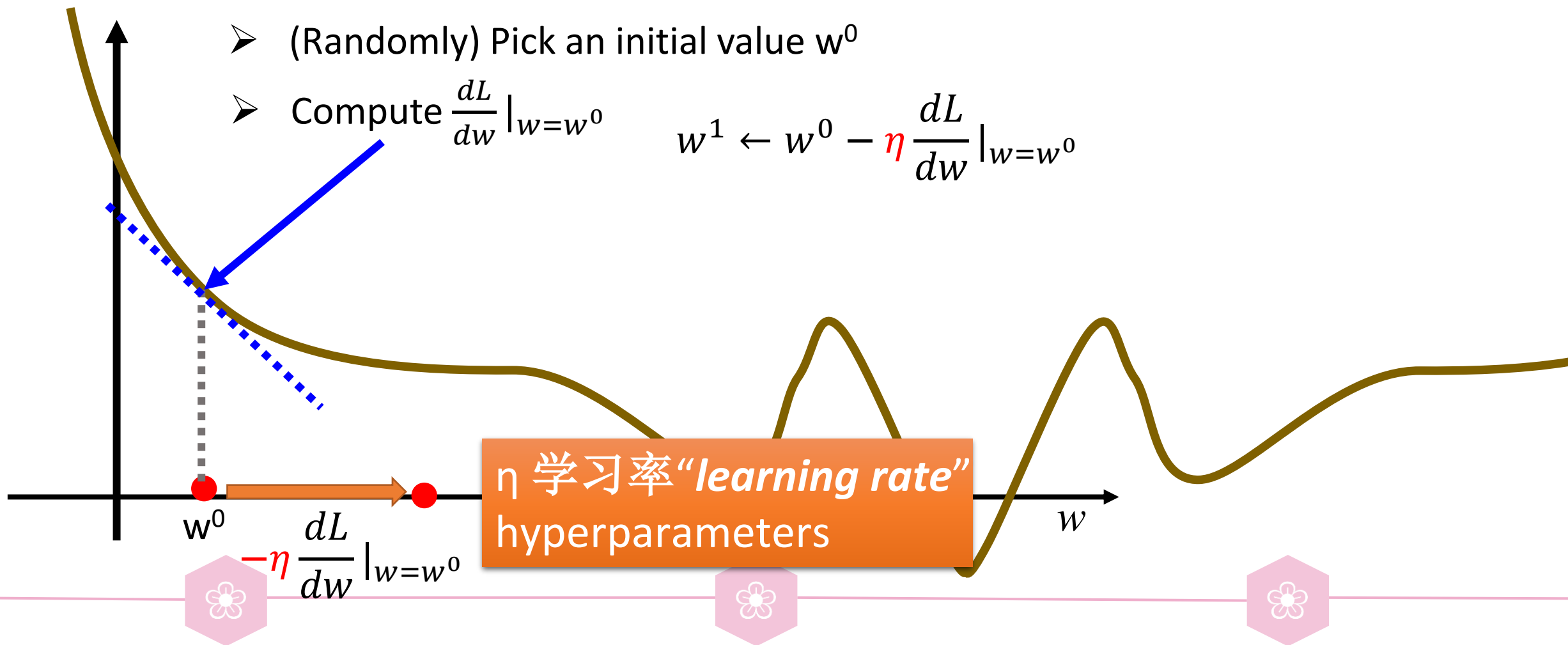
$$w^* = \arg \min_w L(w)$$

- 考虑一个参数 $w$ 的 loss function  $L(w)$  :

➤ (Randomly) Pick an initial value  $w^0$

➤ Compute  $\frac{dL}{dw} \big|_{w=w^0}$

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



# 梯度下降基本步骤

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

- 考虑一个参数 $w$ 的 loss function  $L(w)$  :

➤ (Randomly) Pick an initial value  $w^0$

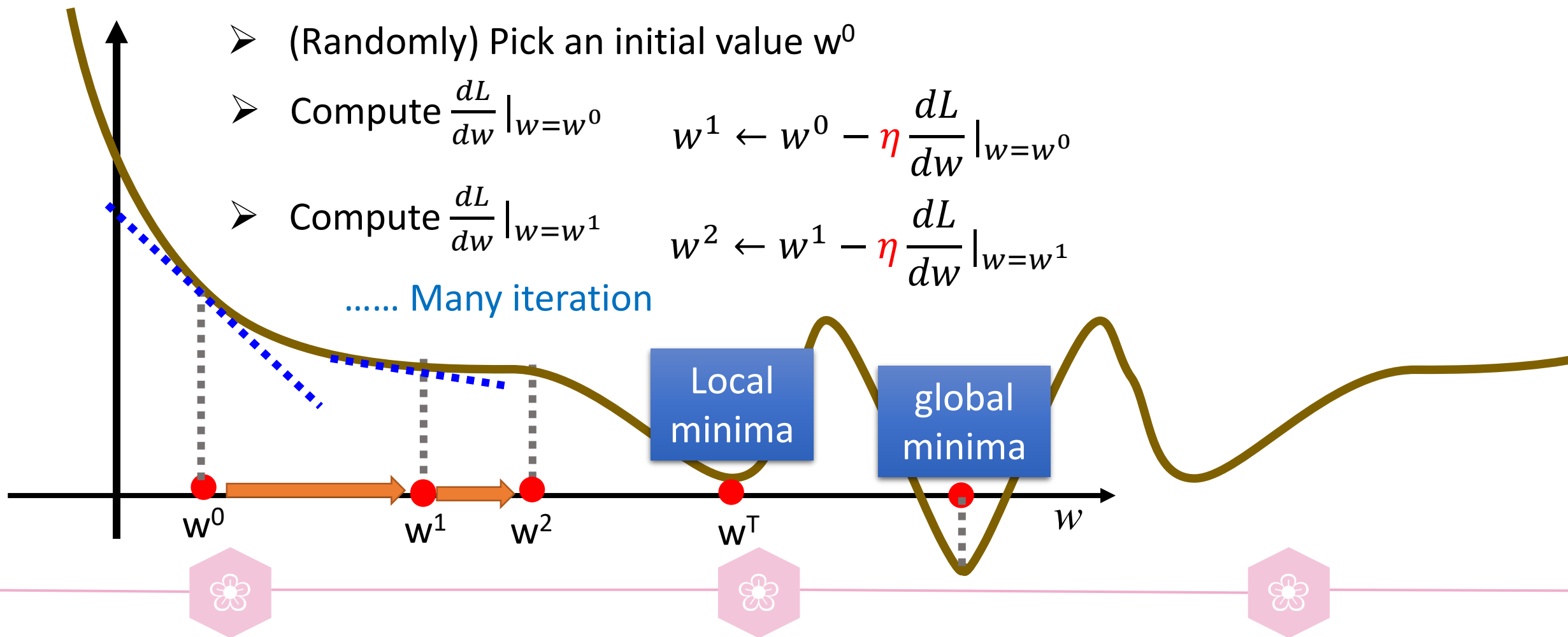
➤ Compute  $\frac{dL}{dw} \big|_{w=w^0}$

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

➤ Compute  $\frac{dL}{dw} \big|_{w=w^1}$

$$w^2 \leftarrow w^1 - \eta \frac{dL}{dw} \big|_{w=w^1}$$

..... Many iteration



# 梯度下降基本步骤

- 两个参数的时候呢?

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

➤ (Randomly) Pick an initial value  $w^0, b^0$

➤ Compute  $\frac{\partial L}{\partial w} \big|_{w=w^0, b=b^0}, \frac{\partial L}{\partial b} \big|_{w=w^0, b=b^0}$

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

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

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





# 梯度下降基本步骤

所有参数初始化 $w, \theta$

for iter=1 to max (迭代次数)

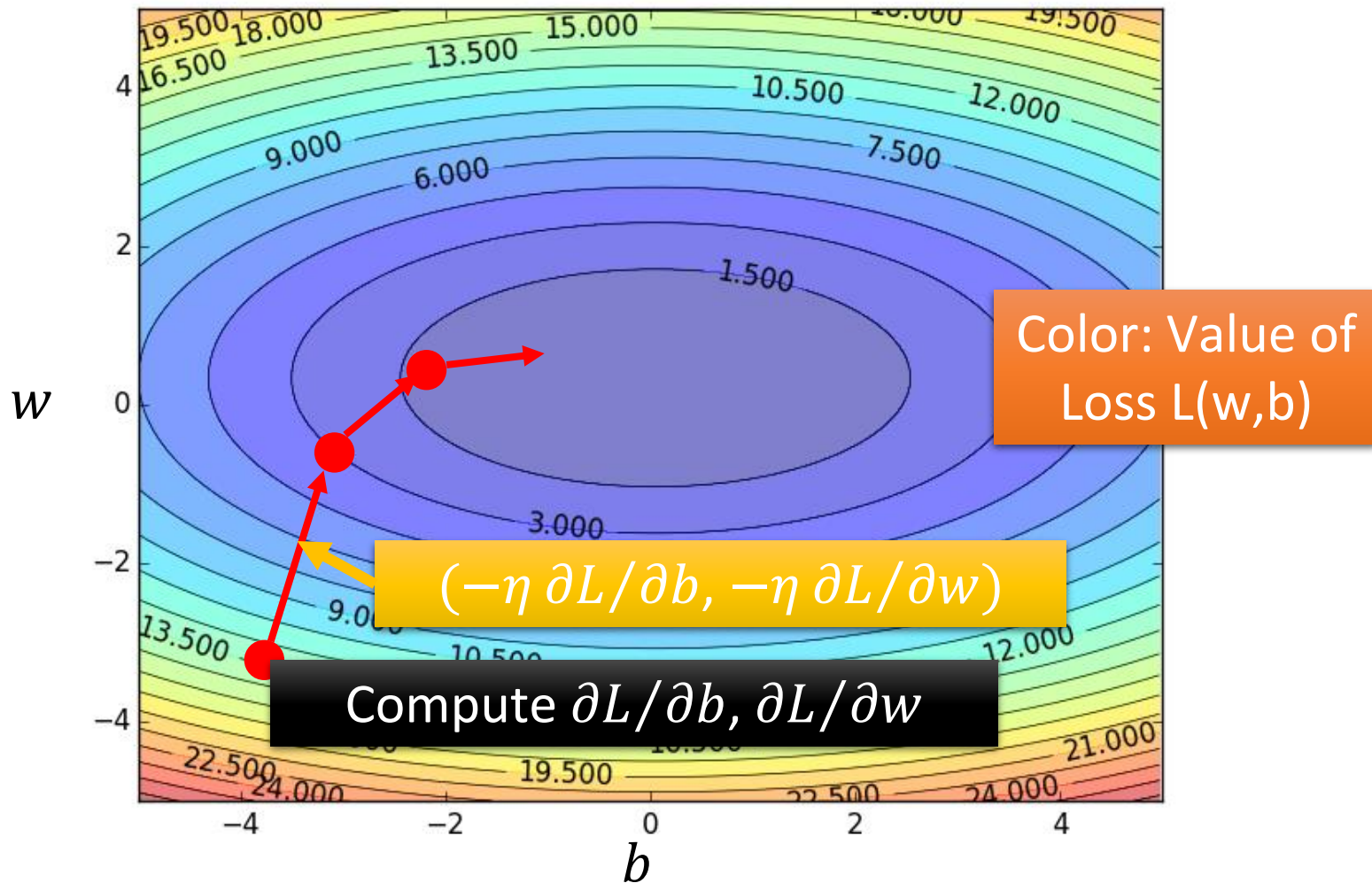
for i=1 to N: (所有样本)

计算 $\Delta b_i$

计算 $\Delta w_i$

$$w \leftarrow w - \eta \sum_i \Delta w_i$$

$$b \leftarrow b - \eta \sum_i \Delta b_i$$



# 模型性能评估

- 回归（预测）问题

- 结果评估：平均误差

- $y = b + w \cdot x$

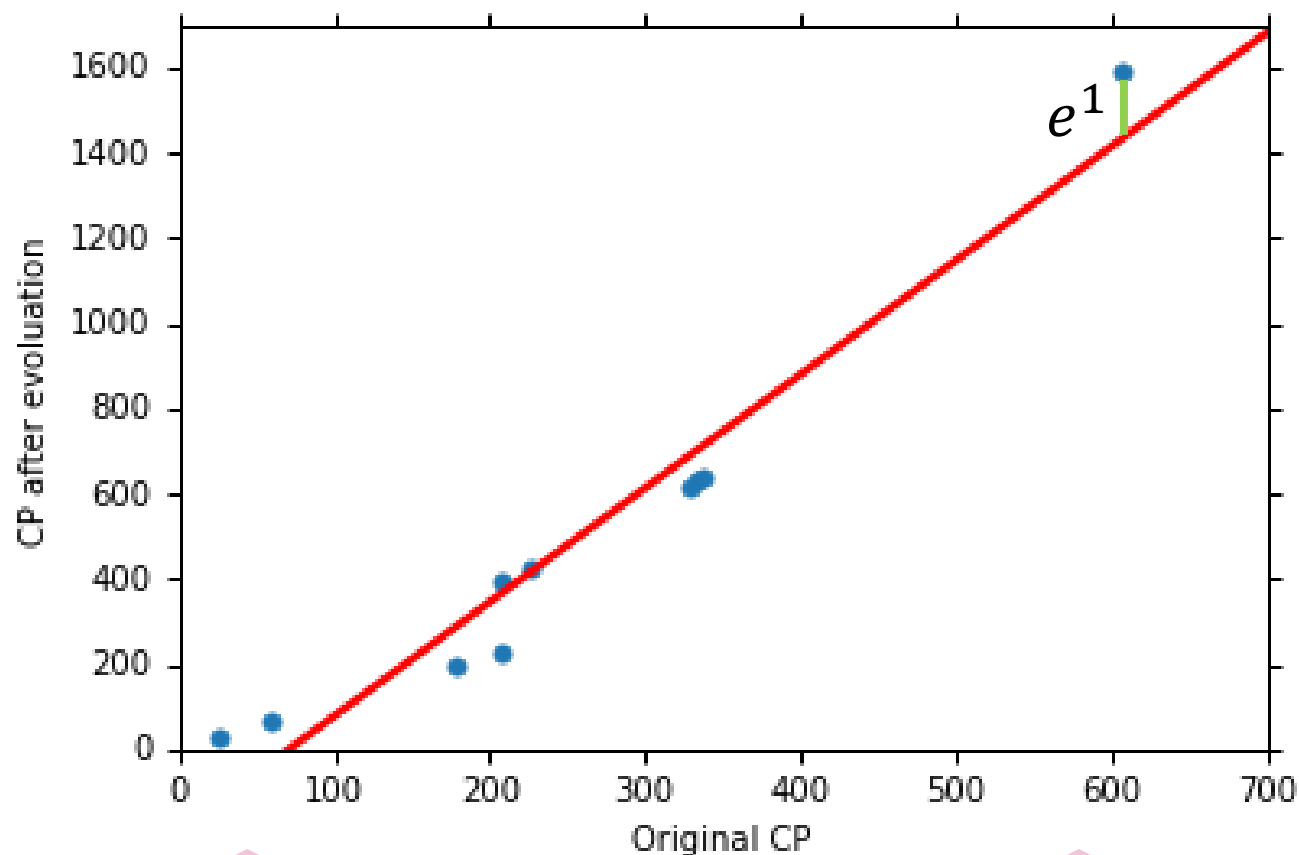
- $b = -188.4$

- $w = 2.7$

平均误差  
(训练数据)

$$= \frac{1}{10} \sum_{n=1}^{10} e^n = 31.9$$

训练集



# 模型性能评估

- 回归（预测）问题

- 结果评估：平均误差

- $y = b + w \cdot x$

$$b = -188.4$$

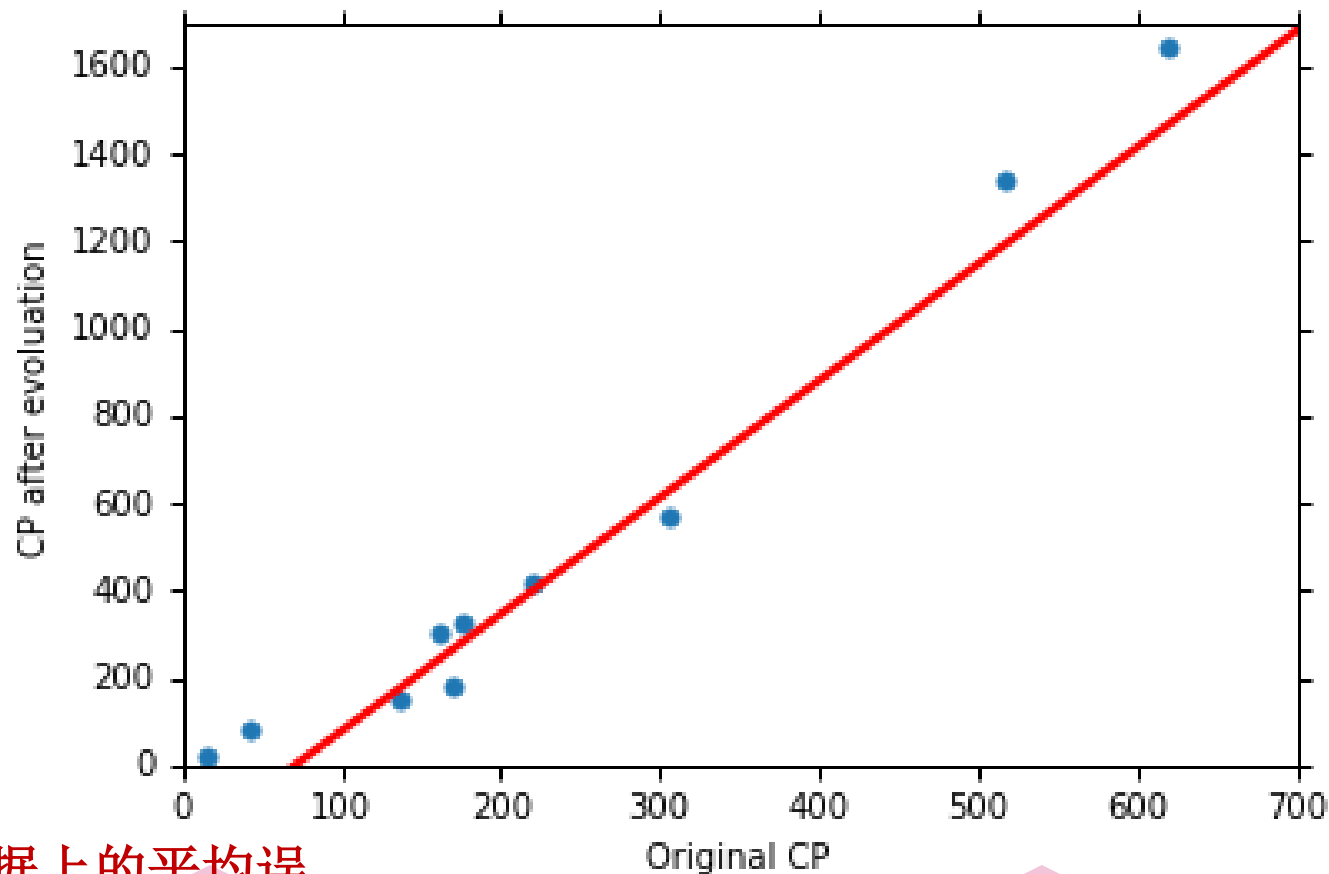
$$w = 2.7$$

平均误差  
(训练数据)

$$= \frac{1}{10} \sum_{n=1}^{10} e^n = 35.0$$

我们更关心的在新的数据集上的学习能力  
(testing data)

测试集



> 训练数据上的平均误差  
(31.9)

# 模型性能评估

$$y = b + w_1 \cdot x_{cp} + w_2 \cdot (x)^2 + w_3 \cdot (x)^3 + w_4 \cdot (x)^4 + w_5 \cdot (x)^5$$

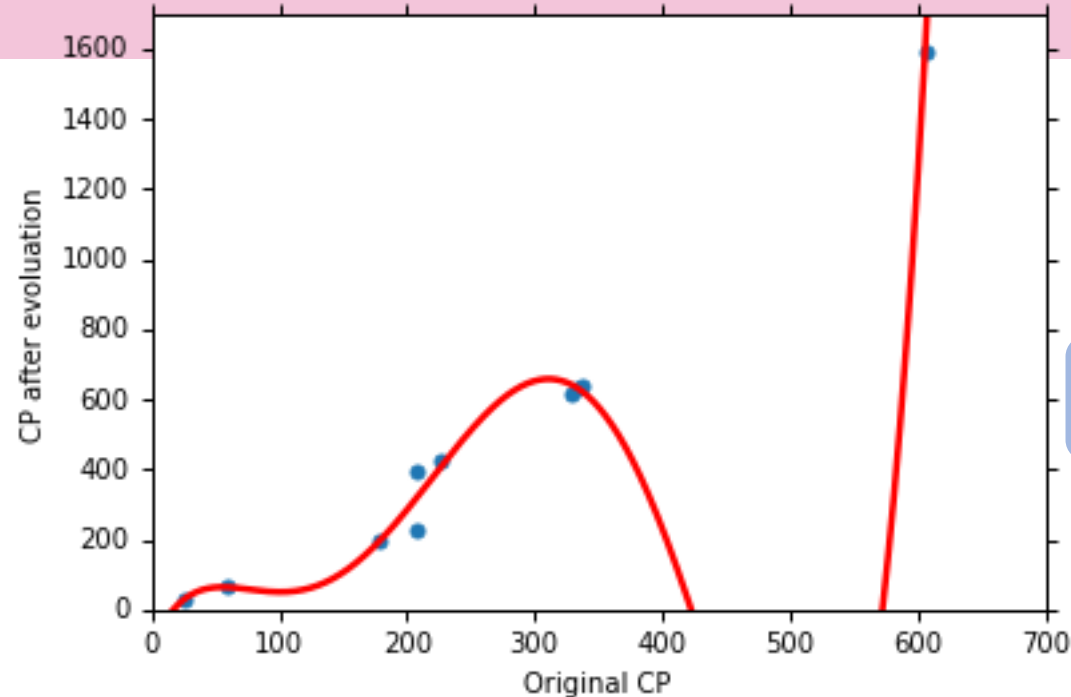
## Training

Average Error = 12.8

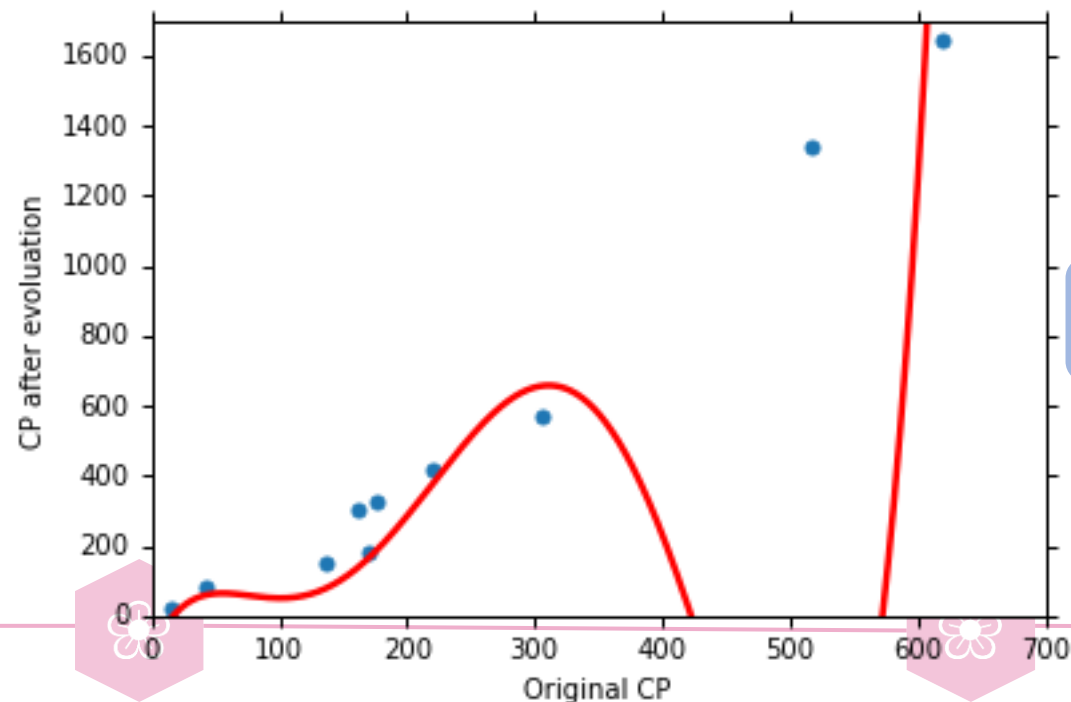
## Testing:

Average Error = 232.1

结果变的更差了!



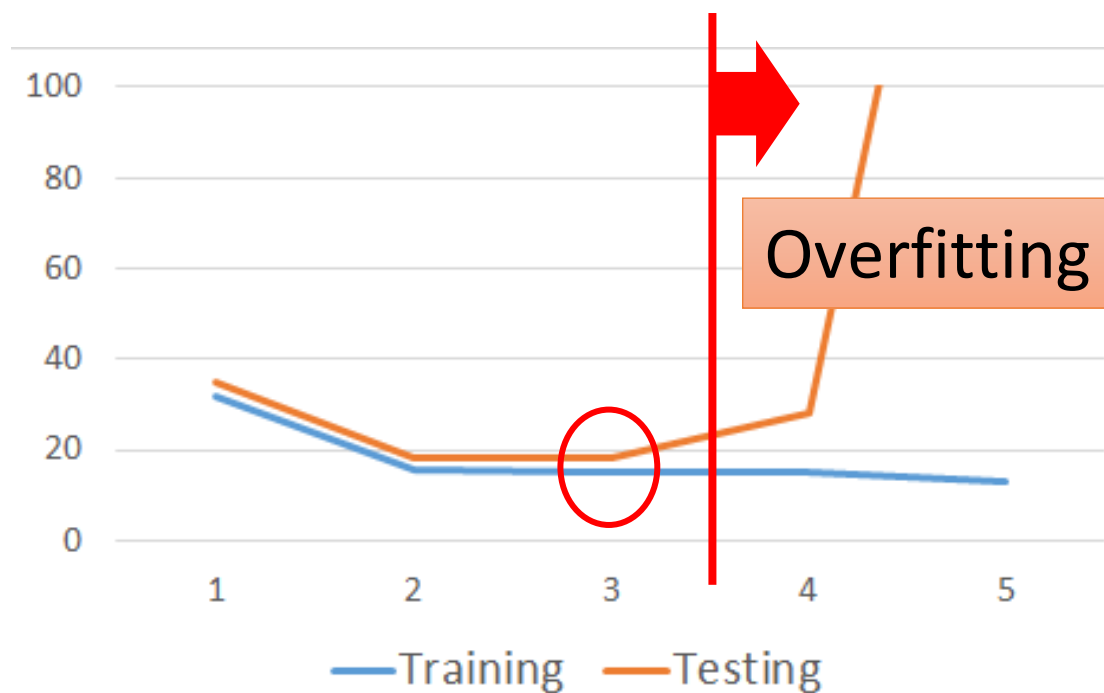
训练集



测试集



# 模型性能评估



	Training	Testing
1	31.9	35.0
2	15.4	18.4
3	15.3	18.1
4	14.9	28.2
5	12.8	232.1

越复杂的模型未必在测试集上获得更好的结果

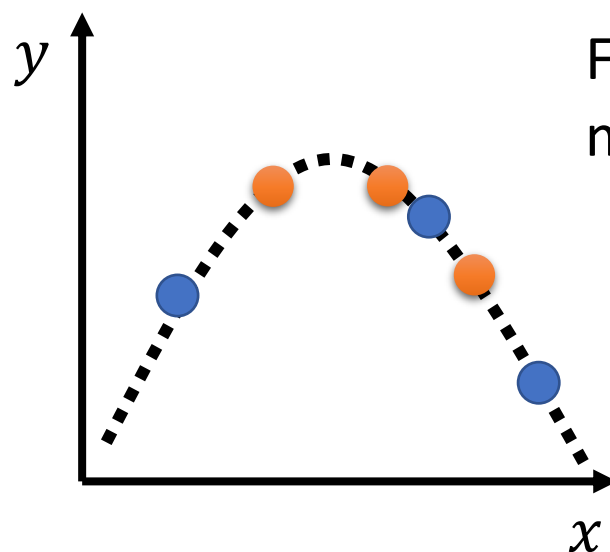
This is Overfitting.



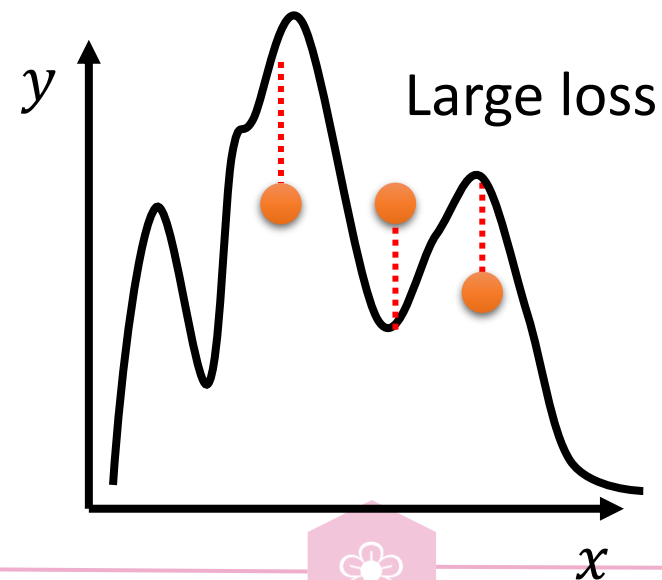
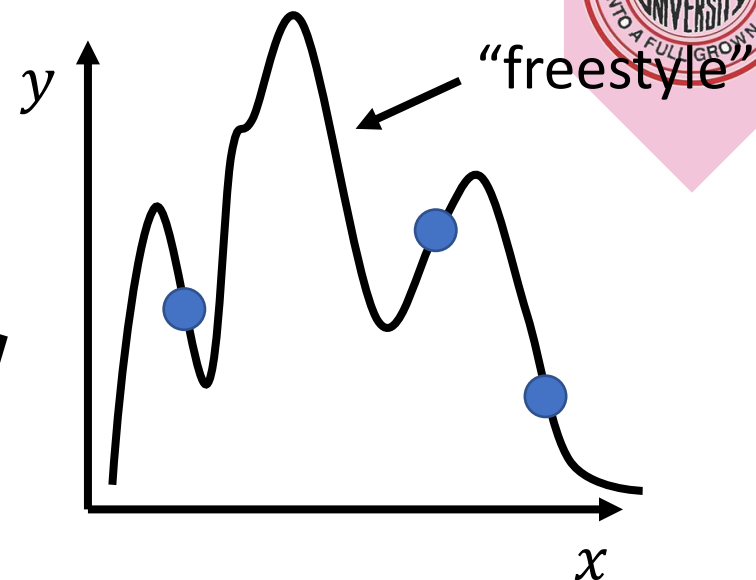
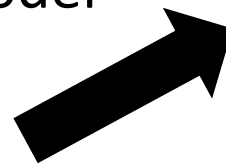
选择合适的模型，  
而不是复杂的模型

# 过拟合

- **过拟合**：模型对已知数据预测很好，对未知数据预测很差
- 当样本容量很小时，经验风险最小化学习的效果未必很好，会产生“**过拟合**”。



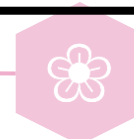
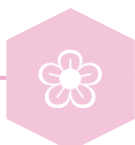
Flexible model



---- Real data distribution (not observable)

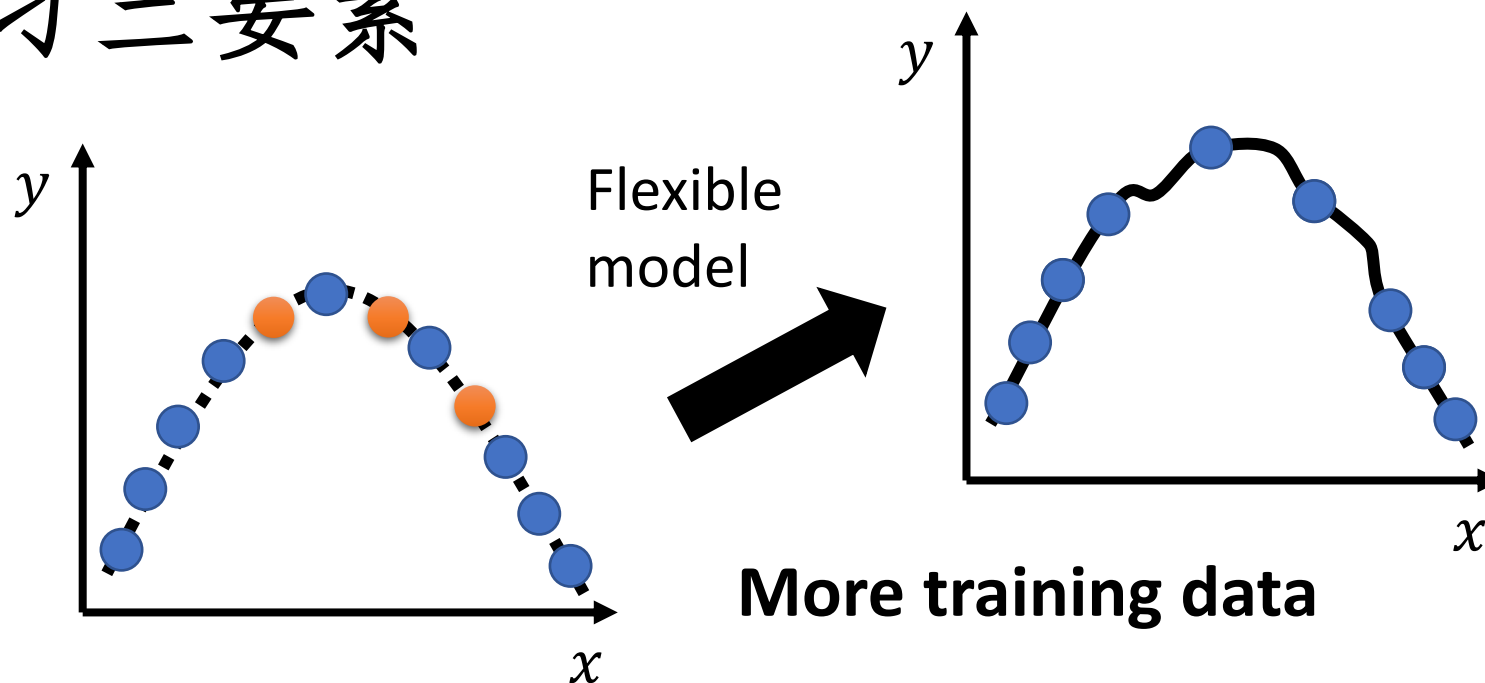
● Training data

● Testing data





# 机器学习三要素



## Data augmentation



# 正则化Regularization

- 为防止过拟合加入正则化项（regularization），或罚项（penalty term）：

$$R_{\text{exp}}(f) = \frac{1}{N} \sum_{i=1}^N L(y_i, f(x_i)) + \lambda J(f) \quad \text{结构风险}$$

- 正则化项一般是模型复杂度的单调递增函数，模型越复杂，正则化值越大。比如：正则化项可以是模型参数向量的范数

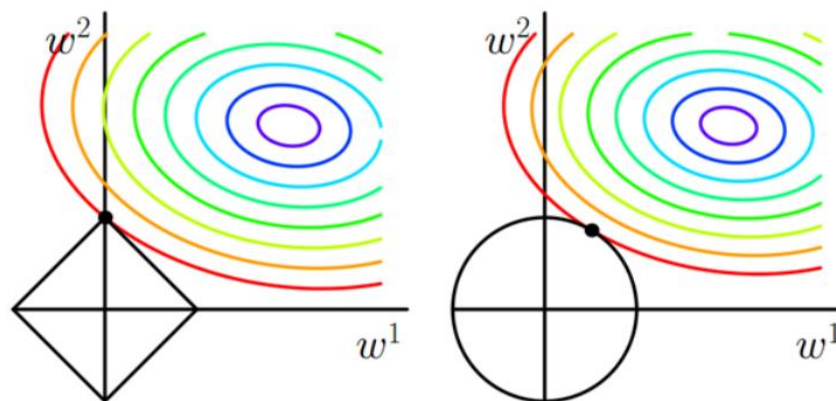
$$L(w) = \frac{1}{2} \sum_{i=1}^N (f(x_i, w) - y_i)^2 + \frac{\lambda}{2} \|w\|_2^2 \quad l_2 \text{ 范数}$$

$$L(w) = \frac{1}{2} \sum_{i=1}^N (f(x_i, w) - y_i)^2 + \lambda \|w\|_1 \quad l_1 \text{ 范数}$$

# Regularization

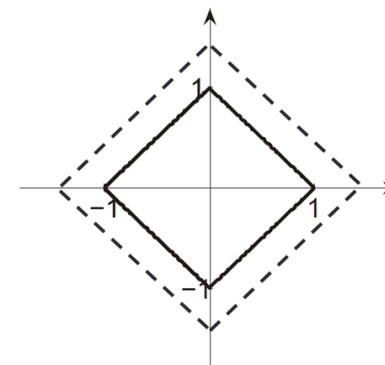
- L2 范式:  $\|w\|_2^2 = w_1^2 + w_2^2 + \dots + w_d^2$
- L1 范式:  $\|w\|_1 = |w_1| + |w_2| + \dots + |w_d|$

- $w_1^2 + w_2^2 \leq 1$
- $|w_1| + |w_2| \leq 1$
- $\lambda$  的意义



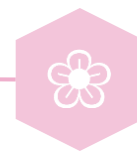
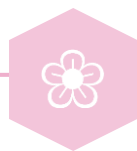
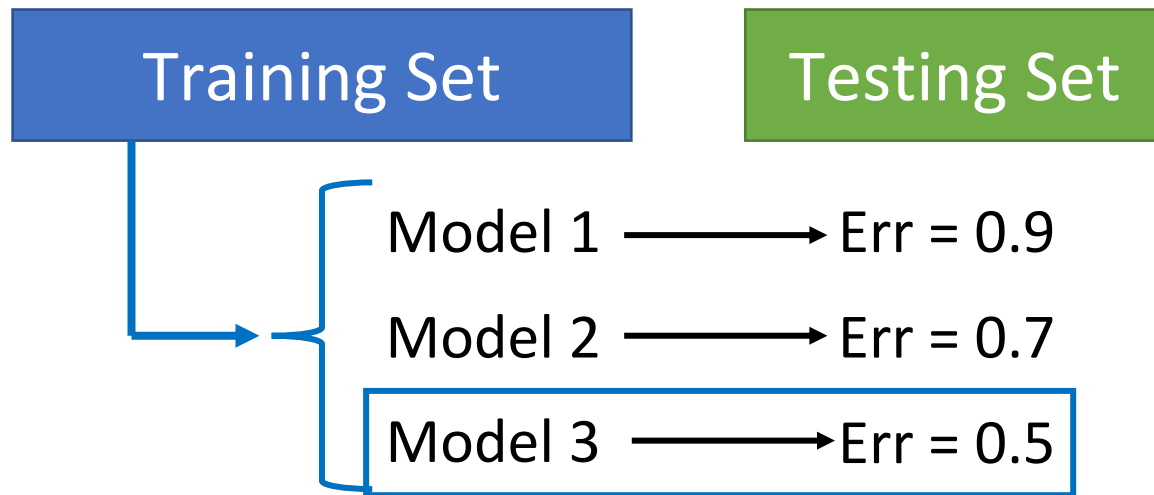
- 特征选择与学习器训练过程融为一体

嵌入式特征选择

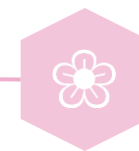
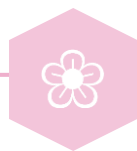
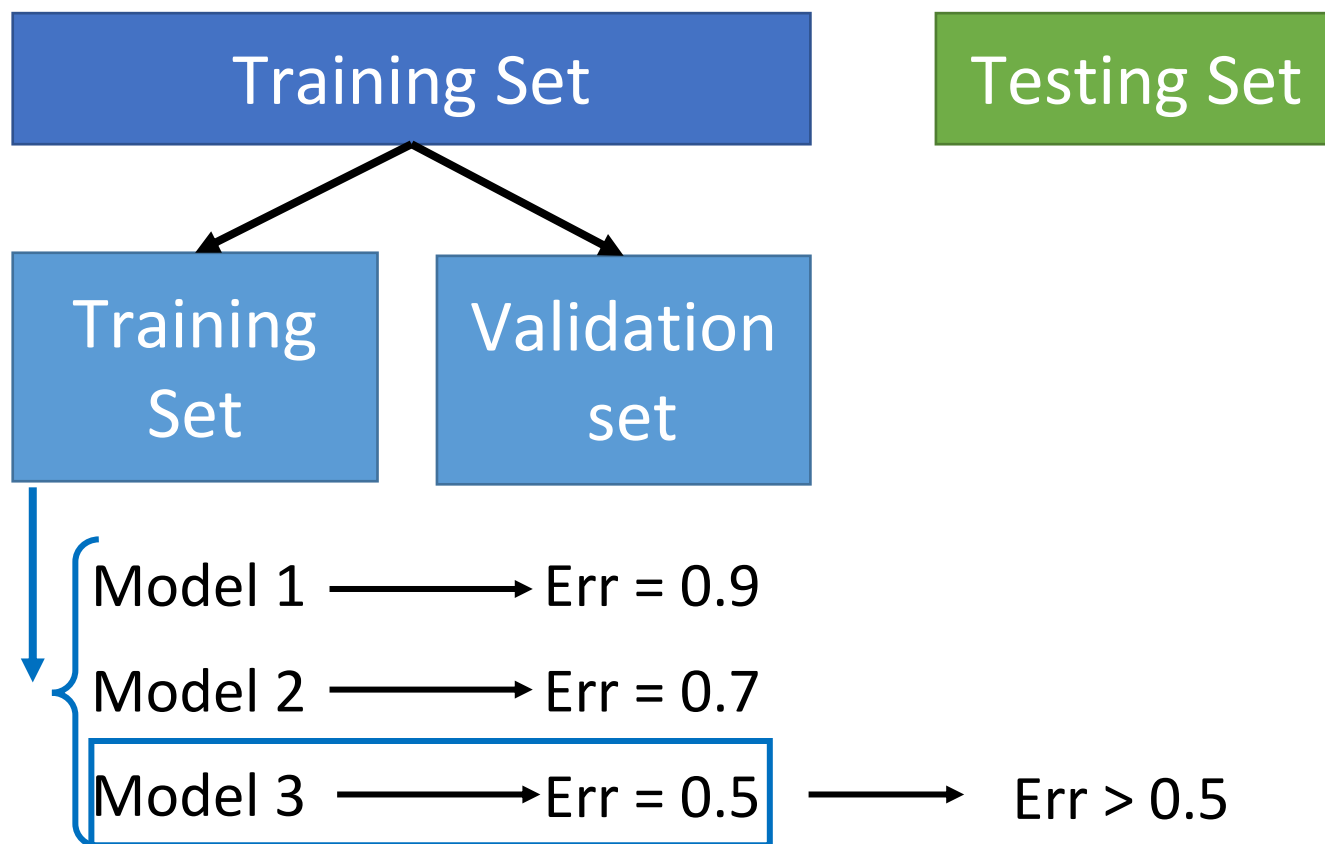


$$\lambda (|\beta_1| + |\beta_2|) \leq 1$$

# 模型测试



# 模型测试



# 怎么理解调参

$$y = xw + b$$

## ❖ 程序流程

- 初始化两个参数 $w, b$ , 设置步长 $\eta$ , 设置循环次数epoch

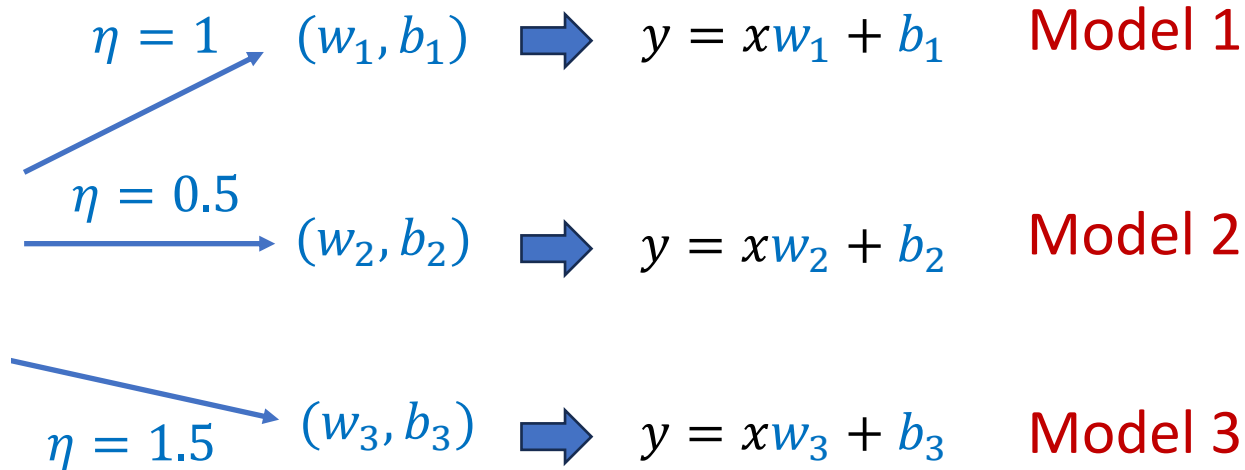
- For  $i = 1$  to epoch

- 计算 $w$ 的梯度 $\Delta w$
- 计算 $b$ 的梯度 $\Delta b$
- $w = w - \eta \Delta w$
- $b = b - \eta \Delta b$

终止:

1. 循环结束
2. 两个参数梯度为0

超参数

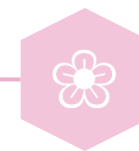
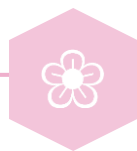
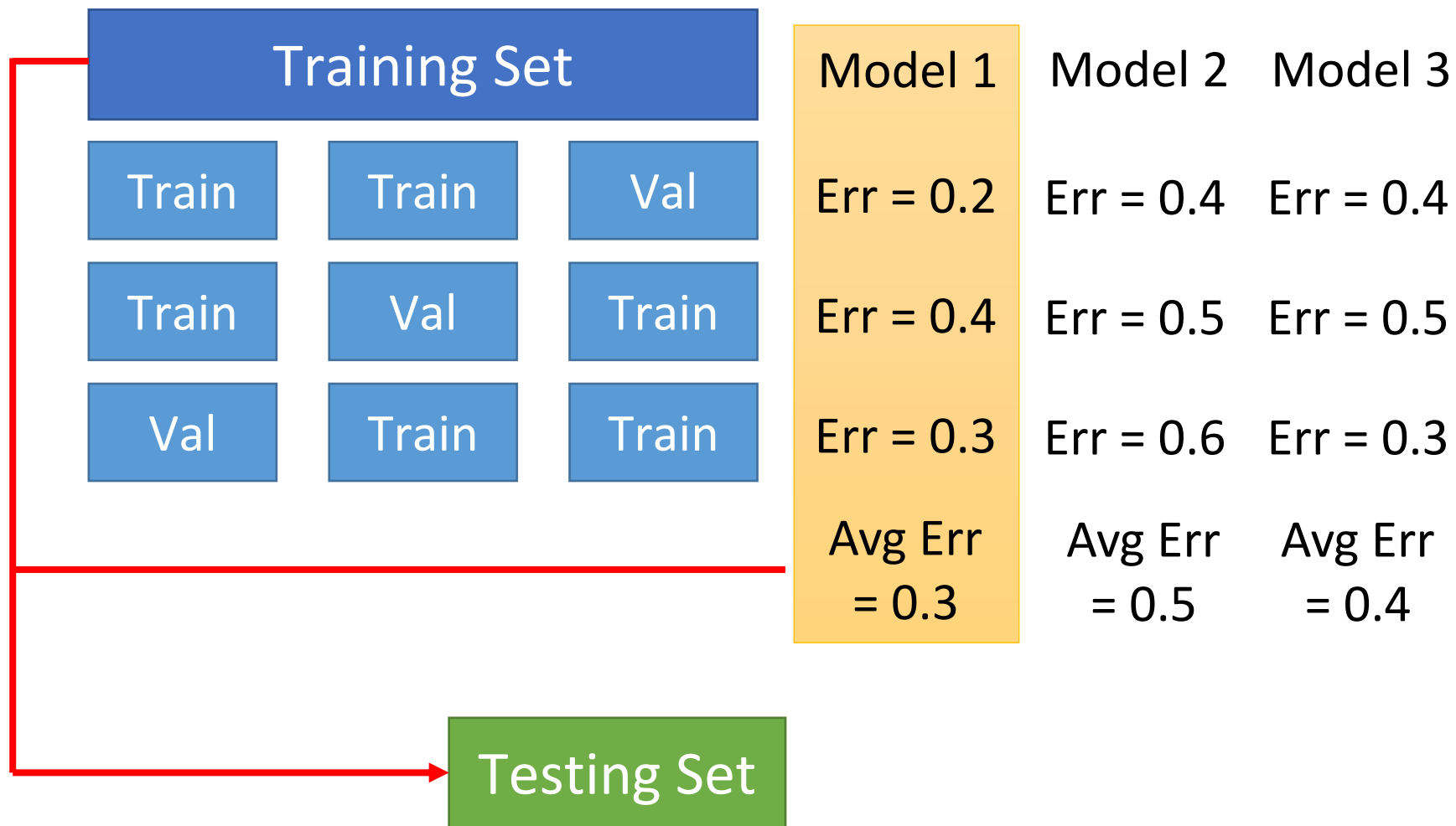


调超参数 $\eta$



# 模型测试

## S折交叉验证

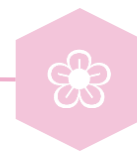
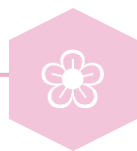




# 性能度量

- 性能度量是衡量模型泛化能力的评价标准，反映了任务需求
  - 对比不同模型的能力，使用不能的性能度量往往会导致不同的评判结果

模型的“好”不仅取决于算法和数据，还取决于任务需求





# 性能度量

- 预测问题

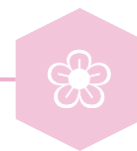
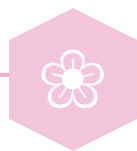
- **MSE(mean squared error)**

- $E = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2$

- 分类问题

- 分类准确率: **accuracy=正确分类的个数/样本总个数**

100张图片，通过模型有80张识别正确，20张错误，准确率=80/100=0.8





# 性能度量

- 分类问题

- 精确率与召回率 **Precision & Recall**
- **TP**-将正类预测为正类数
- **FN**-将正类预测为负类数
- **FP**-将负类预测为正类数
- **TN**-将负类预测为负类数





# 性能度量

- 精确率Precision

- 针对我们预测结果而言的，它表示的是预测为 **正的样本中有多少是真正的正样本**。  
那么预测为正就有两种可能了，一种就是把正类预测为正类(TP)，另一种就是把负类预测为正类(FP)

- $P = \frac{TP}{TP+FP}$

表 2.1 分类结果混淆矩阵

真实情况	预测结果	
	正例	反例
正例	TP (真正例)	FN (假反例)
反例	FP (假正例)	TN (真反例)



# 性能度量

- 召回率Recall

- 针对我们原来的样本而言的，它表示的是**样本中的正例有多少被预测正确了**。那也有两种可能，一种是把原来的正类预测成正类(TP)，另一种就是把原来的正类预测为负类(FN)

表 2.1 分类结果混淆矩阵

真实情况	预测结果	
	正例	反例
正例	TP (真正例)	FN (假反例)
反例	FP (假正例)	TN (真反例)

- $$R = \frac{TP}{TP+FN}$$







# 性能度量

- 二分类（正例，负例）

- 真实数据情况，正例样本100，负例样本40个
- 模型实际预测结果：

- 正例里面预测正确了80个，20个预测错误
- 负例里面预测正确30个，10个预测错误

- 精确率 $p = \frac{80}{90} = 88.89\%$

- 召回率 $r = \frac{80}{100} = 80\%$

表 2.1 分类结果混淆矩阵

真实情况	预测结果	
	正例	反例
正例	TP (真正例)	FN (假反例)
反例	FP (假正例)	TN (真反例)

真实情况	预测结果	
	正例	负例
正例	80	20
负例	10	30





# 性能度量

- 二分类（正例，负例）

- 真实数据情况，正例样本100，负例样本40个
- 模型实际预测结果：
  - 正例里面预测正确了80个，20个预测错误
  - 负例里面预测正确30个，10个预测错误

- 精确率 $p = \frac{80}{90} = 88.89\%$

- 召回率 $r = \frac{80}{100} = 80\%$

表 2.1 分类结果混淆矩阵

真实情况	预测结果	
	正例	反例
正例	TP (真正例)	FN (假反例)
反例	FP (假正例)	TN (真反例)

真实情况	预测结果	
	正例	负例
正例	80	20
负例	10	30

$$F1 = \frac{2 \times P \times R}{P + R} = 84.3\%$$

