

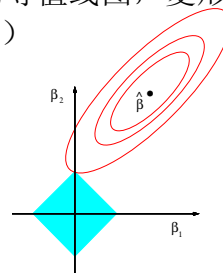
人工智能导论（考试样题）

一、 选择题（每题 4 分，均为单选，请直接在试卷上作答）

1. 关于搜索的表述中，下面说法中不正确的是 ()
 - A. 宽度优先搜索 (BFS) 是一致代价搜索 (UCS) 的一种特殊情况
 - B. 宽度优先搜索 (BFS) 是完备的
 - C. 树搜索的深度优先搜索 (DFS) 是完备的
 - D. 一致代价搜索 (UCS) 是A*搜索的一种特殊情况
2. 对于 3 维迷宫搜索问题，目标为 (x, y, z) 只能沿着坐标轴方向搜索，每步代价均为 1，对于状态 (u, v, w) ，下面哪个启发式函数不是可采纳的 ()
 - A. $h(u, v, w) = \sqrt{|x - u|} + \sqrt{|y - v|} + \sqrt{|z - w|}$
 - B. $h(u, v, w) = \sqrt{(x - u)^2 + (y - v)^2 + (z - w)^2}$
 - C. $h(u, v, w) = |x - u| + |y - v| + |z - w|$
 - D. $h(u, v, w) = (x - u)^2 + (y - v)^2 + (z - w)^2$
3. 对于有向无环图上的搜索，若定义点*i*和点*j*之间的代价为 c_{ij} 时，UCS 算法与 BFS 算法搜索路径完全一致。若重新定义代价 c'_{ij} ，可以使 UCS 算法与 DFS 算法搜索路径完全一致，则 ()
 - A. $c'_{ij} = 1$
 - B. $c'_{ij} = -c_{ij}$
 - C. $c'_{ij} = c_{ij} + \alpha$
 - D. 无法实现
4. 以下关于模型选择 (Model Selection) 的表述中，正确的是 ()
 - A. 应当在测试数据 (Test Data) 上进行模型选择
 - B. 由没有免费的午餐 (No Free Lunch) 定理可知，采用模型选择并不能获得对具体的机器学习问题效果更好的模型
 - C. 模型选择既包括对学习算法的选择，也包括对超参数的调优
 - D. 处理大数据时，通常使用 Leave-One-Out Cross Validation (LOOCV)
5. 以下对于 K-近邻方法的描述中，正确的是 ()
 - A. 当 K 增大时，维度灾难问题更严重
 - B. 既可以用于分类任务，也可以用于回归任务
 - C. K-近邻分类器的错误率是贝叶斯错误率
 - D. 当 K 减小时，模型的分类面一般会更加光滑，不易发生过拟合

6. 下图中 β 为学习器参数，椭圆为无正则项的损失函数的等值线图，菱形区域为正则项的约束区域，该图中使用了哪种正则项？（ ）

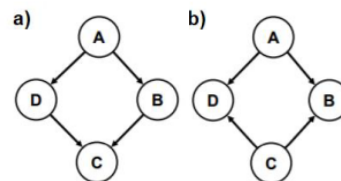
A. L2
C. L1
B. L ∞
D. L1 + L2



7. 下图所示的网络中，以下陈述不成立的是（ ）

(1) $A \perp C | B, D$ (2) $B \perp D | A, C$

A. 图 a) 中 (1) 成立 B. 图 b) 中 (1) 不成立
C. 图 a) 中 (2) 成立 D. 图 b) 中 (2) 成立



8. 下列描述中不正确的一项是（ ）

A. 马尔可夫随机场 (Markov Random Field) 中节点的马尔可夫边界 (Markov boundary) 为该节点的邻居。
B. EM 算法中 M step 可选用梯度下降算法。
C. K-Means 算法可以理解为退化的高斯混合模型。
D. 对于任意的起始分布马尔可夫链具有极限分布。

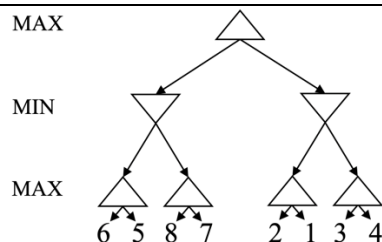
9. 在隐变量模型中， $q(z)$ 是隐藏变量 z 的概率质量函数，请问对数似然函数 $\log p(x|\theta)$ 的 ELBO 下界可表示为（ ）

A. $\sum_z q(z) \log \left(\frac{p(x, z|\theta)}{q(z)} \right)$
B. $\sum_z p(x, z|\theta) \log \left(\frac{p(x, z|\theta)}{q(z)} \right)$
C. $\sum_z q(z) \log \left(\frac{q(z)}{p(x, z|\theta)} \right)$
D. $\sum_z p(x, z|\theta) \log \left(\frac{q(z)}{p(x, z|\theta)} \right)$

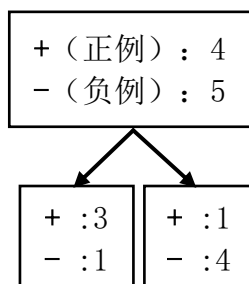
10. 在 Gibbs 采样算法过程中，下列表述有误的一项是（ ）

A. 初始状态 $(x_1^0 \dots x_n^0)$ 可采样均匀分布随机初始化
B. x_1 的更新方式为 $x_1^{t+1} = \arg\max_{x_1} P(x_1 | x_2^t \dots x_n^t)$
C. 对于马尔可夫随机场模型，更新变量仅需考虑变量的邻居节点
D. 算法的最初几轮迭代的结果不能作为有效采样数据

二、填空题（每题 5 分，请直接在试卷上作答）



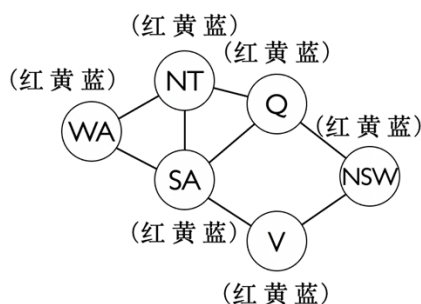
1. 在上图所示的极小极大树中，底层节点从左向右扩展，若使用 Alpha-Beta 剪枝时，则哪些节点将不会被访问到_____（5 分）。
2. 某二分类数据集及其上的一种分割 (Split) 如图所示，请写出该分割的信息增益



- (Information Gain) 的计算式_____（3 分）；
决策树_____（是/否）（2 分）可以用于线性不可分数据的分类。
3. 在使用 Metropolis-Hastings 算法采样时，为满足转移收敛条件 $\pi_i q_{ij} = \pi_j q_{ji}$ ，需要引入 α_{ij} 使得_____（1 分）；但在实际算法中，我们将 α_{ij} 设置为_____（2 分），如此设计的优势/原因为_____（2 分）。

三、解答题（每题 15 分，请选择 3 题作答，若 4 题均答，则仅取得分最高的 3 题计入考试成绩，请在答题纸上作答）

1. 考虑染色问题，要求相邻点不可以染同样颜色，初始状态如下：



- (a) 写出依次对 WA、NT 执行 AC-3 算法的过程及所得结果。（5 分）
- (b) AC-3 在 X_i 的值域中删除任何值时，都把每条边弧 (X_k, X_i) 放回到队列里，即使 X_k 中的每个值都和 X_i 的一些剩余值相容。假设对每条弧 (X_k, X_i) ，记录 X_i 中与 X_k 的每个值都相容的剩余值的个数。如何有效的更新这些数字使得弧相容算法的复杂度为 $O(n^2 d^2)$ 。（10 分）

2. 对于训练数据集 $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)$ 。考虑二分类问题，即 $y \in \{0, 1\}$ ，常使用伯努利分布（Bernoulli distribution）来表示样例 \mathbf{x} 的正负类别的概率：

$$P(y = 1 | \mathbf{w}^T \mathbf{x}) = \frac{1}{1 + e^{-\mathbf{w}^T \mathbf{x}}}.$$

其中， \mathbf{w} 是线性模型的参数向量。请解答以下问题。

- (a) 证明： \mathbf{w} 的极大似然估计方法等价于最小化以下的交叉熵损失函数：

$$-\frac{1}{n} \sum_{i=1}^n \{y_i \log \sigma(\mathbf{x}_i^T \mathbf{w}) + (1 - y_i) \log[1 - \sigma(\mathbf{x}_i^T \mathbf{w})]\}.$$

其中， $\sigma(t) = \frac{1}{1+e^{-t}}$ 。该方法称为逻辑斯特回归（Logistic Regression）。（5 分）

- (b) 在超大规模应用中，通常在上述损失函数上增加 L1 正则项，如下形式：

$$-\frac{1}{n} \sum_{i=1}^n \{y_i \log \sigma(\mathbf{x}_i^T \mathbf{w}) + (1 - y_i) \log[1 - \sigma(\mathbf{x}_i^T \mathbf{w})]\} + \lambda \|\mathbf{w}\|_1$$

请使用次梯度下降法（Subgradient descent）求解上述最小化问题，并给出完整计算公式和伪代码。（10 分）

3. 贝叶斯网络被广泛应用于医学诊断中，医生会通常根据患者的历史行为与症状判断病因。在新冠疫情流行期间，准确有效的医学检测是疫情防控的重中之重。对于新的病人，医生需要判断该病人患有普通肺炎(R_1)或新冠肺炎(R_2)。医生可以了解到该病人是否为密切接触者(A_1)，以及是否吸烟(A_2)，同时，可以获取肺部影像结果(S_1)，核酸检测结果(S_2)，是否干咳(S_3)以及是否呼吸困难(S_4)。已知密切接触者患有新冠肺炎的概率会更高，吸烟会提升患有普通肺炎的概率。新冠肺炎患者的四种症状均可能为阳性，而普通肺炎只可能导致干咳或者呼吸困难。

- (a) 请画出最符合上述描述的贝叶斯网络图，并写出对应的条件概率乘积的分解形式。（3 分）
- (b) 请写出呼吸困难(S_4)的马尔可夫边界（Markov Boundary）。（3 分）
- (c) 请计算描述（1）中的贝叶斯网络，需要多少个独立参数？如果取消所有独立性假设，又需要多少个独立参数？（注：所有变量均为二值变量）（2 分）
- (d) 在实际情况中，医生要根据所有的行为历史(A_1, A_2)与症状(S_1, S_2, S_3, S_4)判断病人是否患有新冠肺炎(R_2)。请写出对应该场景的条件概率，并利用贝叶斯公式与消元法写出计算过程。（7 分）

4. Baum-Welch 算法是 EM 算法的一种，其解决了隐马尔科夫模型（HMM）三大主要问题中的学习问题。HMM 的学习问题可以按如下方式定义：给定观测序列 $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_T\}$ ，在隐藏序列 $\mathbf{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_T\}$ 未知的情况下，如何估计模型的最佳参数 θ ，使得 $P(\mathbf{X}|\theta)$ 最大？

参数 $\theta = \{\pi, A, B\}$ ，包括初始概率分布 $\pi = [\pi_i]_N$ ，转移（Transition）矩阵 $A = [a_{ij}]_{N \times N}$ ，观测/发射（Emission）矩阵 $B = [b_j(k)]_{N \times M}$ ，其中 N 表示隐状态总数， M 表示可观测状态总数。

(a) 首先进行 E 步的计算，请根据 ELBO 写出 $J(\theta)$ ，并证明：

$$\operatorname{argmax}_{\theta} J(\theta) = \operatorname{argmax}_{\theta} \sum_Z P(X, Z | \theta^{(i)}) \log P(X, Z | \theta)$$

其中 $\theta^{(i)}$ 表示第 i 个优化轮得到的参数。（3 分）

(b) 令 $Q(\theta, \theta^{(i)}) = \sum_Z P(X, Z | \theta^{(i)}) \log P(X, Z | \theta)$ ，请用 $\pi_{z_1}, b_j(k), a_{ij}$ 表示 $P(X, Z | \theta)$ ，并将 $Q(\theta, \theta^{(i)})$ 展开为三项之和，每项只与一个参数有关（3 分）。

(c) 上一小题实现了参数之间的解耦，可以进入 M 步的计算。请利用拉格朗日乘子法计算 $Q(\theta, \theta^{(i)})$ 取极大值时， $\pi_i, a_{ij}, b_j(k)$ 的值（提示：先证 $\sum_Z \log \pi_{z_1} P(X, Z | \theta^{(i)}) = \sum_{i=1}^N \log \pi_i P(X, z_i = i | \theta^{(i)})$ ，拉格朗日乘子法的等式是概率之和为 1）。（4 分）

(d) 定义：

$$\gamma_t(i) = P(z_t = q_i | X, \theta)$$

$$\xi_t(i, j) = P(z_t = q_i, z_{t+1} = q_j | X, \theta)$$

其中 q 表示隐状态，请使用前向概率 $\alpha_t(i)$ 和后向概率 $\beta_t(i)$ 表示 $\gamma_t(i)$ 和 $\xi_t(i, j)$ （提示：这个表示中可以使用 $a_{ij}, b_j(k)$ ）。（5 分）