智能软件开发方向基础

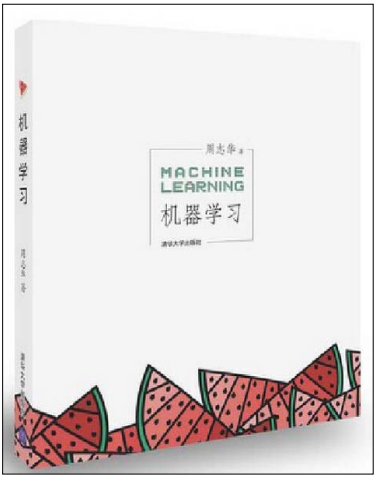
第二章　机器学习基本概念

张朝晖

2023年2月～2023年6月

河北师范大学软件学院
Software College of Hebei Normal University

课程主要参考书



河北师范大学软件学院
Software College of Hebei Normal University

主要内容

1. 什么是机器学习
2. 机器学习的相关术语
3. 机器学习的典型任务
4. 机器学习的学习范式
5. 假设与假设空间
6. 假设的选择原则
7. 机器学习的三要素

关于机器学习的一种形式化定义

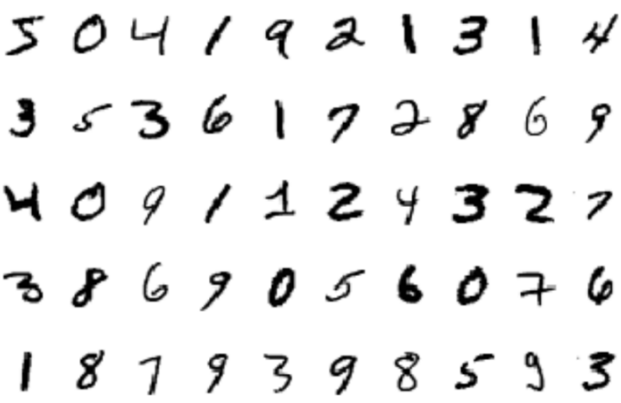
➢ Tom M. Mitchell *(1997).* *Machine Learning*, McGraw Hill
http://www.cs.cmu.edu/~tom/

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E.

以P评价计算机程序关于某类任务T上的性能
如果某程序利用经验E使T中的任务获得了性能改善，则称该程序从经验E中进行了学习.

“机器学习”形式化定义的三要素：

- 明确指定的任务T
- 评价任务的性能的度量指标P
- 用于改善任务性能的经验E



例：手写体数字识别

- 任务T：识别或预测给定的手写体数字图像的类别
- 经验E：已知类别标记的手写体样本图像构成的数据集
- 评价任务性能P的指标：学习系统关于训练样本集的预测正确率

机器学习（Machine Learning）是一门涉及统计学、系统辨识、逼近理论、神经网络、优化理论、计算机科学、脑科学等诸多领域的**交叉学科**，研究计算机怎样**模拟或实现**人类的学习行为，以**获取**新的知识或技能，**重新组织**已有的知识结构使之**不断改善自身的性能**，是**人工智能技术的核心**。

基于数据的机器学习是现代智能技术中的重要方法之一，研究从**观测数据（样本）**出发寻找规律，利用这些规律对未来数据或无法观测的数据进行预测。

--人工智能标准化白皮书(2018版)

河北师范大学软件学院
Software College of Hebei Normal University

Machine learning is a branch of artificial intelligence, concerns the construction and study of systems that can learn from data.

机器学习是人工智能的一个分支，对系统进行构造和研究，使之可以从数据中学习。

Machine learning, a branch of artificial intelligence, is a scientific discipline concerned with the design and development of algorithms that take as input empirical data, and yield patterns or predictions thought to be features of the underlying mechanism that generated the data.

机器学习是人工智能的一个分支，是一门科学学科，涉及算法的开发与设计，该算法以经验数据为输入，并产生(被认为是生成数据的潜在机制特征的)模式或预测。

河北师范大学软件学院
Software College of Hebei Normal University

关于机器学习的更为友好的定义

- **Andrew Ng (吴恩达)**

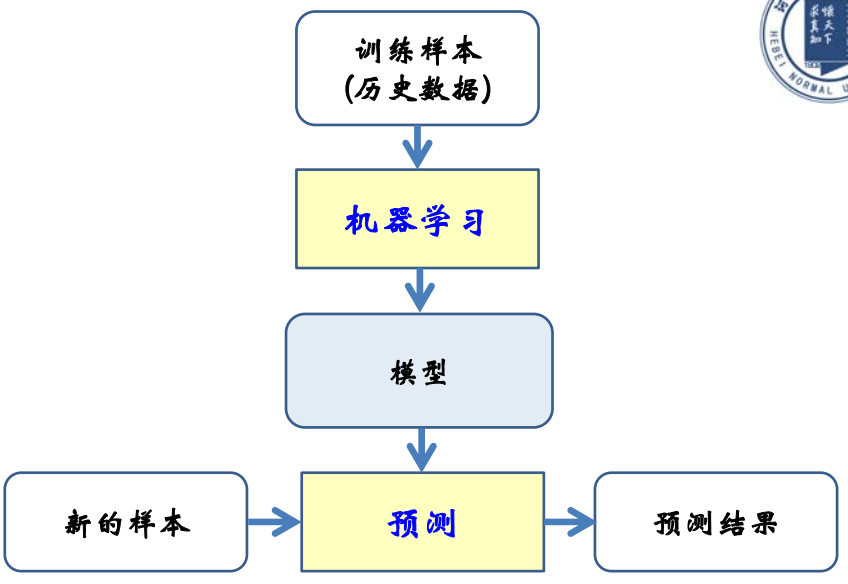
Machine Learning is the science of getting computers to act without being explicitly programmed.

机器学习是一门让计算机无需显式编程即可运行的科学。

- **Microsoft**

Machine learning is **a technique of data science** that helps computers learn from existing data in order to forecast **future behaviors, outcomes, and trends**.

机器学习是一种数据科学技术，它帮助计算机从现有数据中学习，从而预测**未来的行为、结果和趋势**。



河北师范大学软件学院
Software College of Hebei Normal University

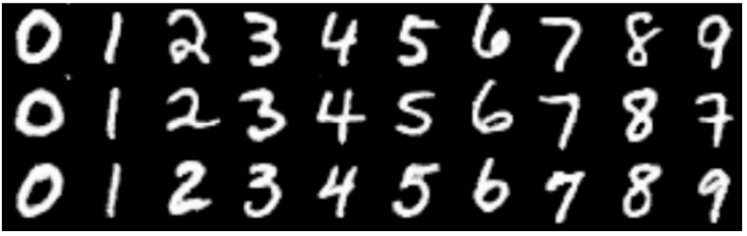
主要内容

1. 什么是机器学习
2. 机器学习的相关术语
3. 机器学习的典型任务
4. 机器学习的学习范式
5. 假设与假设空间
6. 假设的选择原则
7. 机器学习的三要素

- **样本(sample)**、示例(instance)

所研究对象的一个个体。相当于统计学中的实例(example，instance)

- **特征(feature)**、属性(attribute)



➢ **特征向量；特征维数**

用于表征样本的观测，通常是数值表示的某些量化特征，也称属性(attribute)。

分别以每个特征作为一个坐标轴，所有特征所在坐标轴张成一个用于描述不同样本的空间，称为**特征空间**。

在该空间中，每个具体样本就对应空间的一个点，在这个意义下，也称样本为**样本点**。

每个样本点对应特征空间的一个向量，“特征向量”

特征的数目即为**特征空间的维数**。

➢ **特征空间**、样本空间、属性空间、**输入空间**

例：来自d维特征空间的特征向量 $\boldsymbol{x} = [x_1, \ldots, x_d]^T$

➢**样本集 (sample set)**、数据集(data set)

若干样本构成的集合；

该集合的每个元素就是一个样本

例：d维特征空间的m个观测样本构成样本集D

$D = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m\}$

$\boldsymbol{x_i} = [x_{i1}, \ldots, x_{id}]^T \in D$

- 类别(class)与类别标签(label)
- 标记空间(label space)、输出空间
- 已知样本 (known sample)

  标签已知的样本。
- 未知样本(unknown sample)

  标签未知的样本。
- 训练样本、训练样本集
- 测试样本、测试样本集

训练集、测试集
估计集、验证集

独立同分布(i.i.d,independent and identically distributed)

监督式学习，训练集 $\{(\boldsymbol{x}_i, y_i), \boldsymbol{i} = 1,...,\boldsymbol{N}\}$

回归 $\boldsymbol{x}_i = [x_{i1},...,x_{id}]^T \in \boldsymbol{R}^d$， $y_i \in \boldsymbol{R}$

分类

两类别分类 $\boldsymbol{c} = 2$， $\boldsymbol{x}_i \in \boldsymbol{R}^d$， $y_i \in Y = \{1,-1\}$或$\{1,0\}$或$\{1,2\}$

多类别分类 $\boldsymbol{c} = 2$， $\boldsymbol{x}_i \in \boldsymbol{R}^d$， $y_i \in Y = \{1,2,...,C\}$

主要内容

1. 什么是机器学习
2. 机器学习的相关术语
3. 机器学习的典型任务
4. 机器学习的学习范式
5. 假设与假设空间
6. 假设的选择原则
7. 机器学习的三要素

学习任务用于表征通过学习可以解决的基本问题

典型的学习任务包括：

- 分类(classification)
- 回归(regression)
- 聚类(clustering)
- 排序(ranking)
- 关联分析
- 密度估计(density estimation)
- 特征降维(dimensionality reduction)
- 异常检测
- …

3.1 分类(classification)

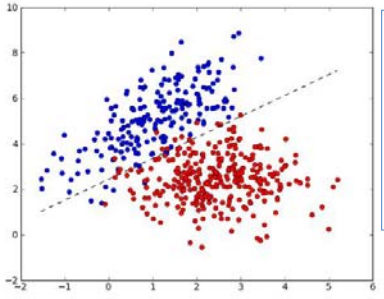
基于**已知类别标签的样本**构成的训练集，学习预测模型；最终使用预测模型，对新的观测样本，预测相应的输出；预测结果为事先指定的两个或多个类别中的某一个，或预测结果来自数目有限的离散值之一。

- **两类别 vs.多类别**

类别数C=2，两类别分类(binary classification)
类别数C>2，多类别分类(multiclass classification)

- **产生式分类模型 vs. 鉴别式分类模型**

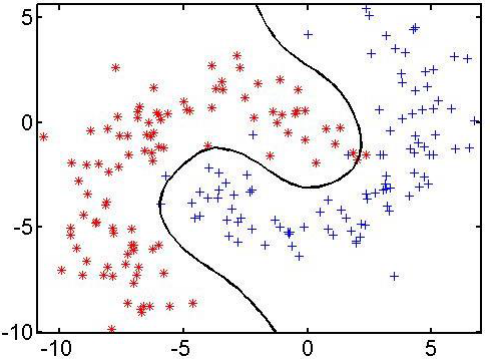
- **线性分类器 vs. 非线性分类器**



- **二维特征空间的两类别线性分类模型**
- **线性判别函数**

$$f(x) = w^T x + b$$

- **二维特征空间的两类别非线性分类**
- **分类边界是关于输入向量的非线性方程**
- **非线性判别函数**

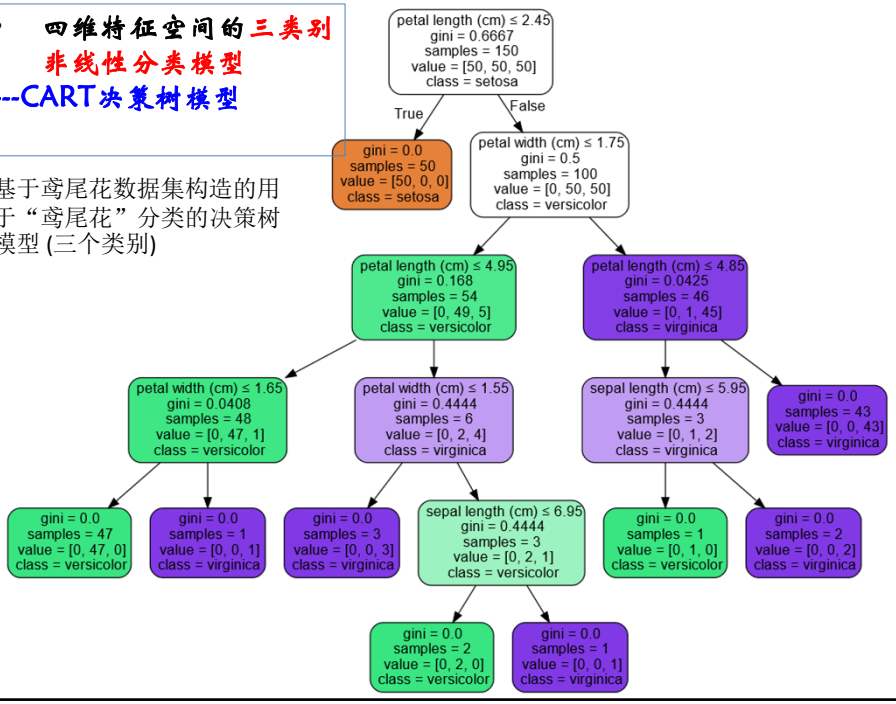


河北师范大学软件学院
Software College of Hebei Normal University

- 四维特征空间的三类别非线性分类模型

---CART决策树模型

基于鸢尾花数据集构造的用于“鸢尾花”分类的决策树模型 (三个类别)



二分类(C=2)应用举例

情况1. “是 vs. 非”

- 垃圾邮件过滤(spam filter)

  文本数据 “垃圾邮件 vs 非垃圾邮件”

- 图像或视频中特定类型的目标检测

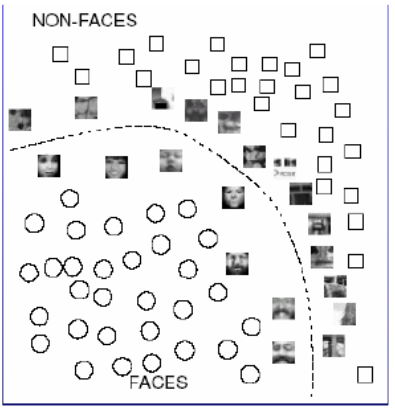
  如：人脸检测、车辆检测、行人检测

  人脸 vs. 非人脸、车辆 vs. 非车辆 行人 vs. 非行人

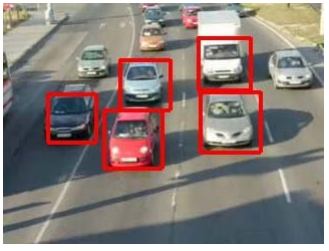
情况2. “第1类 vs 第2类”

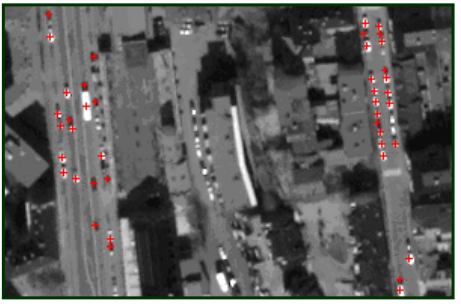
- 基于人脸图像的性别分类 “男vs.女”

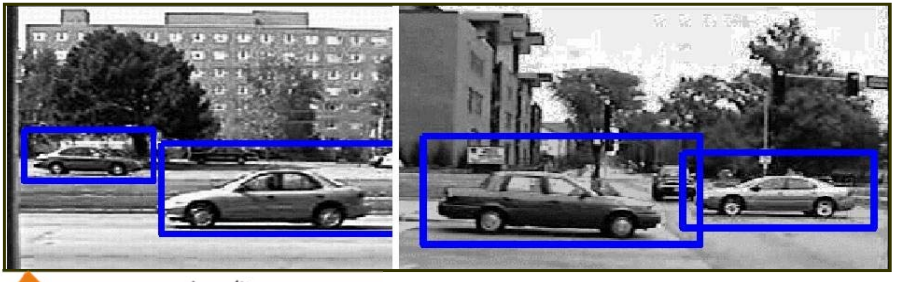
人脸检测 (Face Detection)



车辆检测





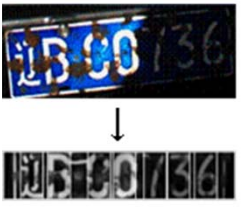


多分类(C>2)应用举例

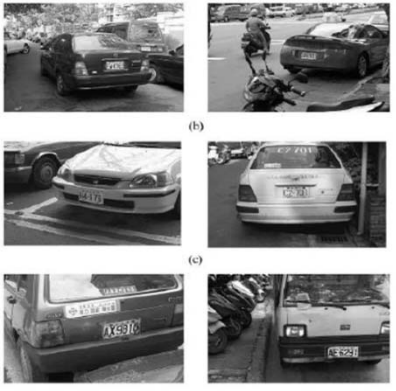
文档分类(document classification)—类别数目

Sports
News
Politics
...

光学字符识别(OCR)—汽车牌照识别



提取到的车牌



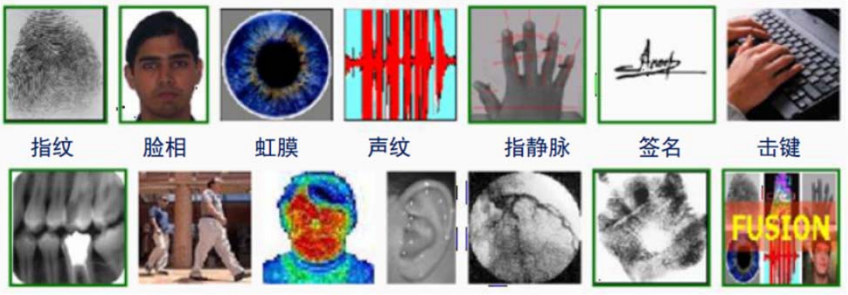
P7-577	P7577	P7577
G4-402	G4402	G4402
DU-3403	DU3403	DU3403
GG-4025	GG4025	GG4025
CX-0166	CX0166	CX0166

S.-L. Chang, et al., Automatic license plate recognition, *IEEE T-ITS*, 2004.

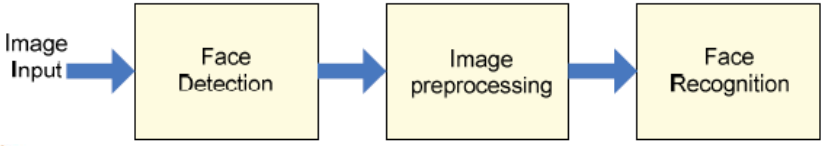
光学字符识别(OCR)—集装箱号码识别

基于生物特征的身份识别

声音、人脸、虹膜、指纹、掌纹、步态,…



➢ 人脸识别
Face Recognition



河北师范大学软件学院
Software College of Hebei Normal University

➢ 1000类的图像分类

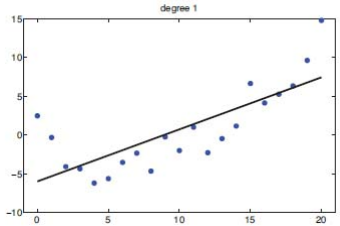
Examples from ImageNet

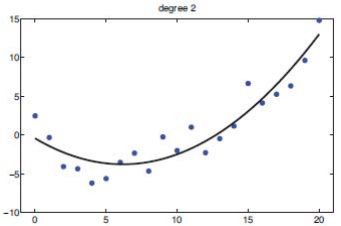
3.2 回归(regression)

回归分析基于已知标签(即：目标答案)的样本构成的训练集，估计自变量与因变量之间关系的统计过程，进而基于该关系对新的观测产生的输出进行预测，预测输出为连续的实数值。

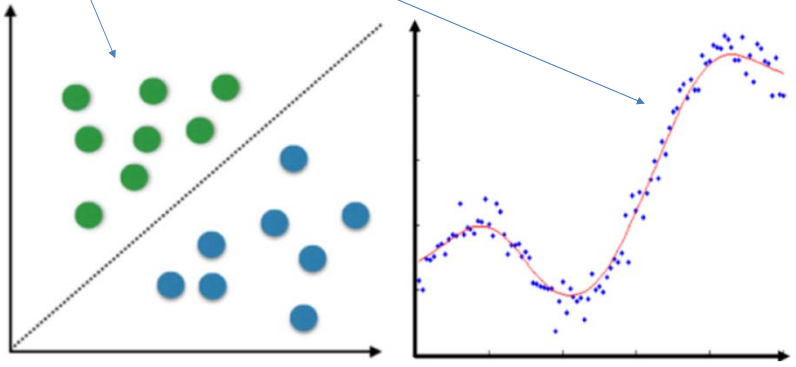
线性函数VS.非线性函数

线性回归 VS.非线性回归





分类任务与**回归任务**的区别：



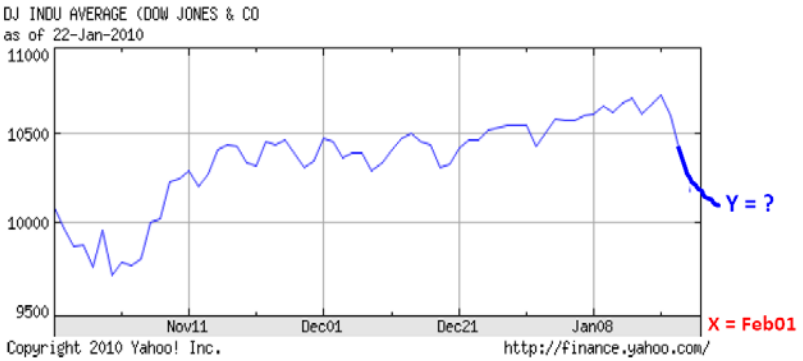
前者，给定已知类别标签(不同颜色对应不同类别)的训练集，学习一种线性分类边界，使之能有效区分不同类别训练样本；最终，基于该边界，将特征空间一分为二。

后者，给定已知标签的训练集(纵轴对应输出空间，横轴对应输入空间，样本标签值为该点纵坐标，样本点的输入对应横坐标)，学习一种预测函数；使得对应每个输入x，能有效预测对应实值输出y

例： 基于人脸图像的年龄估计(face age estimation)

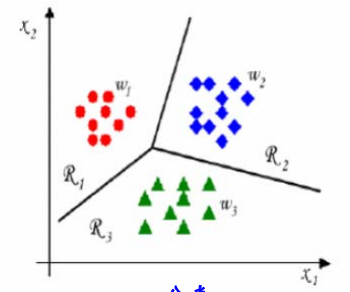
例： 时间序列预测 --Stock market prediction

基于历史观测数据$\{(t, y(t)), t = 1,...,T\}$预测未来时刻$y(T+1)$

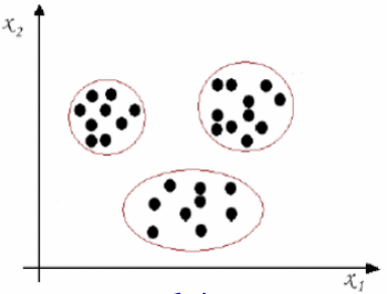


3.3 聚类(clustering)
--discovering cluster、pattern discovery

对给定的数据集进行划分，得到若干“簇”；
使得“簇内”样本之间较“簇间”样本之间更为相似。
通过聚类得到的可能各簇对应一些潜在的概念结构，
聚类是自动为给定的样本赋予标记的过程。



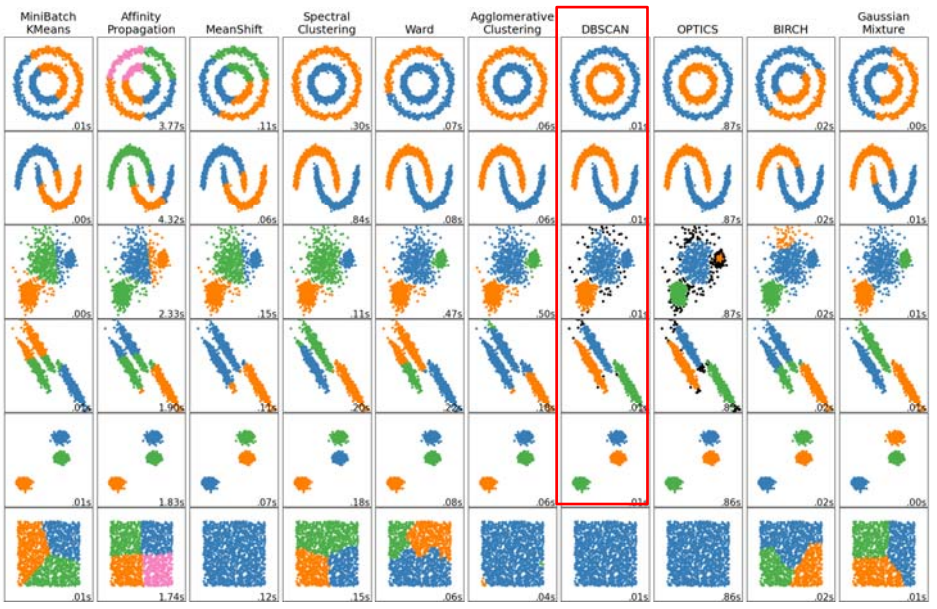
分类



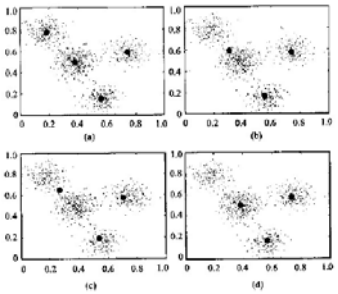
聚类

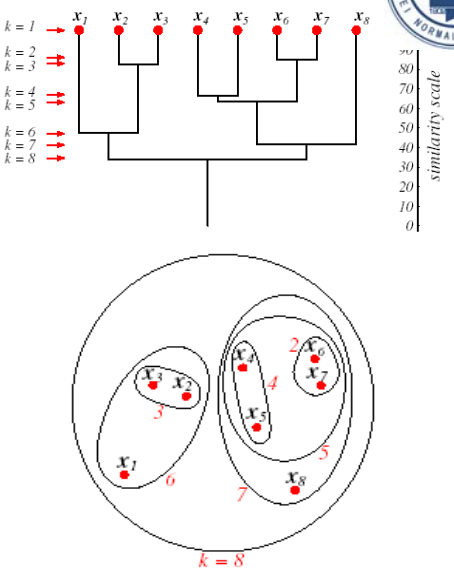
河北师范大学软件学院
Software College of Hebei Normal University

https://scikit-learn.org/stable/auto_examples/cluster/plot_cluster_comparison.html

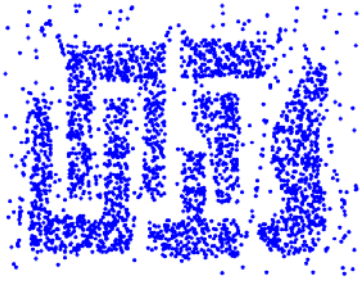


聚类举例

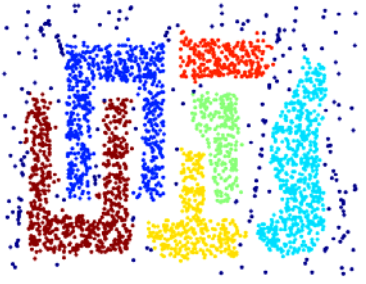




河北师范大学软件学院
Software College of Hebei Normal University



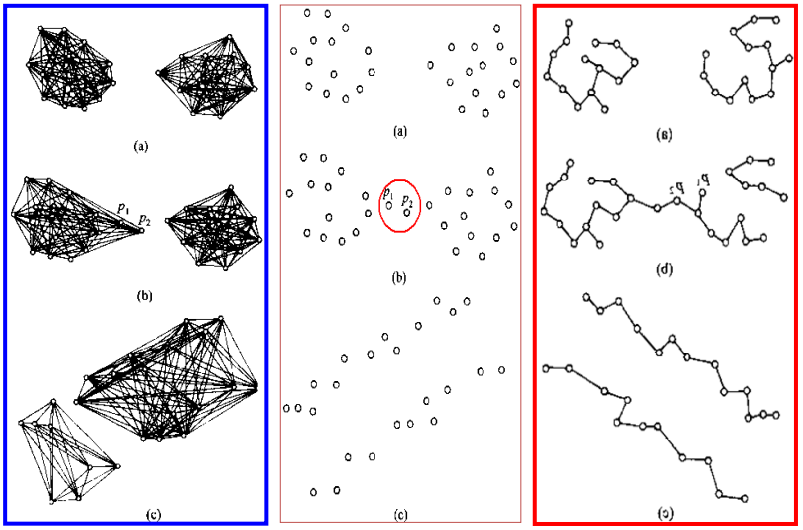
Original Points



Clusters

河北师范大学软件学院
Software College of Hebei Normal University

最远距离法与最近距离法的聚类结果比较



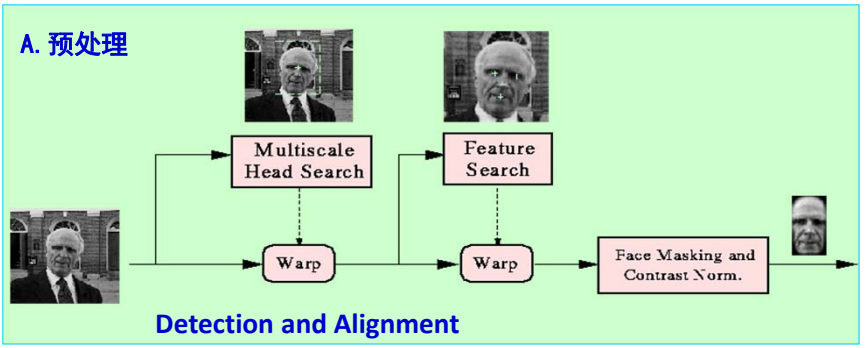
3.4 特征降维与低维可视化

将初始的数据高维表示转化为关于样本的低维表示，借助由高维输入空间向低维空间的映射，来简化输入。

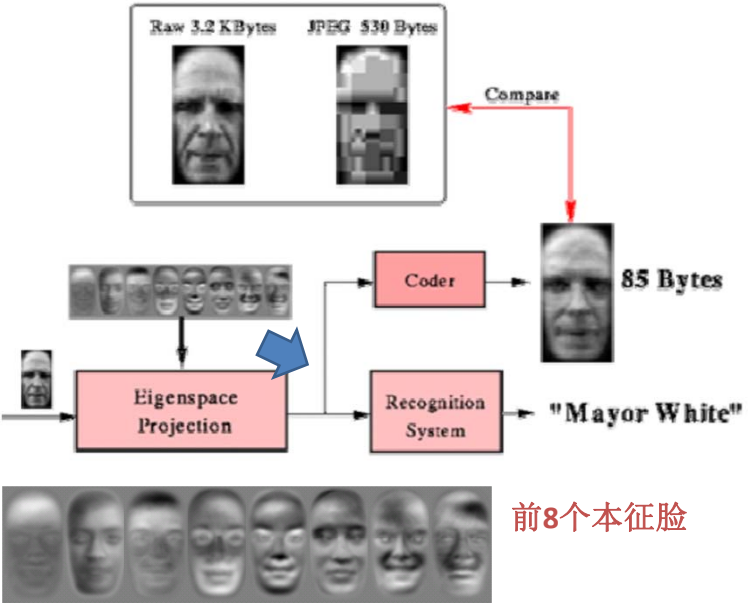
-- 特征提取

--高维数据的低维可视化

例：基于PCA的人脸表示及身份识别(注意：这是早期识别模型，仅用于举例!!!)



基于PCA的非监督式特征提取



前8个本征脸

例：高维数据的低维可视化(注意不同姿势、不同表情人脸对应的位置)

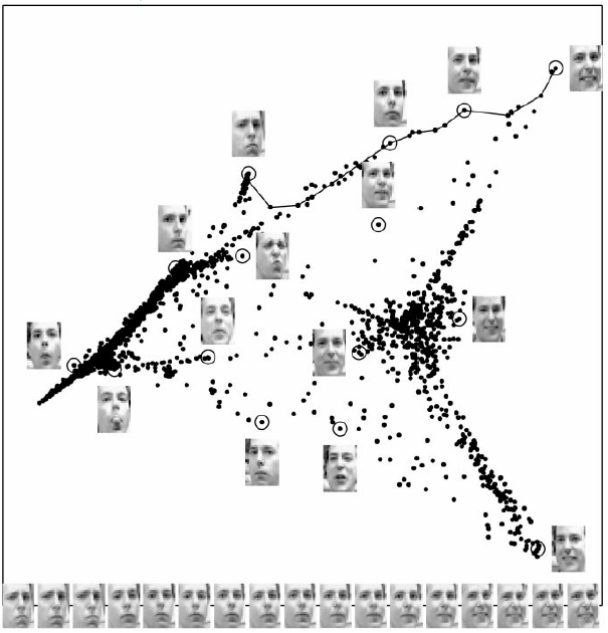
Face pose and Expression

N=2000 Images

k=12 nearest Neighbors

D=20*28=560 Pixels

d=2



主要内容

1. 什么是机器学习
2. 机器学习的相关术语
3. 机器学习的典型任务
4. 机器学习的学习范式
5. 假设与假设空间
6. 假设的选择原则
7. 机器学习的三要素

什么是机器学习的学习范式

--是关于机器学习中典型场景的表示

不同的学习范式体现在：

- 如何从数据中学习
- 如何与场景互动

→ 几种典型的学习范式

4.1 监督式学习
(supervised learning, 也称predictive learning)

--目的在于精确预测

--“预测性能”

--例：面向分类模型、回归模型的学习

基于**已知标签的数据集**学习预测模型，基于该模型对未知样本的输出做出预测。

利用给定的训练集$\boldsymbol{D}=\{(\boldsymbol{x}_i,\boldsymbol{y}_i),\boldsymbol{i}=1,...,\boldsymbol{N}\}$，学习输入$\boldsymbol{x}$与输出$\boldsymbol{y}$的关系，使得对于任意观测$\boldsymbol{x}^*$，该模型尽可能准确预测输出$\boldsymbol{y}^*$

4.2 非监督式学习 (unsupervised learning, 也称 descriptive learning)

-- 发现关于数据的紧致描述、知识发现

-- “描述性能”

算法基于无标签样本集进行模型学习，基于学得的模型对所有未知样本做出预测。

基于给定数据集 $\boldsymbol{D}=\{\boldsymbol{x}_i, \boldsymbol{i}=1,...,\boldsymbol{N}\}$, 寻找关于 $\boldsymbol{D}$ 的更为紧致的描述

4.3 强化学习 (Reinforcement Learning, RL)

- 借助**智能体**与**环境**的连续互动，学习最优行为策略
- 以试错方式，使**智能体**学得当前**环境状态**到**行为**的映射，使得智能体能结合环境状态，选择能够获得**环境最大奖赏**的行为。

- 智能体
- 环境
- 行为、行为空间
- 状态、状态空间
- 奖励

- 结合给定的奖惩机制，算法学习如何与环境交互，以便智能体对环境采取更好的动作行为。
- 典型应用：下棋、无人驾驶

4.4 其它

(1)半监督学习(Semi-Supervised Learning)

基于少量有标签样本(标注成本高)、大量无标签样本(获取容易),学习输入到输出的预测模型。

充分利用无标签样本的信息，辅助有标签的样本，进行监督学习

以较低成本获得较好的学习效果。

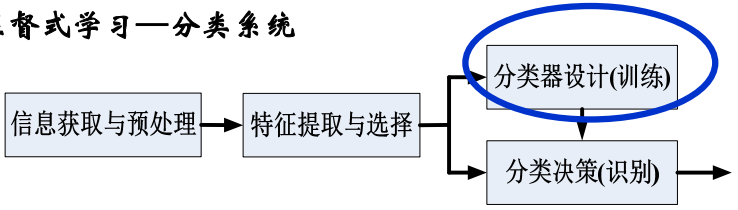
(2)迁移学习(Transfer Learning)

--当在某些领域无法取得规模充足的样本进行模型学习时，基于另一领域样本数据获得的模型进行的学习。

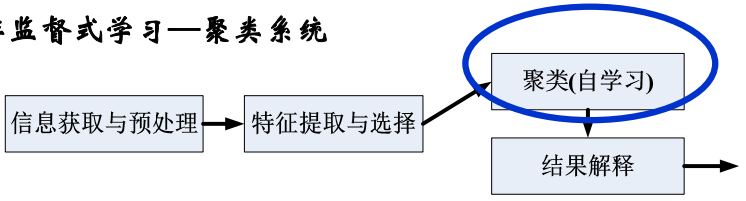
--迁移学习可以把已训练好的模型(“预训练模型”)参数迁移到新的模型，以指导新模型训练，从而更有效地学习底层规则、减少新模型学习时关于样本量的需求。

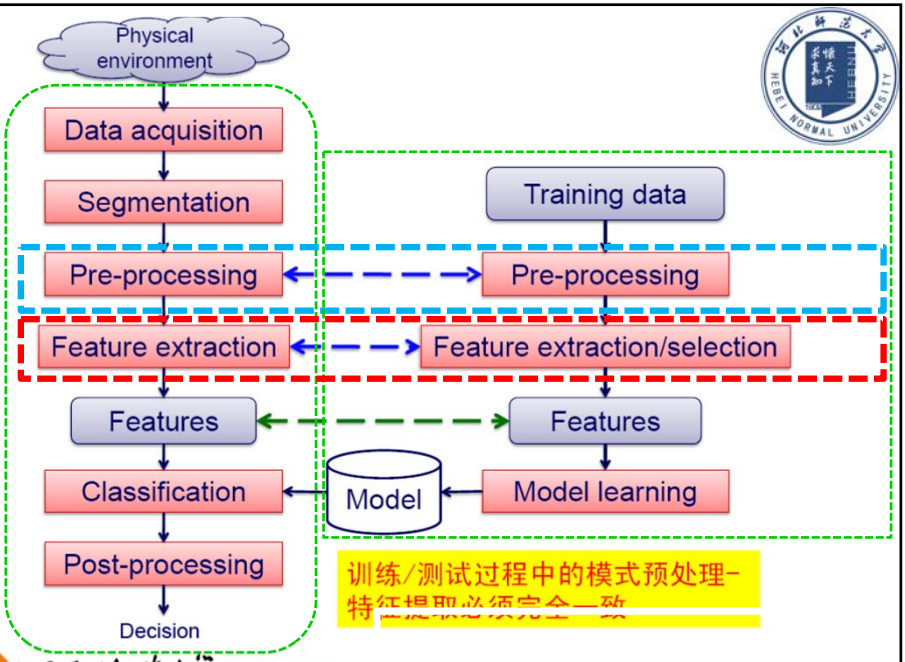
例：典型的机器学习系统

监督式学习—分类系统



非监督式学习—聚类系统





主要内容

1. 什么是机器学习
2. 机器学习的相关术语
3. 机器学习的典型任务
4. 机器学习的学习范式
5. 假设与假设空间
6. 假设的选择原则
7. 机器学习的三要素

➢ 假设(hypothesis)、假设空间(hypothesis space)

每一个具体的模型就是一个“假设(hypothesis)”

模型的学习过程就是一个在所有假设构成的假设空间进行搜索的过程，搜索的目标就是找到与训练集“匹配(fit)”的假设。

➢ 版本空间(version space)

基于有限规模的训练样本集进行假设的匹配搜索，会在假设空间存在多个假设与训练集一致的情况，称这些假设组成的集合为“版本空间”

主要内容

1. 什么是机器学习
2. 机器学习的相关术语
3. 机器学习的典型任务
4. 机器学习的学习范式
5. 假设与假设空间
6. 假设的选择原则
7. 机器学习的三要素

- 主要原则

----“奥克姆剃刀(Occam's Razor)”准则

如无必要，勿增实体

若多个假设与经验观测一致，则选择最简单的那个

- 其它原则

----“多释原则”

保留与经验观察一致的所有假设

(与集成学习的思想一致)

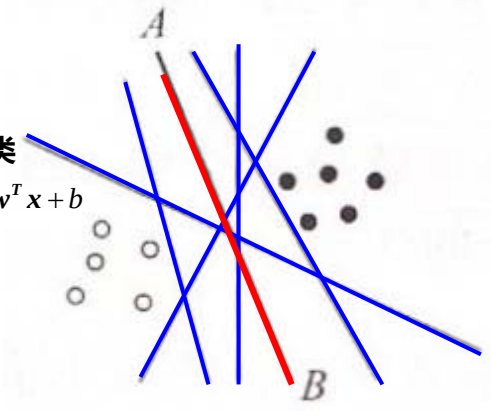
主要内容

1. 什么是机器学习
2. 机器学习的相关术语
3. 机器学习的典型任务
4. 机器学习的学习范式
5. 假设与假设空间
6. 假设的选择原则
7. 机器学习的三要素

机器学习**方法**由**模型**、**策略**、**算法**构成
机器学习的三要素：**方法**=**模型**+**策略**+**算法**

(1)模型

首先明确需要学习什么样的模型



例　d维特征空间的两类别线性分类

判别函数(决策函数)　　$f(x)=w^T x+b$

待估计的参数　$\theta=\left[w^T,b\right]^T$

其中$w\in R^d, b\in R$

假设空间为线性判别函数的集合

$\mathcal{F}=\{f\mid Y=f(X)\}$

分类边界是关于x的线性方程

$w^T x+b=0$

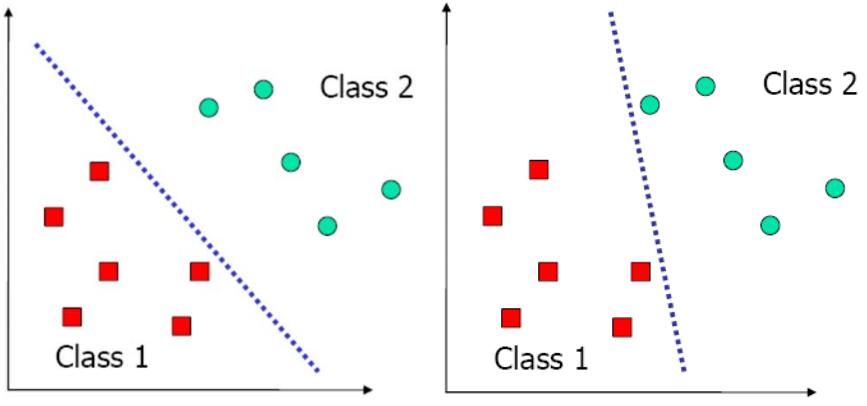
$\omega_1 x_1+\omega_2 x_2+b=0$

函数f依赖于参数向量$\theta=\left[w^T,b\right]^T$

假设空间等价于参数向量决定的函数组

$\mathcal{F}=\{f\mid Y=f_\theta(X),\theta\in R^{d+1}\}$

- **较差的分类边界**



- **较好的分类边界：分类间隔（margin）尽可能大**
  分类边界尽可能远离两类数据

(2)策略

机器学习的目标在于从**假设空间**中选取**最优模型**

策略就是确定**基于什么样的准则**，学习或选择最优模型。

实质：面向具体模型的学习，确定准则函数

准则函数(也称损失函数、代价函数、目标函数)

为便于理解，引出如下概念

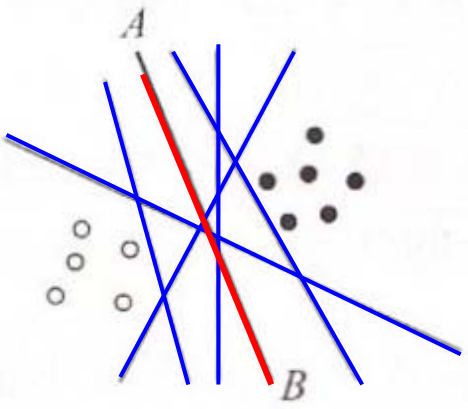
A.损失函数

B.期望风险

C.经验风险

D.经验风险最小化

E.结构风险最小化



A.损失函数(loss function)或代价函数(cost function)

Y ----输入$\boldsymbol{X}$对应的目标答案

$f(\boldsymbol{X}), P(Y \mid \boldsymbol{X})$ ----模型的预测输出

0-1损失函数 $$L(\boldsymbol{Y}, \boldsymbol{f}(\boldsymbol{X})) = \begin{cases} 1 & Y \neq f(\boldsymbol{X}) \\ 0 & Y = f(\boldsymbol{X}) \end{cases}$$ 分类，预测为类别标号

平方损失函数 $$L(\boldsymbol{Y}, \boldsymbol{f}(\boldsymbol{X})) = \left[Y - f(\boldsymbol{X})\right]^2$$

绝对损失函数 $$L(\boldsymbol{Y}, \boldsymbol{f}(\boldsymbol{X})) = \left|Y - f(\boldsymbol{X})\right|$$ 回归，预测为实数值

对数损失函数或对数似然损失函数

$$L(\boldsymbol{Y}, \boldsymbol{f}(\boldsymbol{X})) = -\log P(Y \mid \boldsymbol{X})$$ 分类，预测为后验概率

损失函数值越小，模型越好

B.期望风险 $(expected\ risk)$

模型的输入$\boldsymbol{X}$与输出Y构成输入空间$\mathcal{X}$ 与输出空间$\mathcal{Y}$的联合随机变量$(\boldsymbol{X}, Y)$，遵循联合分布$P(\boldsymbol{X}, Y)$

$\Rightarrow$ 损失函数$L(\boldsymbol{Y}, \boldsymbol{f}(\boldsymbol{X}))$是关于联合随机变量$(\boldsymbol{X}, Y)$的函数

$\Rightarrow$ **期望风险**R_{exp}就是损失函数$L(\boldsymbol{Y}, \boldsymbol{f}(\boldsymbol{X}))$的数学期望。

$$R_{exp} = \boldsymbol{E}_P\left[L(\boldsymbol{Y}, \boldsymbol{f}(\boldsymbol{X}))\right] = \int_{\mathcal{X}\times\mathcal{Y}} L(y, \boldsymbol{f}(\boldsymbol{x}))P(\boldsymbol{x}, y)d\boldsymbol{x}dy$$

机器学习的目标就在于选择具有最小期望风险的模型。

但因联合分布$P(\boldsymbol{X}, Y)$未知，难以计算**期望风险**R_{exp}

C.经验风险 $(empirical\ \ risk)$或经验损失 $(empirical\ \ loss)$

给定训练样本集$T = \{(\boldsymbol{x}_i, y_i), \boldsymbol{i} = 1, ..., \boldsymbol{N}\}$，模型$f(\boldsymbol{X})$关于**训练样本集**$T$**的平均损失**，称为**经验风险**，记为$R_{emp}$

$$R_{emp} = \frac{1}{\boldsymbol{N}}\sum_{\boldsymbol{i}=1}^{\boldsymbol{N}} L(y_i, \boldsymbol{f}(\boldsymbol{x}_i))$$

当容量$\boldsymbol{N}$趋于无穷时，经验风险R_{emp}趋于期望风险R_{exp}

实际问题中，训练样本数目N非常有限，需对经验风险矫正

D.经验风险最小化(*empirical risk minimization, ERM*)

在假设空间、损失函数形式、以及训练样本集确定的前提下，**"经验风险最小化"策略**认为：假设空间$\mathcal{F}$ 中，使**经验风险**R_{emp}最小的模型，就是最优模型。

最优化问题为：$\min_{f\in\mathcal{F}} R_{emp}(f)$

其中：$R_{emp}(f)=\frac{1}{N}\sum_{i=1}^{N}L(y_i,f(x_i))$

N足够大时，采用**"经验风险最小化"策略**，可获得较好学习效果；

N很小时，该策略的学习效果未必好，易产生"过拟合(*overfitting*)"

E.结构风险最小化(*structural risk minimization, SRM*)

为防止模型过拟合，提出**"结构风险最小化"策略**

给定{ 假设空间$\mathcal{F}$；损失函数$L(Y,f(X))$形式；训练样本集$T=\{(x_i,y_i),i=1,...,N\}$

结构风险 = 经验风险 + 表示模型复杂程度的正则化项

$$R_{srm}(f)=R_{emp}(f)+\lambda J(f)$$ 惩罚系数$\lambda\geq 0$

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

$J(f)$表示模型f的复杂度：模型f越复杂，$J(f)$值越大

$R_{emp}(f)$度量模型f关于训练集的学习能力

$R_{emp}(f)$值越小，模型f关于训练集的学习能力越好

E.结构风险最小化(*structural risk minimization, SRM*)

"结构风险最小化"策略认为：假设空间$\mathcal{F}$ 中，使**结构风险**R_{srm}最小的模型，就是最优模型。

最优化问题：$\min_{f\in\mathcal{F}} R_{srm}(f)$

$$R_{srm}(f)=R_{emp}(f)+\lambda J(f)=\frac{1}{N}\sum_{i=1}^{N}L(y_i,f(\boldsymbol{x}_i))+\lambda J(f)$$

河北师范大学软件学院
Software College of Hebei Normal University

"策略"部分的工作：

基于给定的假设空间、损失函数的形式、训练集形成用于模型求解的**最优化问题**：

"经验风险最小化"策略 $\min_{f\in\mathcal{F}} R_{emp}(f)$

"结构风险最小化"策略 $\min_{f\in\mathcal{F}} R_{srm}(f)$

(3)算法

"算法"是指采用何种算法，求解最优化问题。

解析解 **or** 数值解

全局最优 **or** 局部最优

寻优过程是否高效？

例如：基于给定的训练集进行M次多项式函数拟合

训练集 $D_{\text{train}}=\{(x_i,y_i),i=1,\ldots,N_{\text{train}}\}$

M次多项式函数：

$f_M(x;\boldsymbol{\omega})=\omega_0+\omega_1x+\omega_2x^2+\cdots+\omega_Mx^M=\sum_{j=0}^{M}\omega_jx^j$

请结合上述任务，给出机器学习的三要素？

M阶多项式的拟合，是监督式学习，属“回归”任务。

(1)模型

$f_M(x;\boldsymbol{\omega})=\omega_0+\omega_1x+\omega_2x^2+\cdots+\omega_Mx^M=\sum_{j=0}^{M}\omega_jx^j$

(2)策略：结构风险最小化策略

$$R_{\text{srm}}(f)=R_{emp}(f)+\lambda J(f)$$

$$\boldsymbol{F}(\boldsymbol{\omega})=\frac{1}{N_{\text{train}}}\sum_{i=1}^{N_{train}}\boldsymbol{L}\left(y_i,\boldsymbol{f}(x_i;\boldsymbol{\omega})\right)+\lambda\boldsymbol{J}(\boldsymbol{\omega})=\frac{1}{N_{\text{train}}}\sum_{i=1}^{N_{train}}\left[\left(y_i-\sum_{j=0}^{M}\omega_jx_i^{\,j}\right)^2\right]+\lambda\sum_{j=1}^{M}\omega_j^{\,2}$$

$$\boldsymbol{\omega}^*=\arg\min_{\boldsymbol{\omega}}\boldsymbol{F}(\boldsymbol{\omega})$$

(3)算法：

梯度下降法(数值优化、局部寻优)　$\boldsymbol{\omega}(t+1)=\boldsymbol{\omega}(t)-\eta\nabla\boldsymbol{F}(\boldsymbol{\omega}(t))$

本讲小结

1. 什么是机器学习？典型学习任务？学习范式？
2. 机器学习的相关术语

   特征、特征向量、特征空间、特征维数

   样本、样本集、训练集、测试集、估计集、验证集

   有标签样本、无标签样本
3. 模型、假设、假设空间；假设的选择原则
4. 机器学习的三要素(模型、策略、算法)

   损失函数、经验风险、期望风险

   经验风险最小化策略、结构风险最小化策略