1 基础与评估

1.1 偏差-方差分解

泛化误差可分解为偏差、方差与噪声之和：$E(f;D) = \text{bias}^2(\boldsymbol{x}) + \text{var}(\boldsymbol{x}) + \varepsilon^2$。**偏差** $\text{bias}^2(\boldsymbol{x}) = (\bar{f}(\boldsymbol{x}) - y)^2$：度量学习算法的期望预测与真实结果的偏离程度（拟合能力）。**方差** $\text{var}(\boldsymbol{x}) = \mathbb{E}_D[(f(\boldsymbol{x};D) - \bar{f}(\boldsymbol{x}))^2]$：度量同样大小的训练集变动导致的性能变化（稳定性）。**噪声** ε^2：数据本身的难度。

1.2 高斯分布

$\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$，其中 $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \exp(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}))$。

1.3 性能度量

查准率 (Precision): $P = \frac{TP}{TP+FP}$ **查全率 (Recall)**: $R = \frac{TP}{TP+FN}$
F1-Score: $\frac{2\times P\times R}{P+R}$ **ROC 与 AUC**: ROC 曲线坐标为 (FPR, TPR)，AUC 为曲线下覆盖面积，衡量排序质量。

2 线性模型与 SVM

2.1 逻辑回归 (Logistic Regression)

使用对数几率函数（Sigmoid）逼近后验概率：$y = \frac{1}{1+e^{-(\boldsymbol{w}^T\boldsymbol{x}+b)}}$
对数几率 $\ln\frac{y}{1-y} = \boldsymbol{w}^T\boldsymbol{x} + b$。优化目标（最大化对数似然）：$\min_{\boldsymbol{w},b}\sum_{i=1}^m \ln(1 + e^{-y_i(\boldsymbol{w}^T\boldsymbol{x}_i+b)})$ (假设 $y_i \in \{-1, +1\}$ 或对应调整公式)。

2.2 支持向量机 (SVM)

基本型：最大化间隔 $\gamma = \frac{2}{||\boldsymbol{w}||}$。$\min_{\boldsymbol{w},b}\frac{1}{2}||\boldsymbol{w}||^2, \quad s.t.\, y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \geq 1$ **对偶问题**：引入拉格朗日乘子 $\alpha_i \geq 0$：$\max_{\boldsymbol{\alpha}}\sum_{i=1}^m \alpha_i - \frac{1}{2}\sum_{i=1}^m\sum_{j=1}^m \alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j \; s.t. \; \sum_{i=1}^m \alpha_i y_i = 0, \quad \alpha_i \geq 0$ 解得 $\boldsymbol{w} = \sum_{i=1}^m \alpha_i y_i \boldsymbol{x}_i$。支持向量满足 $\alpha_i > 0$。**软间隔**：允许部分样本出错，引入松弛变量 ξ_i 和惩罚参数 C：$\min_{\boldsymbol{w},b}\frac{1}{2}||\boldsymbol{w}||^2 + C\sum_{i=1}^m \xi_i$ **核技巧**：$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi(\boldsymbol{x}_i)^T\phi(\boldsymbol{x}_j)$，解决非线性可分。常用高斯核 $\kappa(\boldsymbol{x}, \boldsymbol{y}) = \exp(-\frac{||\boldsymbol{x}-\boldsymbol{y}||^2}{2\sigma^2})$。

3 集成学习

误差-分歧分解：$E = \bar{E} - \bar{A}$ (集成误差 = 个体平均误差-个体平均分歧)。**AdaBoost** 最小化指数损失 $L(y, f(x)) = e^{-yf(x)}$。**更新权重**：若 h_t 错误率 ϵ_t，则权重 $\alpha_t = \frac{1}{2}\ln\frac{1-\epsilon_t}{\epsilon_t}$。样本分布更新：错分样本权重增加 $D_{t+1}(\boldsymbol{x}) \propto D_t(\boldsymbol{x})e^{-\alpha_t y_i h_t(\boldsymbol{x}_i)}$。
Bagging & Random Forest Bagging: Bootstrap 采样训练基学习器，投票/平均。Random Forest: Bagging + 属性随机选择（基尼指数选择划分时只考虑随机子集）。

4 聚类

4.1 距离度量

四个性质：非负性、同一性、对称性、直递性
无序属性：令 $m_{u,a}$ 表示属性 u 上取值为 a 的样本数，$m_{u,a,i}$ 表示在第 i 个样本簇中在属性 u 上取值为 a 的样本数，k 为样本簇数，则属性 u 上两个离散值 a 与 b 之间的 VDM 距离为 $VDM_p(a,b) = \sum_{i=1}^k \left|\frac{m_{u,a,i}}{m_{u,a}} - \frac{m_{u,b,i}}{m_{u,b}}\right|^p$

4.2 分类

- **原型聚类**：k-means，学习向量量化，高斯混合聚类
  - **k-means**:1. 初始化 k 个中心 $\boldsymbol{\mu}_j$。2. E 步：分配样本到最近中心 $C_j = \{\boldsymbol{x}_i | ||\boldsymbol{x}_i - \boldsymbol{\mu}_j|| \leq ||\boldsymbol{x}_i - \boldsymbol{\mu}_{j'}||\}$. 3. M 步：更新中心 $\boldsymbol{\mu}_j = \frac{1}{|C_j|}\sum_{\boldsymbol{x}\in C_j}\boldsymbol{x}$。4. 重复直到收敛。
  - **学习向量量化 (LVQ)**: 约等于 k-means 但是数据带标记。计算每个样本到各中心 p 的距离，如果 x 和 p 标签相同，$p' = p + \eta(x - p)$，否则 $p' = p - \eta(x - p)$。
  - **高斯混合模型 (GMM)**: 假设 $P(\boldsymbol{x}) = \sum_{i=1}^k \alpha_i p(\boldsymbol{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$。使用 **EM 算法**求解：1. **最大化似然函数** $LL = \sum_{j=1}^m \ln(\sum_{i=1}^k \alpha_i p(x_j|\mu_i, \Sigma_i))$ 2.**E 步**：计算后验概率 $\gamma_{ji} = P(z_j = i|\boldsymbol{x}_j) = \frac{\alpha_i p(\boldsymbol{x}_j|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)}{\sum_l \alpha_l p(\boldsymbol{x}_j|\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)}$。3. **M 步**：更新参数 $\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \alpha_i$。$\boldsymbol{\mu}_i^{new} = \frac{1}{N_i}\sum_{j=1}^m \gamma_{ji}\boldsymbol{x}_j$。$\boldsymbol{\Sigma}_i^{new} = \frac{1}{N_i}\sum_{j=1}^m \gamma_{ji}(\boldsymbol{x}_j - \boldsymbol{\mu}_i^{new})(\boldsymbol{x}_j - \boldsymbol{\mu}_i^{new})^T$。$\alpha_i^{new} = \frac{N_i}{m}$
- **密度聚类**：DBSCAN，OPTICS，DENCLUE
  - **DBSCAN**：**核心对象**：邻域内至少包含 MinPts 个样本的点。**密度直达**：位于核心对象邻域内的点。**密度可达**：存在一条由密度直达连接的路径。**密度相连**：存在共同密度可达的点。
- **层次聚类**：AGNES，DIANA
  - **AGNES**: 每个样本作为一个簇，合并两个最近的簇直到大一统。

5 降维

5.1 PCA (主成分分析)

目标：最近重构性或最大可分性。解：协方差矩阵 $\boldsymbol{X}\boldsymbol{X}^T$ 的前 d' 个最大特征值对应的特征向量。$\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{w}_i = \lambda_i\boldsymbol{w}_i$ 重构：$\hat{\boldsymbol{x}} = \sum_{i=1}^{d'} z_i\boldsymbol{w}_i$。
步骤：1. 数据中心化 $\hat{X} = X - \bar{X}$. 2. 计算协方差矩阵 $\Sigma = \hat{X}^T\hat{X}$。3. 求解特征值与特征向量. $\det(\Sigma - \lambda I) = 0$ 解得特征值，依次代入特征值计算 $(\Sigma - \lambda I)v = 0$ 得到特征向量。4. 选择前 d' 个特征向量构成投影矩阵 $\boldsymbol{W} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{d'}]$。5. 投影降维：$\boldsymbol{Z} = \hat{\boldsymbol{X}}\boldsymbol{W}$。

5.2 流形学习

ISOMAP：MDS + 测地线距离（最短路径算法）。**LLE**：保持局部线性关系（重构权重）。
1. **找近邻**：对 $\boldsymbol{x}_i$ 找 k 近邻集合 Q_i。
2. **算权重**：最小化重构误差求权重 w_{ij}（保持 $\boldsymbol{x}_i$ 由邻居线性表示）。$\min_{\boldsymbol{w}}\sum_i ||\boldsymbol{x}_i - \sum_{j\in Q_i} w_{ij}\boldsymbol{x}_j||^2, \quad s.t. \sum_{j\in Q_i} w_{ij} = 1$
3. **求坐标**：固定 w_{ij}，求低维坐标 $\boldsymbol{z}_i$。$\min_{\boldsymbol{Z}}\sum_i ||\boldsymbol{z}_i - \sum_{j\in Q_i} w_{ij}\boldsymbol{z}_j||^2 = \text{tr}(\boldsymbol{Z}\boldsymbol{M}\boldsymbol{Z}^T)$ 其中 $\boldsymbol{M} = (\boldsymbol{I} - \boldsymbol{W})^T(\boldsymbol{I} - \boldsymbol{W})$。
4. **解**：$\boldsymbol{M}$ 的最小 d' 个非零特征值对应的特征向量。

5.3 距离度量学习

核心：学习马氏距离矩阵 $\boldsymbol{M}$，使度量适应任务。$\text{dist}_{\text{mah}}^2(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i - \boldsymbol{x}_j)^T\boldsymbol{M}(\boldsymbol{x}_i - \boldsymbol{x}_j) = ||\boldsymbol{x}_i - \boldsymbol{x}_j||_{\boldsymbol{M}}^2$ 其中 $\boldsymbol{M} \succeq 0$（半正定对称矩阵）。
基于约束的方法：给定必连集合 $\mathcal{M}$ (同类) 和勿连集合 $\mathcal{C}$ (异类)，求解凸优化：$\min_{\boldsymbol{M}}\sum_{(\boldsymbol{x}_i,\boldsymbol{x}_j)\in\mathcal{M}} ||\boldsymbol{x}_i - \boldsymbol{x}_j||_{\boldsymbol{M}}^2 \; s.t. \sum_{(\boldsymbol{x}_i,\boldsymbol{x}_k)\in\mathcal{C}} ||\boldsymbol{x}_i - \boldsymbol{x}_k||_{\boldsymbol{M}}^2 \geq 1, \quad \boldsymbol{M} \succeq 0$
代表性算法
- **NCA (近邻成分分析)**：优化随机近邻分类器的留一法 (LOO) 正确率。概率 $p_{ij} \propto \exp(-||\boldsymbol{x}_i - \boldsymbol{x}_j||_{\boldsymbol{M}}^2)$。
- **LMNN (最大间隔近邻)**：
  - **拉近**：最小化同类 k 近邻的距离。
  - **推开**：保证异类样本与同类近邻之间有间隔 (Margin)。

6 特征选择与稀疏学习

6.1 特征选择

信息增益公式：$Gain(A) = Ent(D) - \sum_{v=1}^V \frac{|D^v|}{|D|}Ent(D^v)$ 其中 $Ent(D) = -\sum_{k=1}^K \frac{|D_k|}{|D|}\log_2\frac{|D_k|}{|D|}$。

- **过滤式 (Filter)-Relief**: 先用特征选择过程过滤原始数据，再进行模型训练；特征选择与后续学习器无关
- **包裹式 (Wrapper)-LVW**: 直接针对给定学习器进行优化，“量身定做”特征子集。性能通常更好，但计算开销大
- **嵌入式 (Embedded)-LASSO&PGD**: 将特征选择过程与学习器训练过程融为一体，通常通过 **正则化** (岭回归（$+\lambda\frac{1}{2}\|w\|_2^2$），LASSO（$+\lambda\|w\|_1$）) 实现

6.2 稀疏学习

- **字典学习**: 将稠密数据转化为稀疏表示。学习字典 B 和稀疏系数 A. 优化目标：$\min_{B,\alpha_i}\sum_{i=1}^m \|x_i - B\alpha_i\|_2^2 + \lambda\|\alpha_i\|_1$。交替优化 B 和 α_i。
- **压缩感知**：利用部分采样恢复全信号。$y = \Phi x$。条件：信号稀疏，采样矩阵满足限定等距性。求解 $\min_s \|s\|_1 \quad \text{s.t.} \quad y = \Phi\Psi s$, 从 y 中恢复稀疏信号 s，进而恢复 x。
- **矩阵补全**：解决推荐系统等场景下的数据缺失问题。优化：$\min_X \|X\|_* \quad \text{s.t.} \quad X_{ij} = A_{ij}, \quad (i,j) \in \Omega$。其中 $\|X\|_* = \sum\sigma_i$。

7 半监督学习

假设：聚类假设（同簇同类）、流形假设（邻近同类）。

7.1 生成式方法

假设 $P(\boldsymbol{x}, y) = \sum_k \alpha_k p(\boldsymbol{x}|\boldsymbol{\theta}_k)$，将未标记数据视为隐变量，最大化对数似然函数：$\ln p(D_l \cup D_u) =$

$\sum_{(x_j,y_j)\in D_l} \ln(\sum_{i=1}^k \alpha_i p(x_j|\mu_i,\Sigma_i)p(y_j|\Theta = i,x_j)) + \sum_{x_j\in D_u} \ln(\sum_{i=1}^k \alpha_i p(x_j|\mu_i,\Sigma_i))$. **求解 (EM 算法)**

7.2 半监督 SVM (S3VM / TSVM)

原理：在最大化间隔的同时，尽量使超平面穿过数据低密度区域（TSVM）。**优化目标**：$\min_{\boldsymbol{w},b,\hat{\boldsymbol{y}},\boldsymbol{\xi}} \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C_l\sum_{i=1}^l \xi_i + C_u\sum_{i=l+1}^m \xi_i$ s.t. 对 D_l 满足标准约束；对 D_u 满足 $\hat{y}_i(\boldsymbol{w}^T x_i + b) \geq 1-\xi_i$。**求解策略**：1. 用 D_l 训练初始 SVM。2. 对 D_u 进行预测指派伪标记 $\hat{y}$。3. **迭代**：找出可能错误的伪标记样本（$y_i\hat{y}_j < 0$ 且 $\xi > 0$），交换其标记，重新求解，逐步增大 C_u 至 C_l。

7.3 图半监督学习 (Graph-Based Methods)

构图：节点为样本，边权重 W_{ij} 为相似度（如高斯核）。**能量函数**：$E(f) = \frac{1}{2}\sum_{i,j} W_{ij}(f(x_i)-f(x_j))^2 = \boldsymbol{f}^T(\boldsymbol{D}-\boldsymbol{W})\boldsymbol{f}$。**闭式解**：令 $\frac{\partial E}{\partial f_u} = 0$，得 $f_u = (D_{uu}-W_{uu})^{-1}W_{ul}f_l = (I-P_{uu})^{-1}P_{ul}f_l$。**迭代式标记传播** (Label Propagation)：$\boldsymbol{F}(t+1) = \alpha\boldsymbol{S}\boldsymbol{F}(t) + (1-\alpha)\boldsymbol{Y}$ 其中 $\boldsymbol{S} = D^{-1/2}WD^{-1/2}$，收敛解为 $\boldsymbol{F}^* = (1-\alpha)(I-\alpha\boldsymbol{S})^{-1}\boldsymbol{Y}$。

7.4 基于分歧的方法 (Disagreement-based Methods)

代表：协同训练 (Co-training)。**条件**：数据拥有两个充分且条件独立的视图 (View)。**流程**：1. 在两个视图上分别训练分类器 h_1, h_2。2. 每个分类器挑选置信度最高的未标记样本（正/负），赋予伪标记并加入对方的训练集。3. 迭代更新。

7.5 半监督聚类 (Semi-Supervised Clustering)

监督信息类型：
- **约束**：必连 (Must-link, M) 与勿连 (Cannot-link, C)。
- **标记样本**：少量样本已知所属簇。

算法：1. **约束 k-means**：在分配簇时，若违反 M 或 C 约束，则不分配或分配给次优簇。2. **约束种子 k-means**：用有标记样本初始化簇中心 $\mu_j = \frac{1}{|S_j|}\sum_{x\in S_j} x$。

8 概率图模型

8.1 隐马尔可夫模型 (HMM)

动态贝叶斯网，由隐藏状态序列 y 和观测变量序列 x 组成。
- **三组参数** $\lambda = [A, B, \pi]$：
  - 状态转移概率 A: $a_{ij} = P(y_{t+1} = s_j | y_t = s_i)$
  - 输出观测概率 B: $b_{ij} = P(x_t = o_j | y_t = s_i)$
  - 初始状态概率 π: $\pi_i = P(y_1 = s_i)$
- **齐次马尔可夫假设**：$P(y_t|y_{t-1},\dots,y_1) = P(y_t|y_{t-1})$。
- **观测独立性假设**：$P(x_t|y_t,\dots) = P(x_t|y_t)$。

8.2 马尔可夫随机场 (MRF)

基于无向图，使用**团 (Clique)** 和**势函数 (Potential Function)** 定义联合概率。**联合概率分布**：$P(\mathbf{x}) = \frac{1}{Z}\prod_{Q\in\mathcal{C}}\psi_Q(\mathbf{x}_Q)$ 其中 $\mathcal{C}$ 为极大团集合，ψ_Q 为势函数 (通常 $\psi_Q \geq 0$)，$Z = \sum_{\mathbf{x}}\prod_{Q\in\mathcal{C}}\psi_Q(\mathbf{x}_Q)$ 为规范化因子。**马尔可夫性**：
- 全局：给定分离集 x_C，则 $x_A \perp x_B | x_C$。
- 局部/成对：非邻接节点在给定其他节点条件下独立。

8.3 条件随机场 (CRF)

判别式无向图模型，对条件分布 $P(\mathbf{y}|\mathbf{x})$ 建模。**链式 CRF** (线性链)：$P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z}\exp(\sum_j\sum_i \lambda_j t_j(y_{i+1}, y_i, \mathbf{x}, i) + \sum_k\sum_i \mu_k s_k(y_i, \mathbf{x}, i))$ 其中 t_j 为转移特征函数，s_k 为状态特征函数。

8.4 模型推断 (Inference)

推断的核心是计算边际分布或条件分布。
- **精确推断**：
  - **变量消去法 (Variable Elimination)**：利用乘法对加法的分配律，将全局求和转化为局部求和，实质是动态规划。
  - **信念传播 (Belief Propagation)**：通过节点间传递消息 $m_{ij}(x_j)$ 计算边际分布。节点 x_i 的边际分布正比于接收消息的乘积.
- **近似推断**：
  - **采样法**: 通过构造平稳分布为目标分布 $p(\mathbf{x})$ 的马尔可夫链来产生样本。**Metropolis-Hastings (MH) 算法**：1. 根据 $Q(\mathbf{x}^*|\mathbf{x}^{t-1})$ 采样候选 $\mathbf{x}^*$。2. 计算接受率 $\alpha = \min\left(1, \frac{p(\mathbf{x}^*)Q(\mathbf{x}^{t-1}|\mathbf{x}^*)}{p(\mathbf{x}^{t-1})Q(\mathbf{x}^*|\mathbf{x}^{t-1})}\right)$。3. 以概率 α 接受 $\mathbf{x}^*$ 作为 $\mathbf{x}^t$。**吉布斯采样 (Gibbs Sampling)**：MH 的特例。每次固定其他变量，仅对一个变量 x_i 根据 $p(x_i|\mathbf{x}_{\setminus i})$ 进行采样。
  - **变分推断 (Variational Inference)**: 使用简单分布 $q(\mathbf{z})$ 逼近复杂后验分布 $p(\mathbf{z}|\mathbf{x})$。**目标**：最小化 KL 散度 $KL(q\|p)$，等价于最大化证据下界 (ELBO)。$\ln p(\mathbf{x}) = \mathcal{L}(q) + KL(q\|p)$ **平均场理论**：假设 $q(\mathbf{z}) = \prod_i q_i(z_i)$，最优解满足 $\ln q_j^*(z_j) = \mathbb{E}_{i\neq j}[\ln p(\mathbf{x},\mathbf{z})] + \text{const}$。

9 规则学习

序贯覆盖 (Sequential Coverage)：逐条学习规则，覆盖正例并移除，直到覆盖所有正例。**剪枝**：预剪枝（似然率）、后剪枝（REP, IREP, RIPPER）。**一阶规则 (FOIL)**：使用一阶逻辑（谓词）。FOIL 增益：$Gain = \hat{m}_+(\log_2\frac{m_+}{\hat{m}_+ + \hat{m}_-} - \log_2\frac{m_+}{m_+ + m_-})$。$m_+$ 是标记正例中被规则判定为正例，m_- 是标记负例中被规则判定为正例.

冲突消解：顺序规则、缺省规则、元规则。

归纳逻辑程序设计 (ILP)：
- **最小一般泛化 (LGG)**：寻找覆盖两个例子的最特殊的一般规则。
- **逆归结**：演绎的逆过程。

吸收：$\frac{p\leftarrow A\wedge B \quad q\leftarrow A}{p\leftarrow q\wedge B \quad q\leftarrow A}$ 辨识：$\frac{p\leftarrow A\wedge B \quad p\leftarrow A\wedge q}{q\leftarrow B \quad p\leftarrow A\wedge q}$

内构：$\frac{p\leftarrow A\wedge B \quad p\leftarrow A\wedge C}{q\leftarrow B \quad p\leftarrow A\wedge q \quad q\leftarrow C}$ 互构：$\frac{p\leftarrow A\wedge B \quad q\leftarrow A\wedge C}{p\leftarrow r\wedge B \quad r\leftarrow A \quad q\leftarrow r\wedge C}$

10 强化学习 (RL)

四元组 $< X, A, P, R >$。目标：最大化累积回报 $\mathbb{E}[\sum\gamma^t r_t]$。

10.1 K-摇臂赌博机 (K-Armed Bandit)

单步强化学习，最大化单步奖赏。
- **ϵ-贪心**: 以 ϵ 概率随机探索，以 $1-\epsilon$ 概率利用。
- **增量更新公式**: $Q_n(k) = Q_{n-1}(k) + \frac{1}{n}(v_n - Q_{n-1}(k))$。
- **Softmax**: 基于概率分布选择动作，概率 $P(k) = \frac{e^{Q(k)/\tau}}{\sum_i e^{Q(i)/\tau}}$ (τ 为温度参数)。

10.2 值函数与 Bellman 方程

状态值函数：$V^\pi(x) = \mathbb{E}_\pi[\sum\gamma^t r_t | x_0 = x]$。**动作值函数**：$Q^\pi(x,a) = \mathbb{E}_\pi[\sum\gamma^t r_t | x_0 = x, a_0 = a]$。

$V^\pi(x) = \sum_a \pi(a|x)\sum_{x'} P(x'|x,a)[R(x,a,x') + \gamma V^\pi(x')]$

最优 Bellman 方程：$V^*(x) = \max_a Q^*(x,a) = \max_a \sum_{x'} P(x'|x,a)$ $Q^*(x,a) = \sum_{x'} P(x'|x,a)[R + \gamma\max_{a'} Q^*(x',a')]$

10.3 求解算法

- **动态规划 (Model-based)**:
  - **策略迭代 (PI)**: 评估 $V^\pi \to$ 改进 $\pi'(x) = \arg\max Q^\pi(x,a)$。
  - **值迭代 (VI)**: 直接迭代 $V(x) \leftarrow \max_a \sum P(...)$。
- **免模型 (Model-free)**:
  - **蒙特卡罗 (MC)**: 采样轨迹，均值估计期望。**重要性采样**：$\rho_{t:T-1} = \prod_{k=t}^{T-1}\frac{\pi(A_k|S_k)}{b(A_k|S_k)}$。$V(S_t) \approx \rho_{t:T-1}G_t$
  - **时序差分 (TD)**: 结合 MC 和 DP。$V(x) \leftarrow V(x) + \alpha[r + \gamma V(x') - V(x)]$。
  - **Q-Learning (Off-policy)**: $Q(x,a) \leftarrow Q(x,a) + \alpha[r + \gamma\max_{a'} Q(x',a') - Q(x,a)]$。
  - **Sarsa (On-policy)**: $Q(x,a) \leftarrow Q(x,a) + \alpha[r + \gamma Q(x',a') - Q(x,a)]$。

探索与利用：ϵ-greedy (以 ϵ 概率随机)。

10.4 值函数近似 (Value Function Approximation)

针对状态空间巨大或连续的情况，使用参数化模型逼近值函数。
- **线性近似**: $V_\theta(x) = \boldsymbol{\theta}^T\boldsymbol{\phi}(x)$，其中 $\boldsymbol{\phi}(x)$ 为特征向量。
- **梯度更新** (结合 TD 误差): $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \alpha(r + \gamma V_\theta(x') - V_\theta(x))\nabla_\theta V_\theta(x)$ 对于线性近似，$\nabla_\theta V_\theta(x) = \boldsymbol{\phi}(x)$。

10.5 模仿学习 (Imitation Learning)

利用专家示范数据 $\mathcal{D} = \{\tau_1, \tau_2, ...\}$ 进行学习。
- **直接模仿 (Behavior Cloning)**: 将专家轨迹视为分类/回归训练集，直接学习策略 $\pi: X \to A$。
- **逆强化学习 (IRL)**: 假设专家策略是最优的，从数据中反推奖赏函数 R。目标: 寻找 R^* 使得 $\mathbb{E}_{\pi^*}[\sum\gamma^t R^*(s_t)] \geq \mathbb{E}_\pi[\sum\gamma^t R^*(s_t)]$ 得到 R^* 后，再通过标准 RL 算法求解最优策略。