

## 2019 年 秋 研 究 生 机 器 学 习 试 题

下列各题每个大题 10 分，共 8 道大题，卷面总分 80 分。注意：在给出算法时，非标准（自己设计的）部分应给出说明。特别是自己设置的参数及变量意义要说明。

1. 下面是一个例子集合。其中，两个正例，两个反例。“Yes”表示正例、“No”表示反例。例子有 5 个属性，分别是 Sky、AirTemp、Humidity、Wind、Water。

| Example | Sky    | AirTemp | Humidity | Wind   | Water | EnjoySport |
|---------|--------|---------|----------|--------|-------|------------|
| 1       | Sunny  | Warm    | Normal   | Strong | Cool  | Yes        |
| 2       | Cloudy | Warm    | High     | Strong | Cool  | No         |
| 3       | Sunny  | Cold    | Normal   | Strong | Cool  | No         |
| 4       | Sunny  | Warm    | Normal   | Light  | Warm  | Yes        |

各属性可取的具体值为：Sky(Sunny, Cloudy, Rainy), AirTemp(Warm, Cold), Humidity(Normal, High), Wind(Strong, Light), Water(Warm, Cool)。另外，每个属性还可以取 ?（表示接受任意值）和  $\emptyset$ （表示拒绝所有值）。EnjoySport 是目标概念。在本题中假设的形式规定为属性值的合取。

- 1) 根据上述提供的训练样例和假设表示，手动执行候选消除算法。特别是要写出处理了每一个训练样例后变型空间的特殊边界和一般边界；（8 分）
- 2) 列出最后形成的变型空间中的所有假设。（2 分）

2. 1) 给定训练例子集合如下表

| Day | Outlook  | Temperature | Humidity | Wind   | PlayTennis |
|-----|----------|-------------|----------|--------|------------|
| 1   | Overcast | Hot         | High     | Weak   | Yes        |
| 2   | Rain     | Mild        | High     | Weak   | Yes        |
| 3   | Overcast | Cool        | Normal   | Strong | Yes        |
| 4   | Sunny    | Mild        | Normal   | Strong | Yes        |
| 5   | Overcast | Mild        | High     | Strong | Yes        |
| 6   | Overcast | Hot         | Normal   | Weak   | Yes        |
| 7   | Sunny    | Hot         | High     | Weak   | No         |
| 8   | Sunny    | Hot         | High     | Strong | No         |
| 9   | Rain     | Cool        | Normal   | Strong | No         |
| 10  | Sunny    | Mild        | High     | Weak   | No         |

手动执行 GS 算法，求出 PlayTennis 为 Yes 的规则（即求出覆盖正例的规则）。本题要求写出每步使用的选择子以及每步形成的公式，不用写出每步形成的“中间例子集合”。本题可用类似下面的语言回答：“…第 i 步使用某某选择子形成公式，形成的公式为某某公式；…第 j 步使用某某选择子形成公式，形成的公式为某某公式；…”；（5 分）

- 2) 写出 GS 算法。（5 分）

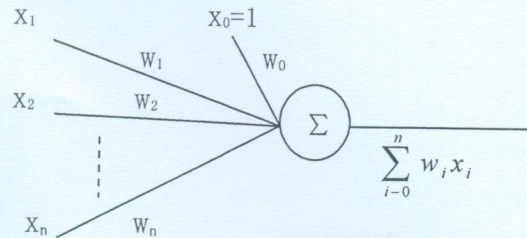
3. 关于神经网络

- (1) 假设一个神经网络有一个隐藏层（有一个隐藏层的神经网络由一个输入层、一个隐藏层、一个输出层组成），写出训练这个神经网络的反向传播算法；（4 分）
- (2) 神经网络训练时，可能出现局部极小的问题，给出两种（可能）克服局部极小的方



法并简要说明为什么这两种方法（可能）克服局部极小；（3分）

- (3) 给定函数  $F(x_1, x_2) = (x_1 + x_2)$ ，其中  $x_1, x_2$  是实数且  $|x_1| \neq |x_2|$ ，规定函数值大于零是正例，函数值小于零为反例。问：“感知器”能否精确划分该函数产生的例子？如果能，说明为什么能？如果不能，说明为什么不能？（提示：F 函数产生的例子有两个属性  $x_1, x_2$ ， $x_1 + x_2 > 0$ ，则这个例子为正例， $x_1 + x_2 < 0$ ，则这个例子为反例。“感知器”结构如下图， $\sum_{i=0}^n w_i x_i > 0$  时，输出 1，否则，输出 -1）（3分）



#### 4. 关于遗传算法

- (1) 写出遗传算法；（6分）
- (2) 给出实现“轮盘赌”的算法（提示：在“轮盘赌”方法中，每个假设  $h$  的适应度  $Fitness(h)$  是已知的）；（4分）

#### 5. 完成下列各题：

- (1) 从你熟悉的角度解释主成分分析(PCA)，并指出它的两种可能应用；（6分）
- (2) 叙述 K-近邻算法的思想。（4分）

#### 6. 请给出 K-means 算法，并指出该方法与高斯混合模型(GMM)之间的联系与差别。（10分）

#### 7. 完成下列各题：

- (1) 给定例子形式如下表：

| Example | Sky   | AirTemp | Humidity | Wind   | Water | Forecast | EnjoySport |
|---------|-------|---------|----------|--------|-------|----------|------------|
| 1       | Sunny | Warm    | Normal   | Strong | Warm  | Same     | Yes        |
| 2       | Rainy | Cold    | High     | Strong | Warm  | Change   | No         |
| 3       | Sunny | Warm    | High     | Strong | Warm  | Same     | Yes        |
| 4       | Sunny | Warm    | High     | Strong | Cool  | Change   | Yes        |

在上面例子集中，各属性可取的具体值为：Sky(Sunny, Cloudy, Rainy), AirTemp(Warm, Cold), Humidity(Normal, High), Wind(Strong, Weak), Water(Warm, Cool), Forecast(Same, Change)。另外，每个属性还可以取？（表示接受任意值）和  $\emptyset$ （表示拒绝所有值）。EnjoySport 是目标概念。在本题中假设的形式规定为属性值的合取。回答如下问题：①“例子空间”中例子的个数是多少（即所有可能的例子的个数是多少）？（要说明如何计算得出的，不能只给出个数）；②语义不同的“假设”个数是多少？（同样要说明如何计算得出的，不能只给出个数。提示：覆盖相同例子集合的“假设”是语义相同的假设，语义不同的假设为覆盖例子集合不同的假设）。

(2) 假定“样例空间” $X$  包含的例子个数为  $n$  个（即所有可能的例子个数为  $n$  个），给定的训练例子集合  $D$  中含有  $m$  个例子（即给我们的训练例子个数为  $m$  个）。如果给定一个无偏的



假设空间（即示例的幂集），问：这个无偏的假设空间中与给定的训练例子一致的假设个数有多少个？并说明为什么（是这么多个）？

8. 关于“增强学习”：

(1) 写出“增强学习”中学习 Q 值的算法。（Q 值调节的公式为：

$$\hat{Q}(s, a) \leftarrow r + \gamma \max_a \hat{Q}(s', a') \quad (5 \text{ 分})$$

(2) 证明确定性马尔可夫决策过程中 Q 学习的收敛性。即令  $\hat{Q}_n(s, a)$  代表第 n 次更新后的 Q

值，证明：当  $n \rightarrow \infty$  时， $\hat{Q}_n(s, a)$  收敛到真实的  $Q(s, a)$ 。（5 分）