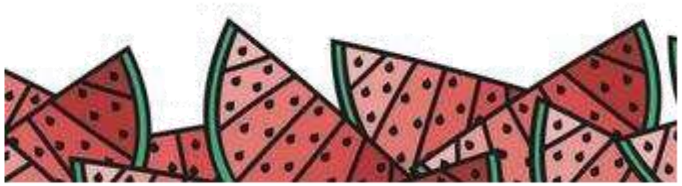
MACHINE LEARNING

机器学习

Bayesian Classification

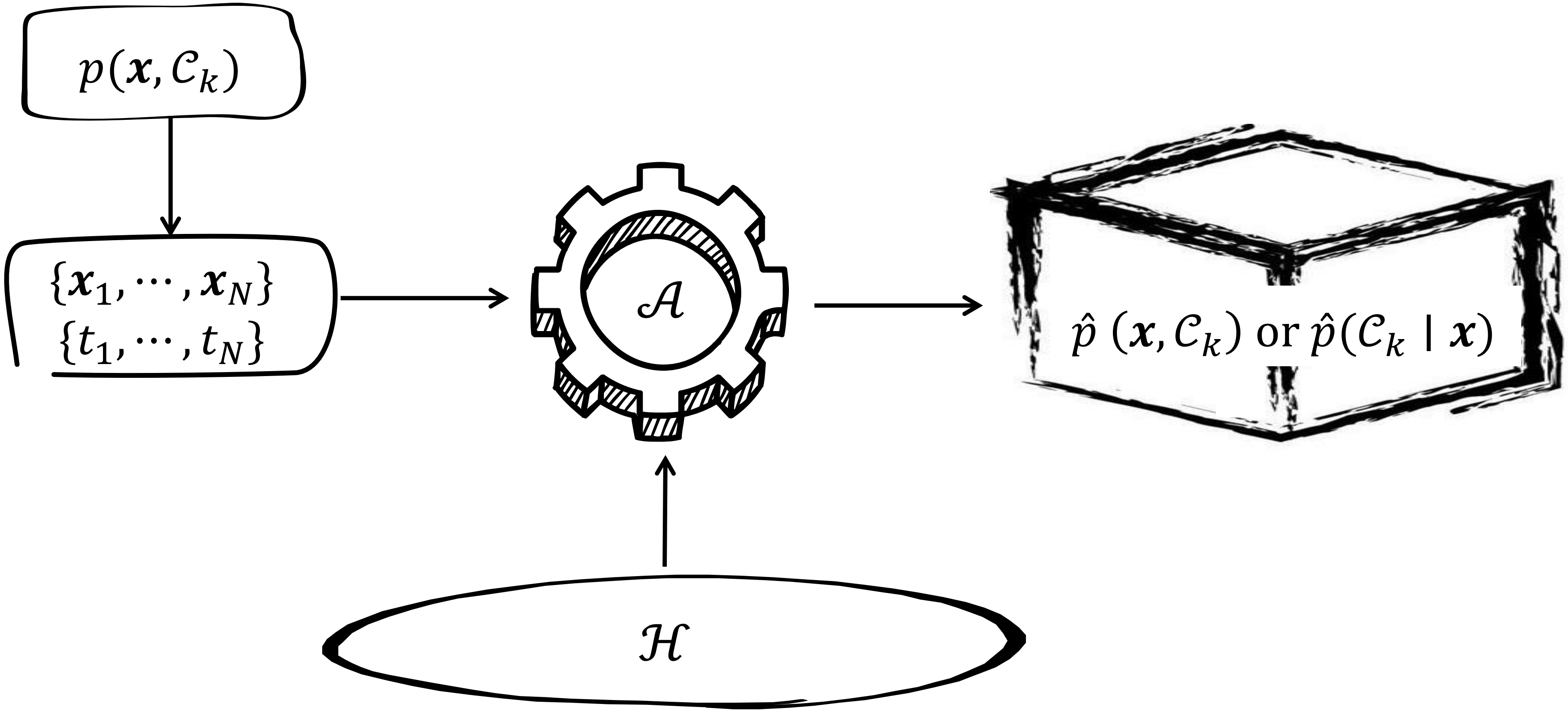
贝叶斯分类



Machine Learning Course

概率论与机器学习

- 分类任务的概率体系框架



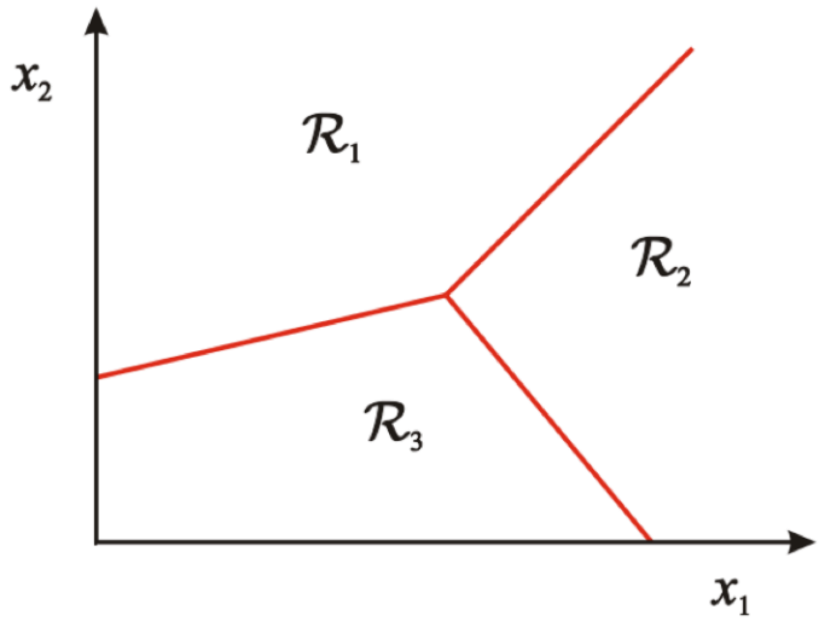
分类任务的决策论

- 关于分类任务的决策理论

- 如何做出最优的分类决策（二分类 $\{\mathcal{C}_k\} = \{-1, +1\}$情况）

$$
\begin{aligned}
p(\text{ mistake }) &= p(\boldsymbol{x} \in \mathcal{R}_1, \mathcal{C}_2) + p(\boldsymbol{x} \in \mathcal{R}_2, \mathcal{C}_1) \\
&= \int_{\mathcal{R}_1} p(\boldsymbol{x}, \mathcal{C}_2) \mathrm{d}\boldsymbol{x} + \int_{\mathcal{R}_2} p(\boldsymbol{x}, \mathcal{C}_1) \mathrm{d}\boldsymbol{x}
\end{aligned}
$$

$$
\begin{aligned}
&\mathcal{R}_1 = \{\boldsymbol{x} \mid p(\boldsymbol{x}, \mathcal{C}_1) > p(\boldsymbol{x}, \mathcal{C}_2)\} \\
&\mathcal{R}_2 = \{\boldsymbol{x} \mid p(\boldsymbol{x}, \mathcal{C}_1) \leqslant p(\boldsymbol{x}, \mathcal{C}_2)\} \\
&p(\boldsymbol{x}, \mathcal{C}_k) = p(\mathcal{C}_k \mid \boldsymbol{x}) p(\boldsymbol{x}) \\
&\mathcal{R}_k = \{\boldsymbol{x} | p(\mathcal{C}_k \mid \boldsymbol{x}) \text{ 是最大的 }\}
\end{aligned}
$$



分类任务的决策论

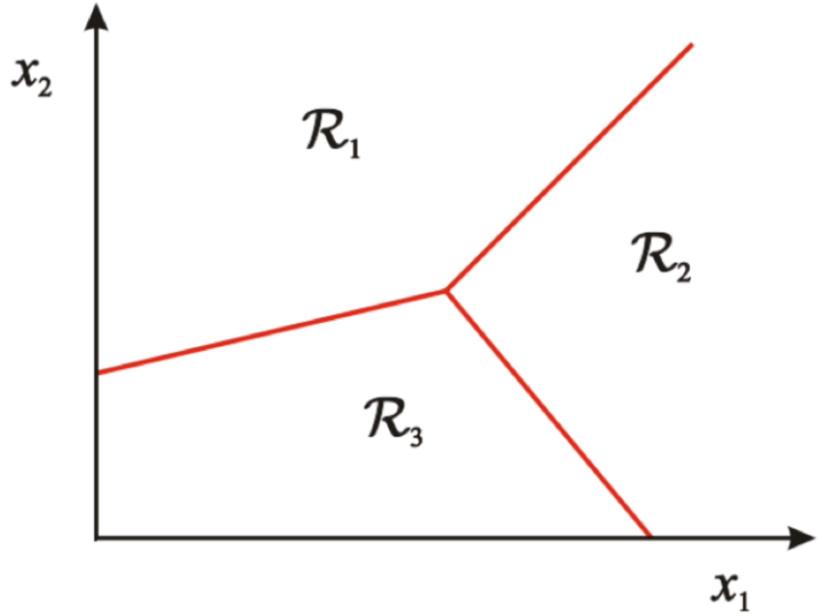
- 关于分类任务的决策理论

- 如何做出最优的分类决策（多分类$\{\mathcal{C}_k\} = \{1, \ldots, K\}$情况）

$$p(\text{ mistake }) = \sum_{k=1}^{K} \sum_{j \neq k} p(\boldsymbol{x} \in \mathcal{R}_j, \mathcal{C}_k) = \sum_{k=1}^{K} \sum_{j \neq k} \int_{\mathcal{R}_j} p(\boldsymbol{x}, \mathcal{C}_k) \mathrm{d}\boldsymbol{x}$$

$$p(\text{ correct }) = \sum_{k=1}^{K} p(\boldsymbol{x} \in \mathcal{R}_k, \mathcal{C}_k) = \sum_{k=1}^{K} \int_{\mathcal{R}_k} p(\boldsymbol{x}, \mathcal{C}_k) \mathrm{d}\boldsymbol{x}$$

$$\mathcal{R}_k = \{\boldsymbol{x} | p(\mathcal{C}_k \mid \boldsymbol{x}) \text{ 是最大的 }\}$$

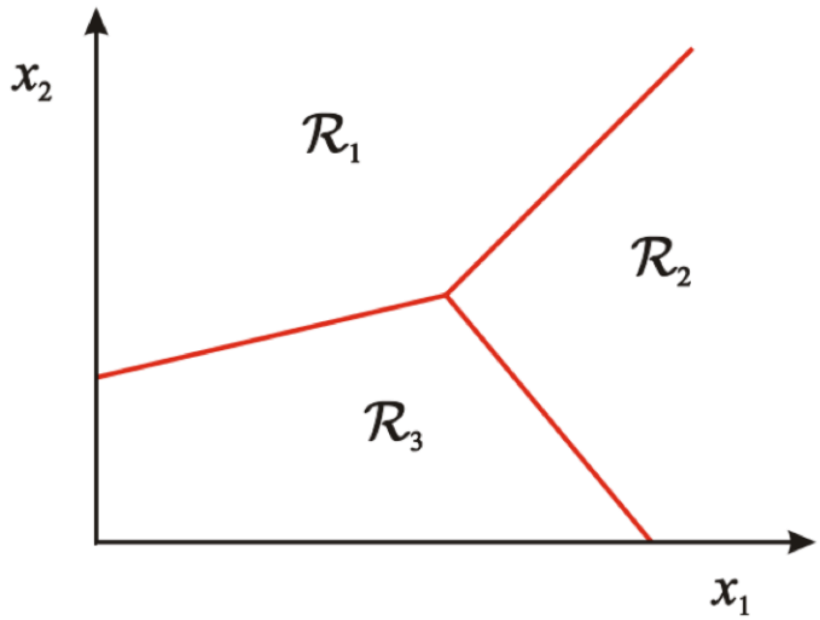


分类任务的决策论

- 关于分类任务的决策理论
- 如何做出最优的分类决策（带损失L_{kj}情况下）

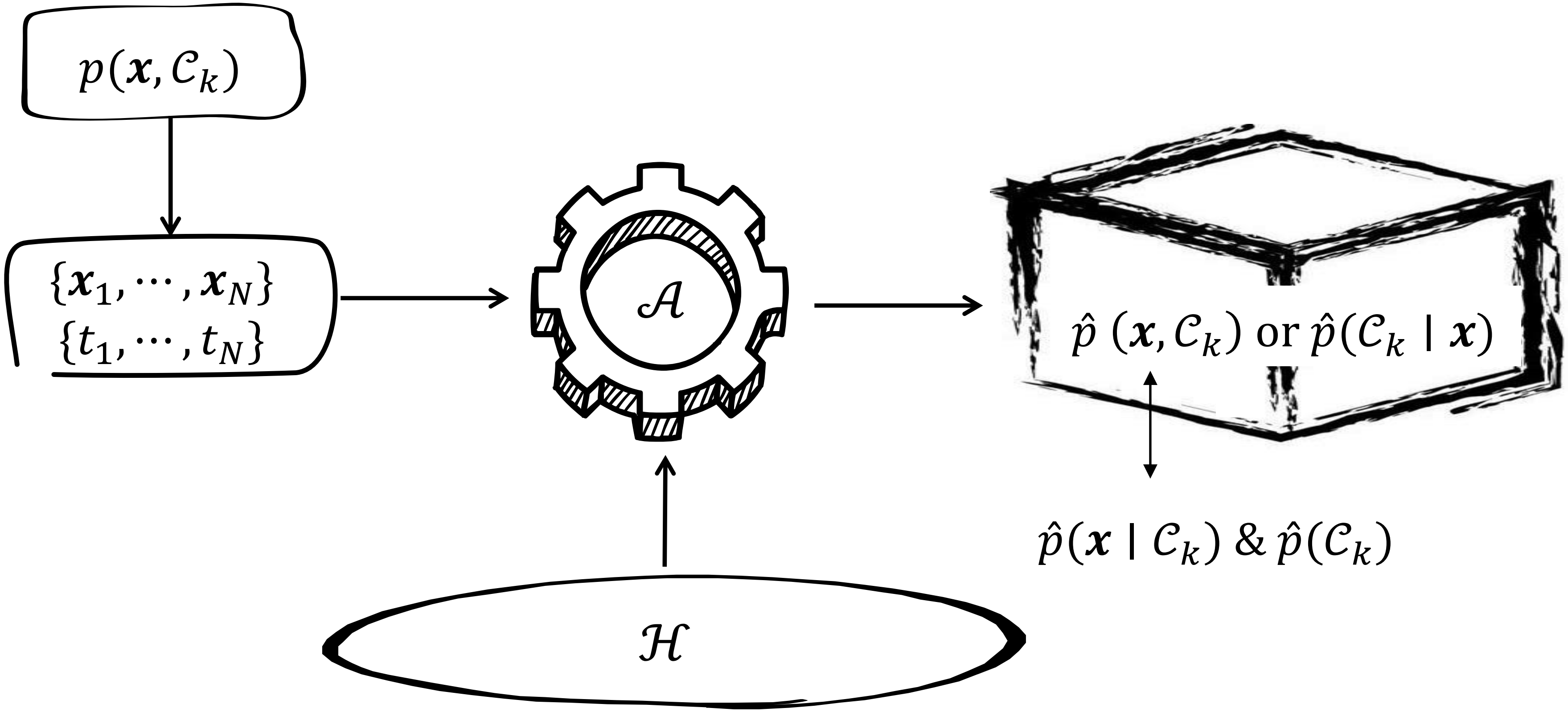
$\boldsymbol{x}$分类为k类的损失为$\sum_{\mathrm{j}=1}^{C} L_{\mathrm{kj}}\, p(\mathcal{C}_{\mathrm{j}} \mid \boldsymbol{x})$

$$\mathcal{R}_k = \{\boldsymbol{x} | \sum_{\mathrm{j}=1}^{C} L_{\mathrm{kj}}\, p(\mathcal{C}_{\mathrm{j}} \mid \boldsymbol{x})\text{是最小的}\}$$



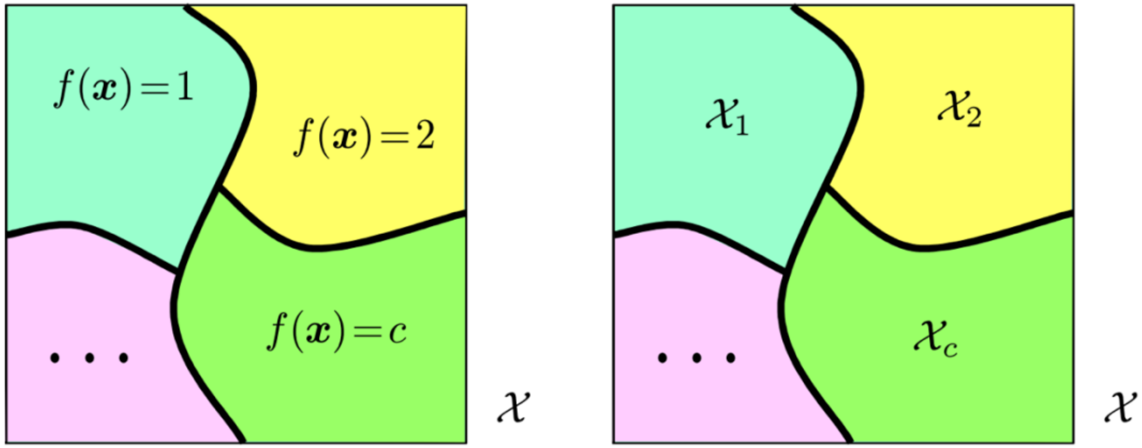
概率论与机器学习

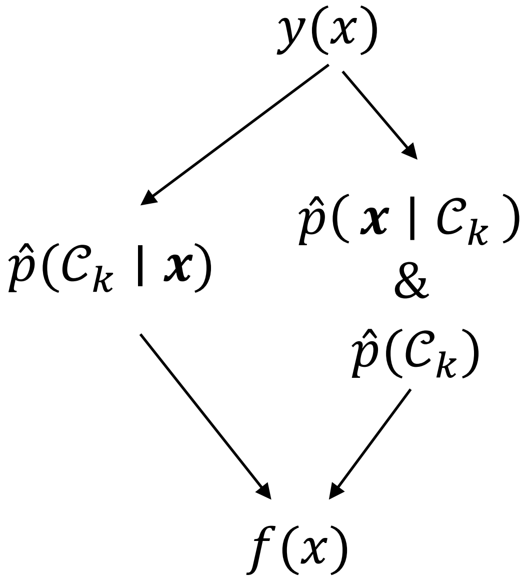
- 生成式算法与判别式算法



分类任务的决策论

- 生成式算法与判别式算法





朴素贝叶斯

- 动机

估计后验概率$P(c \mid \boldsymbol{x})$的主要困难在于: 类条件件概率$P(\boldsymbol{x} \mid c)$ 是所有属性上的联合概率，难以从有限的训练样本直接估计而得.

- 想法

假设所有属性在给定类别的条件下相互独立 —属性条件独立性假设”
attribute conditional independence assumption

$$\hat{p}(c \mid \boldsymbol{x}) = \frac{\hat{p}(c)\hat{p}(\boldsymbol{x} \mid c)}{\hat{p}(\boldsymbol{x})} = \frac{\hat{p}(c)}{\hat{p}(\boldsymbol{x})} \prod_{i=1}^{d} \hat{p}(x_i \mid c)$$

$$h_{nb}(\boldsymbol{x}) = \arg\max_{c \in \mathcal{Y}} \hat{p}(c) \prod_{i=1}^{d} \hat{p}(x_i \mid c)$$

$$\hat{p}(c) = \frac{|D_c|}{|D|}$$

朴素贝叶斯

- 单变量类条件概率$p(x_i \mid c)$ 的估计

- 离散变量情况

$$\hat{p}(x_i \mid c) = \frac{|D_{c,x_i}|}{|D_c|} \xrightarrow[\text{Laplacian correction}]{\text{smoothing}} \hat{p}(x_i \mid c) = \frac{|D_{c,x_i}| + 1}{|D_c| + N_i}$$

$$\left(\hat{p}(c) = \frac{|D_c|}{|D|} \xrightarrow[\text{Laplacian correction}]{\text{smoothing}} \hat{p}(c) = \frac{|D_c| + 1}{|D| + N} \right)$$

- 连续变量情况

$$\hat{p}\,(x_i \mid c) \sim \mathcal{N}(\mu_{c,i}, \sigma_{c,i}^2)$$

生成式算法与判别式算法

- 朴素贝叶斯 vs 逻辑回归

不同于逻辑回归，朴素贝叶斯等方法是通过刻画先验与似然来实现贝叶斯分类的。
这引出了贝叶斯算法的两种类型：
一种是直接对后验概率$p(\mathcal{C}_k \mid \boldsymbol{x})$ 建模
一种是对联合概率$p(\boldsymbol{x}, \mathcal{C}_k)$或$p(\boldsymbol{x} \mid \mathcal{C}_k)$ & $p(\mathcal{C}_k)$建模。
前者我们称为是判别式算法，后者我们称为是生成式算法