ch11 为什么特征选择: 降维, 去除不相关特征降低学习难度. 选特征子集: 前向搜索增加相关特征. 子集评价: 属性子集$A$, 信息增益$\operatorname{Gain}(A) = \operatorname{Ent}(D) - \sum_{v=1}^{V} \frac{\left|D^{v}\right|}{|D|} \operatorname{Ent}\left(D^{v}\right)$, 信息熵$\operatorname{Ent}(D) = -\sum_{i=1}^{|\mathcal{Y}|} p_{k} \log_{2} p_{k}$, $\operatorname{Gain}(A) \uparrow$ 则特征子集有助于分类的信息越多. 搜索+评价=特征选择, 前向搜+信息熵则类似决策树. 特征选择方法: 1 过滤式. Relief: 特征子集重要性: 子集中每个特征所对应的相关统计量分量$\delta^{j} = \sum_{i} -\operatorname{diff}\left(x_{i}^{j}, x_{i, \mathrm{nh}}^{j}\right)^{2} + \operatorname{diff}\left(x_{i}^{j}, x_{i, \mathrm{nm}}^{j}\right)^{2}$之和, 选阈值$\tau$以上的. $\boldsymbol{x}_{i, \mathrm{nm}}$ 猜错近邻, 属性有益则$\boldsymbol{x}_{i}$与$\boldsymbol{x}_{i, \mathrm{nm}}$距离大于$\boldsymbol{x}_{i}$与$\boldsymbol{x}_{i, \mathrm{nh}}$. 2 包裹式. LVM拉斯维加斯方法框架下随机搜索出特征子集, 训练后看误差. 3 嵌入式(特征选择与训练过程融合). LASSO, $\mathbf{L}_{1}$范数. 近端梯度下降PGD: $\nabla f$ 满足L-Lipschitz. 迭代 $\boldsymbol{x}_{k+1} = \underset{\boldsymbol{x}}{\arg \min} \frac{L}{2}\left\|\boldsymbol{x}-\left(\boldsymbol{x}_{k}-\frac{1}{L} \nabla f\left(\boldsymbol{x}_{k}\right)\right)\right\|_{2}^{2} + \lambda\|\boldsymbol{x}\|_{1}$, 闭式解 $x_{k+1}^{i} = z^{i} - \lambda/L, \lambda/L < z^{i}$; $z^{i} + \lambda/L, z^{i} < -\lambda/L$; 0 其他. 稀疏表示与字典学习: 优点 1 使大多数问题线性可分, 2 存储负担小. 字典学习: 为稠密表达找字典→稀疏表达: 1 固定$\mathbf{B}$像LASSO解法: $\min_{\boldsymbol{\alpha}_{i}} \left\|\boldsymbol{x}_{i}-\mathbf{B} \boldsymbol{\alpha}_{i}\right\|_{2}^{2} + \lambda\left\|\boldsymbol{\alpha}_{i}\right\|_{1}$, 2 以$\boldsymbol{\alpha}_{i}$为初值 $\min_{\mathbf{B}}\|\mathbf{X}-\mathbf{B A}\|_{F}^{2} = \min_{\boldsymbol{b}_{i}}\left\|\mathbf{X}-\sum_{j=1}^{k} \boldsymbol{b}_{j} \boldsymbol{\alpha}^{j}\right\|_{F}^{2} = \min_{\boldsymbol{b}_{i}}\left\|\left(\mathbf{X}-\sum_{j \neq i} \boldsymbol{b}_{j} \boldsymbol{\alpha}^{j}\right)-\boldsymbol{b}_{i} \boldsymbol{\alpha}^{i}\right\|_{F}^{2}$ 括号内$\mathbf{E}_{i}$进行奇异值分解, 但直接分解可能破坏$\mathbf{A}$的稀疏性, $\therefore$ KSVD对$\boldsymbol{\alpha}_{i}$保留非零元素 $\mathbf{E}_{i}$仅保留$\boldsymbol{b}_{i}$与$\boldsymbol{\alpha}_{i}$的非零元素乘积项. 压缩感知 $\boldsymbol{y} = \boldsymbol{\Phi} \boldsymbol{x} = \boldsymbol{\Phi} \boldsymbol{\Psi} \boldsymbol{s} = \mathbf{A} \boldsymbol{s}$ 恢复$\boldsymbol{s}$⇒恢复$\boldsymbol{x}$, $k$限定等距性k-RIP $\left(1-\delta_{k}\right)\|\boldsymbol{s}\|_{2}^{2} \leqslant \left\|\mathbf{A}_{k} \boldsymbol{s}\right\|_{2}^{2} \leqslant \left(1+\delta_{k}\right)\|\boldsymbol{s}\|_{2}^{2}$ 此时 $\min_{\boldsymbol{s}}\|\boldsymbol{s}\|_{0} \Rightarrow \|\boldsymbol{s}\|_{1}$ s.t. $\boldsymbol{y} = \mathbf{A} \boldsymbol{s}$. 矩阵补全 $\min_{\mathbf{X}} \operatorname{rank}(\mathbf{X})$ s.t. $(\mathbf{X})_{i j} = (\mathbf{A})_{i j}, (i, j) \in \Omega$, 转化: $\because$ $\operatorname{rank}(\mathbf{X})$在集合$\left\{\mathbf{X} \in \mathbb{R}^{m \times n} : \|\mathbf{X}\|_{F}^{2} \leqslant 1\right\}$上的凸包是$\mathbf{X}$的 “核范数” $\|\mathbf{X}\|_{*} = \sum_{j=1}^{\min\{m, n\}} \sigma_{j}(\mathbf{X})$.

ch12 学习算法$\mathfrak{L}$学得的模型对应的假设$h$尽可能接近目标概念$c$. 不能完全一致因为$D$有限$D$采样偶然. 1PAC辨识: $\mathfrak{L}$从$\mathcal{H}$中PAC辨识$\mathcal{C}$: $P(E(h) \leqslant \epsilon) \geqslant 1-\delta$. $\mathcal{C}$ 2PAC可学习: $\mathfrak{p} = \operatorname{poly}(1/\epsilon, 1/\delta, \operatorname{size}(\boldsymbol{x}), \operatorname{size}(c))\ \forall\, m \geqslant \mathfrak{p}$ 可PAC辨识. 3PAC学习算法: $\mathfrak{L}$ PAC可学习且运行时间$= \mathfrak{p}$. 4样本复杂度: PAC学习算法最小$m \geqslant \mathfrak{p}$. 有限假设空间可分: $\mathfrak{L}$以概率$1-\delta$找到目标假设的$\epsilon$近似: All $i$, $P(h(\boldsymbol{x}_{i} = y_{i})) < (1-\epsilon)^{m}$. 对All $h$, 上式$< |\mathcal{H}|(1-\epsilon)^{m} < |\mathcal{H}| e^{-m \epsilon} \leqslant \delta$, 上式可推$m \geqslant \cdots$. 不可分: $\forall h$ $P\left(|E(h)-\widehat{E}(h)| \leqslant \sqrt{(\ln |\mathcal{H}|+\ln (2/\delta))/(2m)}\right) \geqslant 1-\delta$, 5不可知PAC可学习: $\forall\, m \geqslant \mathfrak{p}$ $P\left(E(h)-\min_{h' \in \mathcal{H}} E\left(h'\right) \leqslant \epsilon\right) \geqslant 1-\delta$. 6增长函数: $\Pi_{\mathcal{H}}(m) = \max_{\{\boldsymbol{x}_{1}, \ldots, \boldsymbol{x}_{m}\} \subseteq \mathcal{X}} \left|\left\{\left(h\left(\boldsymbol{x}_{1}\right), \ldots, h\left(\boldsymbol{x}_{m}\right)\right) \mid h \in \mathcal{H}\right\}\right|$ 集合的数量. 7对分: $\mathcal{H}$中的$h$对$D$中赋予标记的每种可能结果. 8$D$被$\mathcal{H}$打散: 二分类问题$\Pi_{\mathcal{H}}(m) = 2^{m}$ 实现∀对分. 9VC维: $\operatorname{VC}(\mathcal{H}) = \max\left\{m : \Pi_{\mathcal{H}}(m) = 2^{m}\right\}$. VC维$d$: ∃大小为$d$的示例集能被$\mathcal{H}$打散, ∀大小为$d+1$的示例集不能.

ch13 主动学习: 查询专家后有标签加入再训练, 尽量少的”查询”来获得最优性能. 半监督: 不依赖外界交互/自动利用未标记样本: 基于聚类假设/流形假设相似样本相似输出. 半监督分为: 1纯半监督/2直推学习 1中训练数据中未标记样本不是待测数据而2中是, 1基于开放世界假设希望模型适用于未观察的数据, 2封闭世界假设仅预测未标记数据. 半监督的生成式: 未标记数据看成模型缺失参数. EM: $p(\boldsymbol{x}) = \sum_{i=1}^{N} \alpha_{i} \cdot p\left(\boldsymbol{x} \mid \boldsymbol{\mu}_{i}, \boldsymbol{\Sigma}_{i}\right)$ $\therefore$ MAP: $f(\boldsymbol{x}) = \underset{j \in \mathcal{Y}}{\arg \max} p(y = j \mid \boldsymbol{x}) = \underset{j \in \mathcal{Y}}{\arg \max} \sum_{i=1}^{N} p(y = j \mid \Theta = i, \boldsymbol{x}) \cdot p(\Theta = i \mid \boldsymbol{x})$ 其中 $p(\Theta = i \mid \boldsymbol{x}) = \alpha_{i} \cdot p\left(\boldsymbol{x} \mid \boldsymbol{\mu}_{i}, \boldsymbol{\Sigma}_{i}\right) / \sum_{i=1}^{N} \alpha_{i} \cdot p\left(\boldsymbol{x} \mid \boldsymbol{\mu}_{i}, \boldsymbol{\Sigma}_{i}\right)$. $\Theta$ 指向哪个高斯混合成分. $l_{i}$表示第$i$类的有标记样本数目. $D_{l} \cup D_{u}$对数似然: $LL\left(D_{l} \cup D_{u}\right) = \sum_{(\boldsymbol{x}_{j}, y_{j}) \in D_{l}} \ln \left(\sum_{i=1}^{N} \alpha_{i} \cdot p\left(\boldsymbol{x}_{j} \mid \boldsymbol{\mu}_{i}, \boldsymbol{\Sigma}_{i}\right) \cdot p\left(y_{j} \mid \Theta = i, \boldsymbol{x}_{j}\right)\right) + \sum_{\boldsymbol{x}_{j} \in D_{u}} \ln \left(\sum_{i=1}^{N} \alpha_{i} \cdot p\left(\boldsymbol{x}_{j} \mid \boldsymbol{\mu}_{i}, \boldsymbol{\Sigma}_{i}\right)\right)$, E步: $\gamma_{ji} = \alpha_{i} \cdot p\left(\boldsymbol{x}_{j} \mid \boldsymbol{\mu}_{i}, \boldsymbol{\Sigma}_{i}\right) / \sum_{i=1}^{N} \alpha_{i} \cdot p\left(\boldsymbol{x}_{j} \mid \boldsymbol{\mu}_{i}, \boldsymbol{\Sigma}_{i}\right)$, M 步: $\boldsymbol{\mu}_{i} = \frac{1}{\sum_{\boldsymbol{x}_{j} \in D_{u}} \gamma_{ji}+l_{i}}\left(\sum_{\boldsymbol{x}_{j} \in D_{u}} \gamma_{ji} \boldsymbol{x}_{j}+\sum_{(\boldsymbol{x}_{j}, y_{j}) \in D_{l} \wedge y_{j}=i} \boldsymbol{x}_{j}\right)$, $\boldsymbol{\Sigma}_{i} = 1/(\sum_{\boldsymbol{x}_{j} \in D_{u}} \gamma_{ji}+l_{i})\left(\sum_{\boldsymbol{x}_{j} \in D_{u}} \gamma_{ji}\left(\boldsymbol{x}_{j}-\boldsymbol{\mu}_{i}\right)\left(\boldsymbol{x}_{j}-\boldsymbol{\mu}_{i}\right)^{\mathrm{T}} + \sum_{(\boldsymbol{x}_{j}, y_{j}) \in D_{l} \wedge y_{j}=i}\left(\boldsymbol{x}_{j}-\boldsymbol{\mu}_{i}\right)\left(\boldsymbol{x}_{j}-\boldsymbol{\mu}_{i}\right)^{\mathrm{T}}\right)$ $\alpha_{i} = \frac{1}{m}\left(\sum_{\boldsymbol{x}_{j} \in D_{u}} \gamma_{ji}+l_{i}\right)$. 此类生成式方法关键: 模型假设必须准确, 假设的生成式模型必须与真实数据分布吻合. 半监督SVM中TSVM: 二分类, 式子① $\min_{\boldsymbol{w}, b, \hat{\boldsymbol{y}}, \boldsymbol{\xi}} \frac{1}{2}\|\boldsymbol{w}\|_{2}^{2} + C_{l} \sum_{i=1}^{l} \xi_{i} + C_{u} \sum_{i=l+1}^{m} \xi_{i}$ s.t. $y_{i}\left(\boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_{i}+b\right) \geqslant 1-\xi_{i}, \hat{y}_{i}\left(\boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_{i}+b\right) \geqslant 1-\xi_{i}, \xi_{i} \geqslant 0$. 试图考虑对未标记样本进行尝试各种可能的标记指派: 局部搜索来迭代地寻找近似解 1$D_{l}$学得SVM, 2给$D_{u}$伪标记, 3代入①求解出新的SVM, 4找预测为异类且很可能发生错误的$D_{u}$: $\exists\left\{i, j \mid \left(\hat{y}_{i} \hat{y}_{j}<0\right) \wedge \left(\xi_{i}>0\right) \wedge \left(\xi_{j}>0\right) \wedge \left(\xi_{i}+\xi_{j}>2\right)\right\}$, 再求解①, 这样找一遍$D_{u}$后$C_{u} = \min\left\{2 C_{u}, C_{l}\right\}$, 直到$C_{u} = C_{l}$时停止. 初始化$C_{u} \ll C_{l}$. 对未标记样本进行调整过程中可能类别不平衡拆分$C_{u}$并改进初始化 $C_{u}^{+} = u_{-}/u_{+} \cdot C_{u}^{-}$. 图半监督: 结点$V$ $x_{i}$, 边集$E$亲和矩阵$(\mathbf{W})_{ij} = \exp\left(-\left\|\boldsymbol{x}_{i}-\boldsymbol{x}_{j}\right\|_{2}^{2}/2\sigma^{2}\right)$, if $i \neq j$ 其他为0, 期望学得$f : V \rightarrow \mathbb{R}$, 通过$\operatorname{sign}(f(x_{i}))$分类, 二分类标记传播: 定义能力函数及其矩阵表达(有点像损失函数): $E(f) = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m}(\mathbf{W})_{ij}\left(f\left(\boldsymbol{x}_{i}\right)-f\left(\boldsymbol{x}_{j}\right)\right)^{2} = \frac{1}{2}\left(\sum_{i=1}^{m} d_{i} f^{2}\left(\boldsymbol{x}_{i}\right)+\sum_{j=1}^{m} d_{j} f^{2}\left(\boldsymbol{x}_{j}\right)-2 \sum_{i=1}^{m} \sum_{j=1}^{m}(\mathbf{W})_{ij} f\left(\boldsymbol{x}_{i}\right) f\left(\boldsymbol{x}_{j}\right)\right) = \sum_{i=1}^{m} d_{i} f^{2}\left(\boldsymbol{x}_{i}\right) - \sum_{i=1}^{m} \sum_{j=1}^{m}(\mathbf{W})_{ij} f\left(\boldsymbol{x}_{i}\right) f\left(\boldsymbol{x}_{j}\right) = \boldsymbol{f}^{\mathrm{T}}(\mathbf{D}-\mathbf{W}) \boldsymbol{f}$. 这里$\mathbf{D} = \operatorname{diag}\left(d_{1}, d_{2}, \ldots, d_{l+u}\right)$ $d_{i} = \sum_{j=1}^{l+u}(\mathbf{W})_{ij}$. 拉普拉斯矩阵$\Delta = \mathbf{D}-\mathbf{W}$. 按照$l$和$u$分块矩阵表达: $E(f) = \boldsymbol{f}_{l}^{\mathrm{T}}\left(\mathbf{D}_{ll}-\mathbf{W}_{ll}\right) \boldsymbol{f}_{l} - 2 \boldsymbol{f}_{u}^{\mathrm{T}} \mathbf{W}_{ul} \boldsymbol{f}_{l} + \boldsymbol{f}_{u}^{\mathrm{T}}\left(\mathbf{D}_{uu}-\mathbf{W}_{uu}\right) \boldsymbol{f}_{u}$, 对$\boldsymbol{f}_{u}$求偏导等于0得式②: $\boldsymbol{f}_{u} = \left(\mathbf{D}_{uu}-\mathbf{W}_{uu}\right)^{-1} \mathbf{W}_{ul} \boldsymbol{f}_{l}$. 进一步分块, 令$\mathbf{P}_{uu} = \mathbf{D}_{uu}^{-1} \mathbf{W}_{uu}, \mathbf{P}_{ul} = \mathbf{D}_{uu}^{-1} \mathbf{W}_{ul}$, ②化简为: $\boldsymbol{f}_{u} = \left(\mathbf{D}_{uu}\left(\mathbf{I}-\mathbf{D}_{uu}^{-1} \mathbf{W}_{uu}\right)\right)^{-1} \mathbf{W}_{ul} \boldsymbol{f}_{l} = \left(\mathbf{I}-\mathbf{P}_{uu}\right)^{-1} \mathbf{P}_{ul} f_{l}$. 多分类标记传播: $\mathbf{F}$矩阵迭代, $\mathbf{F}(0) = (\mathbf{Y})_{ij} =$ 当$i$是有标记样本(第$i$个)并且类别为$j$时为1. 标记传播矩阵$\mathbf{S} = \mathbf{D}^{-\frac{1}{2}} \mathbf{W} \mathbf{D}^{-\frac{1}{2}}$, $\mathbf{D}^{-\frac{1}{2}} = \operatorname{diag}(1/\sqrt{d_{1}}, \cdots, 1/\sqrt{d_{l+u}})$, 于是有迭代式: $\mathbf{F}(t+1) = \alpha \mathbf{S} \mathbf{F}(t)+(1-\alpha) \mathbf{Y}$ 基于此迭代收敛得: $\mathbf{F}^{*} = \lim_{t \mapsto \infty} \mathbf{F}(t) = (1-\alpha)(\mathbf{I}-\alpha \mathbf{S})^{-1} \mathbf{Y}$. 对于所有未标记样本$i$: $y_{i} = \arg \max_{1 \leqslant j \leqslant |\mathcal{Y}|}\left(\mathbf{F}^{*}\right)_{ij}$. 对应于正则化框架: $\min_{\mathbf{F}} \frac{1}{2}\left(\sum_{i, j=1}^{l+u}(\mathbf{W})_{ij}\left\|\frac{1}{\sqrt{d_{i}}} \mathbf{F}_{i}-\frac{1}{\sqrt{d_{j}}} \mathbf{F}_{j}\right\|^{2}\right) + \mu \sum_{i=1}^{l}\left\|\mathbf{F}_{i}-\mathbf{Y}_{i}\right\|^{2}$, 上式中$\mu = (1-\alpha)/\alpha$时最优解恰为迭代收敛解$\mathbf{F}^{*}$. 基于分歧的方法: 学习器之间的分歧作用于未标记数据. multi-view中的co-training: 图像属性集是一个视图. 不同视图相容性: 包含的关于输出空间$\mathcal{Y}$(标签)的信息是一致的. 互补性: 不同视图信息互补. 充分且条件独立视图: 充分每个视图包含足以产生最优学习器的信息, 条件独立在给定类别标记下两个视图独立. 训练: 先在每个视图上训, 让训好的分类器挑选最有把握的未标记数据赋予伪标记, 迭代. 评价基于分歧的方法: 只需采用合适的基学习器具有显著分歧性能尚可的, 就能较少受到模型假设/损失函数非凸性和数据规模问题的影响, 学习方法简单有效理论基础相对坚实适用范围较为广泛. 半监督聚类: 聚类任务中的监督信息 1”必连”(样本属于同一个簇)和”勿连”约束, 2少量的有标记样本. 约束$k$均值算法利用第1类信息(注意这只有部分样本的约束信息): $\mathcal{M}$是必连关系集合($(x_{i}, x_{j})$属于它表示同类) $\mathcal{C}$是勿连关系集合, 训过程: 在k-means聚类过程中保证$\mathcal{M}$和$\mathcal{C}$中的约束满足即可. 约束种子$k$均值算法利用第2类信息: 少量有标记样本: $S = \cup_{j=1}^{k} S_{j} \subset D$, $S_{j}$ 表示隶属于第$j$个聚类簇的样本, 训过程: 直接用有监督信息的样本作为初始化聚类中心, 并在样本簇归属的迭代更新过程中不更新$S$中的类别隶属.

ch14 EM算法: $\log p(x_{i} \mid \theta)$=$\log(p(x_{i}, z_{i}))$-$\log(p(z_{i} \mid x_{i}))$ =$E_{z_{i}}\left(\log p\left(x_{i}, z_{i}\right)\right) - H\left(q_{i}\right) + KL\left(q_{i}\left(z_{i}\right) \| p\left(z_{i} \mid x_{i}\right)\right)$, ⇒最大化ELBO, E步把KL设为0; M步最大化ELBO. 概率模型: 一种框架, 将学习任务归结于计算变量的概率分布, ”推断”: 利用已知变量推测未知变量的分布; (定义集合: $Y$ 所关心的变量; $O$ 可观测变量; $R$ 其他变量) 则生成式模型考虑联合分布 $P(Y, R, O)$, 判别式模型考虑条件分布 $P(Y, R \mid O)$. 给定$O$, 需由$P(Y, R, O)$或$P(Y, R \mid O)$得到$P(Y \mid O)$. 概率图模型: 一类用图来表达变量相关关系的概率模型, 一个截断表示一个或一组随机变量, 结点之间的边表示变量间的概率相关关系. 概率图模型分为: 1使用有向无环图表示变量间的依赖关系, 有向图模型/贝叶斯网; 2使用无向图表示变量间的相关关系, 无向图模型/马尔可夫网. HMM隐马尔可夫模型结构最简单的动态贝叶斯网 (生成式): 变量有 1状态变量$\mathcal{Y}$ 是隐藏的不可观测的隐变量, 离散; 2观测变量$\mathcal{X}$观测值, 离散或连续. 确定HMM所需的参数: 1结构信息联合概率分布: $P$(指向的结点 | 指出来的结点) 的相乘; 2状态转移概率: $a_{ij} = P\left(y_{t+1} = s_{j} \mid y_{t} = s_{i}\right)$; 3输出观测概率 $b_{ij} = P\left(x_{t} = o_{j} \mid y_{t} = s_{i}\right)$, 根据当前状态获得各个观测值的概率; 4初始状态概率 $\pi_{i} = P\left(y_{1} = s_{i}\right)$. HMM产生观测序列$\{x_{i}\}$: 1)设置 $t = 1$, 并根据初始状态概率 $\boldsymbol{\pi}$ 选择初始状态 $y_{1}$; 2)根据状态 $y_{t}$ 和输出观测概率 $\mathbf{B}$ 选择观测变量取值$x_{t}$; 3)根据状态 $y_{t}$ 和状态转移矩阵 $\mathbf{A}$ 转移模型状态, 即确定 $y_{t+1}$; 4)若 $t < n$, 设置 $t = t+1$, 并转到第 (2) 步, 否则停止. HMM三个基本问题 (基于联合概率分布条件独立性可高效求解): 1)给定模型 $\lambda = [\mathbf{A}, \mathbf{B}, \pi]$, 如何有效计算 $P(\mathbf{x} \mid \lambda)$ (产生观测序列的概率)? 2)给定模型和观测序列 $\{x_{1} \cdots x_{n}\}$, 如何找到最匹配的状态序列(观测序列推隐藏模型状态)? (例子: 语音识别) 3) 给定观测序列如何调整模型参数使 $P(\mathbf{x} \mid \lambda)$ 最大(更好地描述观测数据)? MRF马尔可夫随机场无向图模型(生成式): 一组势函数/”因子”: 定义在变量子集上的非负实函数(用于定义概率分布函数). ”团”clique: 其中任意两点间有边连接; ”极大团”: 加入相邻任一结点不再形成团. 多个变量之间的联合概率分布能基于团分解为多个因子的乘积, 每个因子仅与一个团有关. 例如: 所有团的集合为$\mathcal{C}$, 联合概率: $P(\mathbf{x}) = \frac{1}{Z} \prod_{Q \in \mathcal{C}} \psi_{Q}\left(\mathbf{x}_{Q}\right)$, 其中$\psi_{Q}$是团$Q$的势函数, $Z$是规范化(归一化)因子, 因为变量个数$\mathbf{x}$较多造成团较多, 所以$P(x)$中$\mathcal{C}^{*}$可基于极大团定义. 一个结点可能出现在多个团中. 全局马尔可夫性: 其中条件独立性借助 ”分离”概念即结点集$A \rightarrow B$ 都要经过结点集$C$, 则$C$是分离集, $A$和$B$条件独立: $\mathbf{x}_{A} \perp \mathbf{x}_{B} \mid \mathbf{x}_{C}$. 此时联合概率 $P\left(x_{A}, x_{B}, x_{C}\right) = \frac{1}{Z} \psi_{AC}\left(x_{A}, x_{C}\right) \psi_{BC}\left(x_{B}, x_{C}\right)$; 条件独立的数学证明 $P\left(x_{A}, x_{B} \mid x_{C}\right) = P\left(x_{A} \mid x_{C}\right) P\left(x_{B} \mid x_{C}\right)$: 等式左边利用此时: $P(x_{C}) = \sum_{x_{A}'} \sum_{x_{B}'} P(x_{A}', x_{B}', x_{C}) = \sum_{x_{A}'} \sum_{x_{B}'} \psi_{AC}\left(x_{A}', x_{C}\right) \psi_{BC}\left(x_{B}', x_{C}\right)/Z$, 等式右边: $P\left(x_{A}, x_{C}\right) = \sum_{x_{B}'} P\left(x_{A}, x_{B}', x_{C}\right) = \sum_{x_{B}'} \psi_{AC}\left(x_{A}, x_{C}\right) \psi_{BC}\left(x_{B}', x_{C}\right)/Z$, $P(\cdot \mid \cdot)$ 展开约分即可. 局部马尔可夫性: 给定某变量的邻接变量, 则它条件独立于其他变量: $n^{*}(v) = n(v) \cup \{v\}$, 有 $\mathbf{x}_{v} \perp \mathbf{x}_{V \backslash n^{*}(v)} \mid \mathbf{x}_{n(v)}$. 成对马尔可夫性: 若 $\langle u, v\rangle \notin E$, 则 $\mathbf{x}_{u} \perp \mathbf{x}_{v} \mid \mathbf{x}_{V \backslash\langle u, v\rangle}$. 势函数: 标定模型偏好变量之间的相关性, 非负⇒指数函数: $\psi_{Q}\left(\mathbf{x}_{Q}\right) = e^{-H_{Q}\left(\mathbf{x}_{Q}\right)}$, $H_{Q}\left(\mathbf{x}_{Q}\right)$ 是一个定义在变量 $\mathbf{x}_{Q}$ 上的实值函数常见形式为 $H_{Q}\left(\mathbf{x}_{Q}\right) = \sum_{u, v \in Q, u \neq v} \alpha_{uv} x_{u} x_{v} + \sum_{v \in Q} \beta_{v} x_{v}$. CRF条件随机场判别式无向图模型: 目标构建条件概率模型 $P(y \mid x)$, $y$可以是结构型变量(分量之间有相关性). $n(v)$表示$v$的邻接结点, 每个变量$y_{v}$都满足马尔可夫性: $P\left(y_{v} \mid \mathbf{x}, \mathbf{y}_{V \backslash\{v\}}\right) = P\left(y_{v} \mid \mathbf{x}, \mathbf{y}_{n(v)}\right)$, 则$(\mathbf{y}, \mathbf{x})$构成CRF. CRF类似MRF 用势函数和图结构上的团定义条件概率 $P(\mathbf{y} \mid \mathbf{x}) = \frac{1}{Z} \exp \left(\sum_{j} \sum_{i=1}^{n-1} \lambda_{j} t_{j}\left(y_{i+1}, y_{i}, \mathbf{x}, i\right)+\sum_{k} \sum_{i=1}^{n} \mu_{k} s_{k}\left(y_{i}, \mathbf{x}, i\right)\right)$ 其中 $t_{j}$是定义在观测序列的两个相邻标记位置上的转移特征函数: 刻画相邻标记变量间的相关关系和观测序列对它们的影响; $s_{k}$是定义在观测序列的标记位置$i$上的状态特征函数: 用于刻画观测序列对标记变量的影响. 例子: $t_{j}\left(y_{i+1, i}, \mathbf{x}, i\right) = 1$, if $x_{i} =$ ”knock”, then $y_{i+1} = [P], y_{i} = [V]$ 其他为0; $s_{k}\left(y_{i}, \mathbf{x}, i\right) = 1$, if $x_{i} =$ ”knock”, then $y_{i} = [V]$, 形式上MRF和CRF没有区别都是势函数,但MRF是联合概率. 学习与推断: 基于联合概率分布对目标变量的边际分布或以某些可观测变量为条件的条件分布进行推断, 边际分布是对无关变量求和或积分得到的结果. 概率图模型需要确定具体分布的参数⇒参数估计用MAP/MLE, 也可以看出待推测的变量用推断做. 推断: $P\left(\mathbf{x}_{F} \mid \mathbf{x}_{E}\right) = P\left(\mathbf{x}_{E}, \mathbf{x}_{F}\right)/P\left(\mathbf{x}_{E}\right) = P\left(\mathbf{x}_{E}, \mathbf{x}_{F}\right)/\sum_{\mathbf{x}_{F}} P\left(\mathbf{x}_{E}, \mathbf{x}_{F}\right)$, $P\left(\mathbf{x}_{E}, \mathbf{x}_{F}\right)$由概率图模型获得, 关键在于高效地计算边际分布$P(\mathbf{x}_{E})$. 精确推断: 计算精确值, 一类动规算法, 利用图模型描述的条件独立性来削减计算目标概率的计算量. 变量消去: 设贝叶斯网络结构: 1→2→3→4 3→5. 推断的目标是计算$P(x_{5})$: $P\left(x_{5}\right) = \sum_{x_{4}} \sum_{x_{3}} \sum_{x_{2}} \sum_{x_{1}} P\left(x_{1}, x_{2}, x_{3}, x_{4}, x_{5}\right)$

$= \sum_{x_4}\sum_{x_3}\sum_{x_2}\sum_{x_1} P(x_1)P(x_2\mid x_1)P(x_3\mid x_2)P(x_4\mid x_3)P(x_5\mid x_3)$, 一层一层地加, 采用加法顺序1-4: $\{x_1,\cdots,x_4\}$: $P(x_5) = \sum_{x_3}P(x_5\mid x_3)\sum_{x_4}P(x_4\mid x_3)\sum_{x_2}P(x_3\mid x_2)\sum_{x_1}P(x_1)P(x_2\mid x_1) = \sum_{x_3}P(x_5\mid x_3)\sum_{x_4}P(x_4\mid x_3)\sum_{x_2}P(x_3\mid x_2)m_{12}(x_2) = \sum_{x_3}P(x_5\mid x_3)m_{23}(x_3)\sum_{x_4}P(x_4\mid x_3) = m_{35}(x_5)$. 最后仅与$x_5$的取值有关. 对无向图也同样适用. 变量消去缺点: 计算多个边际分布, 重复使用变量消去将会造成大量冗余计算. 信念传播: 将变量消去法中的求和操作看作一个消息传递过程, 较好地解决了求解多个边际分布时的重复计算问题: $m_{ij}(x_j)=\sum_{x_i}\psi(x_i,x_j)\prod_{k\in n(i)\setminus j}m_{ki}(x_i)$, 其中 $n(i)$ 是邻接结点, 一个结点仅在接收到来自其他所有结点的消息后才能向另一个结点发送消息, 且结点的边际分布正比于它所接收的消息的乘积: $P(x_i)\propto\prod_{k\in n(i)}m_{ki}(x_i)$. 如果图结构中没有环: 则信念传播两个步骤可完成所有消息传递, 进而计算所有变量上的边际分布: 1)指定一个根结点, 从所有叶结点开始向根结点传递消息, 直到根结点收到所有邻接结点的消息; 从根结点开始向叶结点传递消息, 直到所有叶结点均收到消息. $x_4\to x_3: m_{43}(x_3)$. 近似推断: 两大类 1采样 2变分推断. MCMC采样: 计算概率分布是为了计算期望, 不如直接计算/逼近期望. 关键在构造"平稳分布为$p$的马尔可夫链"来产生样本: 平稳条件如下 $\mathbf{x},\mathbf{x}'$ 是两个状态则 $p(\mathbf{x}^t)T(\mathbf{x}^{t-1}\mid\mathbf{x}^t)=p(\mathbf{x}^{t-1})T(\mathbf{x}^t\mid\mathbf{x}^{t-1})$ 此时$p(\mathbf{x})$ 是该马尔可夫链的平稳分布且满足条件时已收敛到平稳状态. MH算法: MCMC的重要代表, 基于"拒绝采样"来逼近平稳分布$p$, 算法每次根据上一轮采样结果 $\mathbf{x}^{t-1}$ 来采样获得候选状态样本 $\mathbf{x}^*$, 但这个候选样本会以一定的概率被"拒绝"掉. 假定从状态 $\mathbf{x}^{t-1}$ 到状态 $\mathbf{x}^*$ 的转移概率为 $Q(\mathbf{x}^*\ \ \mathbf{x}^{t-1})A(\mathbf{x}^*\mid\mathbf{x}^{t-1})$, 其中 $Q(\mathbf{x}^*\mid\mathbf{x}^{t-1})$ 是用户给定的先验概率, $A(\mathbf{x}^*\mid\mathbf{x}^{t-1})$ 是 $\mathbf{x}^*$ 被接受的概率. 若 $\mathbf{x}^*$ 最终收敛到平稳状态, 则根据上述平稳条件有: $p(\mathbf{x}^{t-1})Q(\mathbf{x}^*\mid\mathbf{x}^{t-1})A(\mathbf{x}^*\mid\mathbf{x}^{t-1}) = p(\mathbf{x}^*)Q(\mathbf{x}^{t-1}\mid\mathbf{x}^*)A(\mathbf{x}^{t-1}\mid\mathbf{x}^*)$. MH算法伪代码: 循环{ 根据$Q(x^*\mid x^{t-1})$采样出候选样本$x^*$, 根据均匀分布采样出阈值$u$, 如果$u\leqslant A(\mathbf{x}^*\mid\mathbf{x}^{t-1})$, $x^t=x^*$, 否则 $x^t=x^{t-1}$ }, 最后返回采样的样本序列$x^1\ldots$. 为了达到平稳状态, 设置接受率为: $A(\mathbf{x}^*\mid\mathbf{x}^{t-1}) = \min\left(1, p(\mathbf{x}^*)Q(\mathbf{x}^{t-1}\mid\mathbf{x}^*)/(p(\mathbf{x}^{t-1})Q(\mathbf{x}^*\mid\mathbf{x}^{t-1}))\right)$. Gibbs采样是MH算法的特例, 以下步骤: 1)随机或以某个次序选取某变量 $x_i$; 2)根据 $\mathbf{x}$ 中除 $x_i$ 外的变量的现有取值, 计算条件概率 $p(x_i\mid\mathbf{x}_{\bar{i}})$, 其中 $\mathbf{x}_{\bar{i}}=\{x_1,x_2,\ldots,x_{i-1},x_{i+1},\ldots,x_N\}$, 3)根据 $p(x_i\mid\mathbf{x}_{\bar{i}})$ 对变量 $x_i$ 采样, 用采样值代替原值. 变分推断: 通过使用已知简单分布来逼近需推断的复杂分布, 并通过限制近似分布的类型, 从而得到一种局部最优但具有确定解的近似后验分布. $p(\mathbf{x}\mid\Theta)=\prod_{i=1}^N\sum_{\mathbf{z}}p(x_i,\mathbf{z}\mid\Theta)$. 上式取对数似然, 用EM算法: 在 E 步, 根据 $t$ 时刻的参数 $\Theta^t$ 对 $p(\mathbf{z}\mid\mathbf{x},\Theta^t)$ 进行推断, 并计算联合似然函数 $p(\mathbf{x},\mathbf{z}\mid\Theta)$; 在 M 步, 基于 E 步的结果进行最大化寻优, 即对关于变量 $\Theta$ 的函数 $\mathcal{Q}(\Theta;\Theta^t)$ 进行最大化从而求取 $\Theta^{t+1}=\arg\max\limits_{\Theta}\mathcal{Q}(\Theta;\Theta^t)=\arg\max\limits_{\Theta}\sum_{\mathbf{z}}p(\mathbf{z}\mid\mathbf{x},\Theta^t)\ln p(\mathbf{x},\mathbf{z}\mid\Theta)$. $\ln p(x)=\int q(z)\ln p(x)dz=\int q(z)\left(\ln\frac{p(x,z)}{q(z)}-\ln\frac{p(z|x)}{q(z)}\right)=\mathcal{L}(q)+\mathrm{KL}(q\|p)$, 在现实任务中E步对$p(z\mid x,\Theta^t)$的推断很可能因$z$模型复杂而难以进行, 此时可借助变分推断, 假设$z$服从: $q(\mathbf{z})=\prod_{i=1}^M q_i(\mathbf{z}_i)$. LDA话题模型: 词袋词频的直方图, $\Theta_{t,k}$ 即表示文档$t$中包含话题$k$的比例: 1) 根据参数为 $\alpha$ 的狄利克雷分布随机采样一个话题分布 $\Theta_t$; 2) 生成文档中的 $N$ 个词: 1根据 $\Theta_t$ 进行话题指派, 得到文档 $t$ 中词 $n$ 的话题 $z_{t,n}$ 2根据指派的话题所对应的词频分布 $\boldsymbol{\beta}_k$ 随机采样生成词. 文档中的词频是唯一的已观测变量, 它依赖于对这个词进行的话题指派$z_{t,n}$, 以及话题对应的词频$\beta_k$, 同时话题指派$z_{t,n}$依赖于话题的分布$\Theta_t$, $\Theta_t$依赖于狄利克雷分布的参数 $\alpha$, 而话题词频依赖于参数$\eta$. 对应的概率分布 $p(\mathbf{W},\mathbf{z},\boldsymbol{\beta},\Theta\mid\boldsymbol{\alpha},\boldsymbol{\eta}) = \prod_{t=1}^T p(\Theta_t\mid\boldsymbol{\alpha})\prod_{i=1}^K p(\boldsymbol{\beta}_k\mid\boldsymbol{\eta})\left(\prod_{n=1}^N P(w_{t,n}\mid z_{t,n},\boldsymbol{\beta}_k)P(z_{t,n}\mid\Theta_t)\right)$ 其中 $p(\Theta_t\mid\boldsymbol{\alpha})$ 和 $p(\boldsymbol{\beta}_k\mid\boldsymbol{\eta})$ 通常分别设置为以 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\eta}$ 为参数的$K$维和$N$维狄利克雷分布, 例如 $p(\Theta_t\mid\boldsymbol{\alpha})=\frac{\Gamma(\sum_k\alpha_k)}{\prod_k\Gamma(\alpha_k)}\prod_k\Theta_{t,k}^{\alpha_k-1}$. 寻找$\alpha$和$\eta$以最大化对数似然 $LL(\boldsymbol{\alpha},\boldsymbol{\eta})=\sum_{t=1}^T\ln p(\boldsymbol{w}_t\mid\boldsymbol{\alpha},\boldsymbol{\eta})$, $p(\boldsymbol{w}_t\mid\boldsymbol{\alpha},\boldsymbol{\eta})$ 不易计算, 采用Gibbs采样或变分法来求取近似解. 参数 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\eta}$ 已确定, 则根据词频 $w_{t,n}$ 来推断 $\Theta_t,\boldsymbol{\beta}_k$ 和 $z_{t,n}$ 求解 $p(\mathbf{z},\boldsymbol{\beta},\Theta\mid\mathbf{W},\boldsymbol{\alpha},\boldsymbol{\eta})=\frac{p(\mathbf{W},\mathbf{z},\boldsymbol{\beta},\Theta|\boldsymbol{\alpha},\boldsymbol{\eta})}{p(\mathbf{W}|\boldsymbol{\alpha},\boldsymbol{\eta})}$.

ch15 规则学习, 优点: 1更好的可解释性, 2数理逻辑具有极强的表达能力. 冲突消解: 多条规则判别结果不同. 命题规则由"原子命题"和逻辑连接词(与或非蕴含)构成的简单陈述句; 一阶规则加入任意, 存在量词. 序贯覆盖: 逐条归纳, 在训练集上学到一条规则, 就将其覆盖的训练样例去除. 刚开始以第一个样例(青绿蜷缩..)属性赋值生成规则: "色泽=青绿"如果无法仅覆盖正例, 则继续加"根蒂=蜷缩", 加起来判断直到仅覆盖正例. 但常常组合爆炸. 改进: 1自顶向下: 从比较一般的规则开始, 逐渐添加新文字缩小覆盖范围. 2自底向上, 从比较特殊的规则开始, 逐渐删除文字以扩大覆盖范围. 适用于训练样本更少的情形; 一阶规则学习(复杂情况). "生成-测试"法, 适用于产生泛化性能好的规则, 命题规则学习. 覆盖准确率: $m_+/m$, 训过程: 1算出所有属性的所有取值对应的准确率, 2选最高并删掉覆盖的样例, 继续第1步. 如上是贪心搜索, 还可以集束搜索: 每层搜索保留$b$个, 在下一层加入新的时候都要评估. 剪枝: 增/删规则逻辑文字前后的性能. 借助统计显著性检验: CN2算法用似然率统计量LRS 衡量规则集覆盖的样例分布与训练集经验分布的差别, LRS越大⇒预测与直接猜(在正/负类比例下)的差别. 后剪枝: REP剪错剪枝复杂度$\mathcal{O}(m^4)$, 对规则集穷举所有可能的剪枝操作, 然后评估最好的. RIPPER: 剪枝与后处理优化(将$\mathcal{R}$中所有规则放在一起优化)相结合, 缓解贪心的局部性, 循环: $\mathcal{R}'=\mathrm{PostOpt}(\mathcal{R})$ $D_i=\mathrm{NotCovered}(\mathcal{R}',D)$ $\mathcal{R}_i=\mathrm{IREP}^*(D_i)$ $\mathcal{R}=\mathcal{R}'\cup\mathcal{R}_i$. 一阶规则学习(此时数据集也变了, 标签变为更好$(X,Y)$): 命题逻辑难以处理对象之间的关系, 这里相互比较: 比···更好. 关系数据集例子: 根蒂更蜷$(X,Y)$. 优点容易引入领域知识. FOIL算法: 遵循序贯覆盖框架自顶向下, 最初始空规则: 更好$(X,Y)\leftarrow\cdot$然后候选文字是所有属性于$X,Y$的比较, 使用F_Gain $=\hat{m}_+\times\left(\log_2\frac{\hat{m}_+}{\hat{m}_++\hat{m}_-}-\log_2\frac{m_+}{m_++m_-}\right)$来选择文字, 其中$\hat{m}_+$表示增加候选文字后新规则所覆盖的正样例数, 这里F_Gain仅考虑正例的信息量. ILP归纳逻辑程序设计: 在一阶规则学习中引入了函数和逻辑表达式嵌套: 1具备更强表达能力, 2解决基于背景知识的逻辑程序归纳. 难点: 函数和逻辑表达式的嵌套使候选原子公式可能无穷多. ILP一般自底向上直接将一个或多个正例所对应的具体事实作为初始规则, 再对规则逐步进行泛化以增加其对样例的覆盖率. LGG最小一般泛化: 规则对应的特殊关系数据样例是特殊的 $\to$ 一般的. 在归纳逻辑程序设计中, 获得LGG之后可将其看做单条规则加入规则集. 逆归结: 演绎是从一般性规律出发来探讨具体事物, 归纳是从个别事物出发概括出一般性规律.

ch16 MDP: 环境$E$, 动作空间$A$, 状态空间$X$, 潜在的状态转移函数$P$, 潜在的奖赏函数$R$. 状态转移/奖赏返回不受机器控制, 机器只能选择执行的动作. 学策略 $\boldsymbol{\pi}$, $\sum_a\boldsymbol{\pi}(x,a)=1$. 找长期累积奖赏最大化的策略. $T$步累积奖赏$\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^T r_t\right]$, $\gamma$折扣累积奖赏$\mathbb{E}\left[\sum_{t=0}^{+\infty}\gamma^t r_{t+1}\right]$, 强化学习是延迟标记信息的监督学习. 探索-利用窘境: 探索平均分配每种动作最后得到期望的近似估计, 利用动作为当前平均奖赏最大的. $\epsilon$-贪心: 以$\epsilon$的概率进行探索(从均匀分布选动作), 其他时候用当前最优: $v_n$为当前获得的奖赏则同步更新的平均奖赏为式$\star$: $Q_n(k)=\frac{1}{n}((n-1)\times Q_{n-1}(k)+v_n)$. Softmax: 基于当前已知的平均奖赏来对探索和利用进行折中, 由Boltzmann分布: $P(k)=e^{\frac{Q(k)}{\tau}}/\sum_{i=1}^K e^{\frac{Q(i)}{\tau}}$, 其中$\tau>0$是温度, $\to 0$是仅利用 $\to +\infty$是仅探索. model-based: 在已知模型的环境中学习, 对$\forall$状态$x,x'$和执行动作$a$ 转移概率 $P_{x\to x'}^a$, 以及带来的奖赏 $R_{x\to x'}^a$, 也是已知的. 策略评估: 状态值函数 $V^\pi(x)$ 为从状态$x$出发使用策略$\pi$所带来的的累积奖赏, 状态-动作值函数 $Q^\pi(x,a)$表示从状态$x$出发执行动作$a$后再使用策略$\pi$带来的累积奖赏, eg: $\gamma$折扣累积奖赏: $V_\gamma^\pi(x)=\mathbb{E}_\pi\left[\sum_{t=0}^{+\infty}\gamma^t r_{t+1}\mid x_0=x\right]$, $Q_\gamma^\pi(x,a)=\mathbb{E}_\pi\left[\sum_{t=0}^{+\infty}\gamma^t r_{t+1}\mid x_0=x,a_0=a\right]$. MDP具有马尔可夫性, Bellman等式: $V_T^\pi(x)=\mathbb{E}_\pi\left[\frac{1}{T}\sum_{t=1}^T r_t\mid x_0=x\right]=\mathbb{E}_\pi\left[\frac{1}{T}r_1+\frac{T-1}{T}\frac{1}{T-1}\sum_{t=2}^T r_t\mid x_0=x\right]=\sum_{a\in A}\pi(x,a)\sum_{x'\in X}P_{x\to x'}^a\left(\frac{1}{T}R_{x\to x'}^a+\frac{T-1}{T}\mathbb{E}_\pi\left[\frac{1}{T-1}\sum_{t=1}^{T-1}r_t\mid x_0=x'\right]\right)=\sum_{a\in A}\pi(x,a)\sum_{x'\in X}P_{x\to x'}^a\left(\frac{1}{T}R_{x\to x'}^a+\frac{T-1}{T}V_{T-1}^\pi(x')\right)$. 对于 $\gamma$ 折扣累积奖赏: $V_\gamma^\pi(x)=\sum_{a\in A}\pi(x,a)\sum_{x'\in X}P_{x\to x'}^a\left(R_{x\to x'}^a+\gamma V_\gamma^\pi(x')\right)$. 由上述递归等式计算值函数实际上为动规算法迭代, 停止准则为$V(x)\ V'(x)$对于$\max_x$相差一个阈值, 由此计算状态-动作值函数(式③): $Q_T^\pi(x,a)=\sum_{x'\in X}P_{x\to x'}^a\left(\frac{1}{T}R_{x\to x'}^a+\frac{T-1}{T}V_{T-1}^\pi(x')\right)$, $Q_\gamma^\pi(x,a)=\sum_{x'\in X}P_{x\to x'}^a\left(R_{x\to x'}^a+\gamma V_\gamma^\pi(x')\right)$. 策略改进: 理想策略是最大化累积奖赏, 对Bellman等式做改进对动作的求和取最优(式④): $V_T^*(x)=\max_{a\in A}\sum_{x'\in X}P_{x\to x'}^a\left(\frac{1}{T}R_{x\to x'}^a+\frac{T-1}{T}V_{T-1}^*(x')\right)$ $V_\gamma^*(x)=\max_{a\in A}\sum_{x'\in X}P_{x\to x'}^a\left(R_{x\to x'}^a+\gamma V_\gamma^*(x')\right)$, $V^*(x)=\max_{a\in A}Q^{\pi^*}(x,a)$, 代入上面式③得到最优Bellman等式: $Q_T^*(x,a)=\sum_{x'\in X}P_{x\to x'}^a\left(\frac{1}{T}R_{x\to x'}^a+\frac{T-1}{T}\max_{a'\in A}Q_{T-1}^*(x',a')\right)$ $Q_\gamma^*(x,a)=\sum_{x'\in X}P_{x\to x'}^a\left(R_{x\to x'}^a+\gamma\max_{a'\in A}Q_\gamma^*(x',a')\right)$. 改变动作的条件为 $V^\pi(x)\leqslant Q^\pi\left(x,\pi^i(x)\right)$, $V^\pi(x)\leqslant Q^\pi(x,\pi'(x))=\sum_{x'\in X}P_{x\to x'}^{\pi'(x)}\left(R_{x\to x'}^{\pi'(x)}+\gamma V^\pi(x')\right)\leqslant\sum_{x'\in X}P_{x\to x'}^{\pi'(x)}\left(R_{x\to x'}^{\pi'(x)}+\gamma Q^\pi(x',\pi'(x'))\right)=\cdots=V^{\pi'}(x)$, 所以值函数对于策略的改进是单调增的, 所以$\pi'(x)=\arg\max_{a\in A}Q^\pi(x,a)$的迭代是可的, 直到$\pi'$和$\pi$差不多即满足最优Bellman等式. 策略迭代: 评估策略的值函数与策略评估后改进值最优策略结合的迭代. 值迭代算法: 根据式④更新值函数直到更新后的相差一个阈值. model-based方法能归结为基于动规的寻优问题, 估计状态值函数$V$, 最终策略通过状态-动作值函数$Q$获得. model-free: 不依赖环境建模, 策略无法评估因为模型未知导致无法做全概率展开, 蒙特卡罗强化学习: 多次采样替代策略评估, 平均累积奖赏作为期望累积奖赏的近似. 模型未知时$V\to Q$的转换是困难的. 多次采样就是综合多个轨迹的奖赏后得到状态-动作值函数$Q$的估计, 同策略(on-policy, 被评估与被改进是同一个策略) $\epsilon$-贪心: 策略评估用式$\star$, 改进用得到不同的轨迹需要用不同的策略$\pi^\epsilon$: 即以$\epsilon$概率均匀概率选取动作, 否则采用原始策略 $\arg\max_{a'}Q(x,a')$. 异策略(off-policy)蒙特卡罗强化学习: 使用策略$\pi'$的采样轨迹来评估策略$\pi$ 则对累积奖赏加权: $Q(x,a)=\frac{1}{m}\sum_{i=1}^m\frac{P_i^\pi}{P_i^{\pi'}}r_i$, $P^\pi=\prod_{i=0}^{T-1}\pi(x_i,a_i)P_{x_i\to x_{i+1}}^{a_i}$是策略产生第$i$条轨迹的概率, $R_i$是第$i$条轨迹的累积奖赏. 伪代码: 因为$\epsilon$贪心, $p_i$为 1选择原策略的概率: $1-\epsilon+\epsilon/|A|$, 或2平均取的概率: $\epsilon/|A|$; 策略评估: $R=\frac{1}{T-t}\sum_{i=t+1}^T\left(r_i\times\prod_{j=i}^{T-1}\frac{1}{p_j}\right)$, $Q(x_t,a_t)=(Q(x_t,a_t)\times\mathrm{count}(x_t,a_t)+R)/(\mathrm{count}(x_t,a_t)+1)$, 在一条轨迹后进行策略改进$\arg\max_{a'}Q(x,a')$. TD时序差分学习: (因为蒙特卡罗强化学习没有充分利用MDP结构, 采样一个轨迹后才能更新值估计不够高效), 利用值函数$Q$的增量更新来评估策略 (Sarsa算法): 迭代 1: $a'=\pi^\epsilon(x')$ (因为 $Q_{t+1}^\pi(x,a)=Q_t^\pi(x,a)+\frac{1}{t+1}(r_{t+1}-Q_t^\pi(x,a))$, $Q^\pi(x,a)=\sum_{x'\in X}P_{x\to x'}^a\left(R_{x\to x'}^a+\gamma\sum_{a'\in A}\pi(x',a')Q^\pi(x',a')\right)$) 2: 通过增量求和有 $Q_{t+1}^\pi(x,a)=Q_t^\pi(x,a)+\alpha\left(R_{x\to x'}^a+\gamma Q_t^\pi(x',a')-Q_t^\pi(x,a)\right)$, 其中更新步长$\alpha$越大则越靠后的累积奖赏越重要 3: 式$\star$的$Q$更新. Q学习算法中仅在策略评估时$a'=\pi(x')$不同. 值函数近似: 值函数是关于有限状态的表格值函数, 状态空间离散化后, $E_{\boldsymbol{\theta}}=\mathbb{E}_{\boldsymbol{x}\sim\pi}\left[(V^\pi(\boldsymbol{x})-V_{\boldsymbol{\theta}}(\boldsymbol{x}))^2\right]$ 梯度下降 $-\frac{\partial E_{\boldsymbol{\theta}}}{\partial\boldsymbol{\theta}}=\mathbb{E}_{\boldsymbol{x}\sim\pi}\left[2(V^\pi(\boldsymbol{x})-V_{\boldsymbol{\theta}}(\boldsymbol{x}))\frac{\partial V_{\boldsymbol{\theta}}(\boldsymbol{x})}{\partial\boldsymbol{\theta}}\right]=\mathbb{E}_{\boldsymbol{x}\sim\pi}\left[2(V^\pi(\boldsymbol{x})-V_{\boldsymbol{\theta}}(\boldsymbol{x}))\boldsymbol{x}\right]$, $\boldsymbol{\theta}=\boldsymbol{\theta}+\alpha(V^\pi(\boldsymbol{x})-V_{\boldsymbol{\theta}}(\boldsymbol{x}))\boldsymbol{x}$ 是对于单个样本的更新规则. 借助时序差分学习, 不知道策略的真实函数$V^\pi$时用当前估计的值函数代替真实值函数: $\boldsymbol{\theta}=\boldsymbol{\theta}+\alpha\left(r+\gamma\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{x}'-\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{x}\right)\boldsymbol{x}$. 模仿学习: 从人类专家的决策过程范例中学习, 直接模仿学习: 模仿人类专家"状态-动作对"数据集, 状态: 特征, 动作: 标记, 学策略模型可作为强化学习初始策略. 逆强化学习: 从人类专家提供的反例数据中反推出奖赏函数. 基本思想: 欲使机器做出与范例一致的行为, 等价于在某个奖赏函数的环境中求解最优策略, 该最优策略产生的轨迹与范例数据一致. 寻找某种奖赏函数使范例数据最优, 使用该奖赏函数训练强化学习策略. $\rho^\pi=\mathbb{E}\left[\sum_{t=0}^{+\infty}\gamma^t R(\boldsymbol{x}_t)\mid\pi\right]=\mathbb{E}\left[\sum_{t=0}^{+\infty}\gamma^t\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_t\mid\pi\right]=\boldsymbol{w}^{\mathrm{T}}\mathbb{E}\left[\sum_{t=0}^{+\infty}\gamma^t\boldsymbol{x}_t\mid\pi\right]$, 对于最优奖赏函数 $R(\boldsymbol{x})=\boldsymbol{w}^{*\mathrm{T}}\boldsymbol{x}$ 和任意其他策略产生的 $\overline{\boldsymbol{x}}^\pi$, 有 $\boldsymbol{w}^{*\mathrm{T}}\overline{\boldsymbol{x}}^*-\boldsymbol{w}^{*\mathrm{T}}\overline{\boldsymbol{x}}^\pi=\boldsymbol{w}^{*\mathrm{T}}(\overline{\boldsymbol{x}}^*-\overline{\boldsymbol{x}}^\pi)\geqslant 0$, $\boldsymbol{w}^*=\arg\max\limits_{\boldsymbol{w}}\min_{\boldsymbol{\pi}}\boldsymbol{w}^{\mathrm{T}}(\overline{\boldsymbol{x}}^*-\overline{\boldsymbol{x}}^\pi)$ s.t. $\|\boldsymbol{w}\|\leqslant 1$. 加油冲冲冲, 越哥罩我!!