第一章 绪论

机器学习定义

吃瓜群众希望根据经验从无数不同形态特点的瓜中选择出一个好瓜，于是把 **经验** 作为 **数据** 输入计算机，学习经验数据生成 **算法模型**，并作出判断。

重要会议

- 国际机器学习会议 **ICML**
- 国际神经信息处理系统会议 **NIPS**
- 国际学习理论会议 **COLT**
- 欧洲机器学习会议 **ECML**
- 亚洲机器学习会议 **ACML**
- 人工智能领域会议 **IJCAI**， **AAAI**
- 数据挖掘领域会议 **KDD**， **ICDM**
- 机器视觉模式识别领域 **CVPR**
- 中国机器学习大会 **CCML**
- 机器学习及应用研讨会 **MLA**

重要期刊

- 国际学术期刊 Journal of Machine Learning Research , Machine Learning
- 人工智能领域 Artifical Intelligence , Journal of Artificial Intelligence Research
- 数据挖掘领域 ACM Transaction on Knowledge Discovery from Data , Data Mining and Knowledge Discovery
- 计算机视觉与模式识别 IEEE Transaction on Pattern Analysis and Machine Intelligence
- 神经网络 Neural Computation, IEEE Transactions on Neural Networks and Learning Systems
- 统计学 Annals of Statustion

基本术语

- `sample`

  样本，也叫示例 `instance`

  如描述一个瓜的特性的一条记录：（色泽=青绿；根蒂=蜷缩；敲声=浊响）

- `attribute`

  属性，或 `feature`，如色泽、根蒂、敲声

- `attribute value`

  属性值，不多解释

- `data set`

  数据集，由样本组成

  $D = \{x_1, x_2, \ldots, x_m\}$

- `attribute space`

  属性空间，或样本空间 `sample space`，作为输入空间。比如西瓜的三个属性分别作为空间上的 x, y, z 轴形成的一个三维空间

- `feature vctor`

  特征向量，在样本空间中每个样本对应一个

  $x_i = \{x_{i1}, x_{i2}, \ldots, x_{id}\}$, xi 的维度 `dimensionality` 为 `d`

- 机器学习基本过程为学习或训练。使用由训练数据 `training data` 中的训练样本 `trainning sample` 组成的训练集 `training set` 进行训练。

- `learner`

  学习器，学习训练得到的模型

- 学习训练得到的模型对应某种潜在的规律 —假设`hypohesis`或真实、真相 `ground-tuth`。

- `lable`

  标记，关于训练结果好坏的信息。比如（（色泽=青绿；根蒂=蜷缩；敲声=浊响）好瓜）

- `example`

  拥有标记的样本 $(x_i, y_i) y_i \in Y$， Y 为标记空间/输出空间 `lable space`

- `classification`

  分类。预测 `prediction` 的是 **离散值**，如好瓜、坏瓜。

  可以分为二分类 `binary classification` 和多分类 `multi-class classification`，二分类中有正类 `positive class` 和反类 `negative class`。

- `regression`

  回归。预测的是 **连续值**，如瓜的成熟度0.95，0.37.

- `clustering`

  聚类。把训练集中的西瓜分为若干组/簇 `cluster`.

- 监督学习 `supervised learning`

  包括 **分类** 和 **回归**，有 **标记样本**

- 无监督学习 `unsupervised learning`

  如 **聚类**， 没有标记样本信息

- `generalization`

  泛化能力，在无监督学习中模型适用于新样本的能力

- 假设空间

  学习就是一个在假设空间中搜索，找到与训练集匹配 `fit` 的假设的过程

- 版本空间

  可能有多个**假设与训练集一致**，这些假设集合为版本空间

- 与训练集一致的假设

  表示能对训练集中所有的样本进行正确的判断

- 归纳偏好

  对于多个版本空间，需要选择一个更加合适的，最常见的是选择图线更加平滑的，算法有奥卡姆剃刀 `Occam's razor` 等

对于一个预测任务，对训练集$\{(x_1,y_1),(x_2,y_2),\dots,(x_i,y_i)\}$进行学习，建立输入空间 **X** 到输出空间 **Y** 的映射 $f:X\mapsto Y$,得到模型之后对被预测样本 `testing sample` **x** 进行测试得到其预测标记 **y = f(x)**。

没有免费午餐定理NFL

考虑到两种策学习算法 $\mathcal{E}_a,\mathcal{E}_b$，其中前者较为优越，后者为"随机胡猜"。

- 样本空间 χ 和假设空间 $\mathcal{H}$ 都是离散的；
- $P(h|X,\mathcal{E}_a)$: 算法 **a** 给予训练数据 **X** 产生假设 **h** 的概率；
- f : 希望学习的真实目标函数；
- $\mathbb{I}(\cdot)$: 只是函数，真为1假为0；

则 $\mathcal{E}_a$ 的 **训练集外误差** ——在训练机之外所有样本上的误差为

$$E_{ote}(\mathfrak{L}a|X,f)=\sum_h\sum_{x\in\mathcal{X}-X}P(x)\mathbb{I}(h(x)\neq f(x))P(h|X,\mathfrak{L}a)$$

对于二分问题，f 按照均匀分布对误差求和：

$$\begin{aligned}\sum_f E_{ote}(\mathfrak{L}a|X,f)&=\sum_f\sum_h\sum_{x\in\mathcal{X}-X}P(x)\mathbb{I}(h(x)\neq f(x))P(h|X,\mathfrak{L}a)\\&=\sum x\in\mathcal{X}-XP(x)\sum_h P(h|X,\mathfrak{L}a)\sum_f\mathbb{I}(h(x)\neq f(x))\\&=\sum x\in\mathcal{X}-XP(x)\sum_h P(h|X,\mathfrak{L}a)\frac{1}{2}2^{|\mathcal{X}|}\\&=\frac{1}{2}2^{|\mathcal{X}|}\sum x\in\mathcal{X}-XP(x)\sum_h P(h|X,\mathfrak{L}a)\\&=2^{|\mathcal{X}|-1}\sum x\in\mathcal{X}-XP(x)\cdot 1\end{aligned}$$

所以 **总误差与算法无关！**，它们的期望相同，但是是在 **均匀分布** 的前提下。所以 **NFL** 告诉我们 **学习算法自身与归纳偏好相匹配起到决定性作用**。

习题

1.1

若只包（青绿，蜷缩，浊响，是）、（乌黑、稍蜷，沉闷，否）的两个样例，试给出相应的版本空间

首先 3 * 3 * 3 + 1 = 28 个，然后枚举一下，把假设空间中能对所有的样本进行正确的判断的保留，剩下

色泽=青绿 根蒂=蜷缩 敲声=浊响

色泽=青绿 根蒂=蜷缩 敲声=*

色泽=青绿 根蒂=* 敲声=浊响

色泽=* 根蒂=蜷缩 敲声=浊响

色泽=青绿 根蒂=* 敲声=*

色泽=* 根蒂=蜷缩 敲声=*

1.3

若数据包含噪声，则假设空间中可能不存在与所有训练样本都一致的假设。在此情形下，试设计一种归纳偏好用于假设选择

去掉一些有欧相同属性却不同分类的数据

1.4

本章1.4节在论述“没有免费的午餐”定理时，默认使用了“分类错误率”作为性能度量来对分类器进行评估。若换用其他性能度量l,试证明没有免费的午餐”定理仍成立

证明：NFL首先要保证目标函数均匀分布...