

题目 1

我想强人工智能在相当长的一段时间内是无法实现的。这是因为目前我们的人工神经网络仍然处在极其简单的水平，我们没有研究清楚人脑中意识、推理能力产生的机制，也没有足够大的计算能力对这样复杂的机制进行模拟。现在的 AI 模型都是仅仅实现某个特定的功能，而不是通用人工智能。因此强人工智能短期内无法实现。

题目 2

我们说“学习意味着变化”，是因为：

- 对于参数化的模型来说，学习意味着通过某种优化方法（例如基于梯度下降的算法）调整模型中的参数，使得损失函数减小。参数在变化就代表着模型正在从数据中学习信息。
- 对于非参数化的模型来说，在模型学习时，模型的结构（表达函数的方式）在变化。例如决策树模型的学习过程就是树模型的生成过程。

题目 3

奥卡姆剃刀原则鼓励模型在尽可能地正确拟合数据的前提下，尽量使用简单的模型。简单的模型事实上限制了模型拟合训练集中能力，这会使得：

- 模型不会关注数据中细枝末节的特征，而是关注数据体现的普遍规律。
 - 模型不容易受到数据中的噪声影响。
- 因此，模型会有更好的泛化能力，不容易过拟合。事实上，数学上的一些概念也反映了奥卡姆剃刀原则的效果，例如在使用多项式模型拟合一个函数（或者使用多项式对某个函数进行插值）时，如果多项式的次数过高，容易发生“龙格现象”——在拟合（或插值）区间的边缘，拟合（插值）的结果与原函数偏离很大，这也可以被理解为一种过拟合。限制多项式的次数可以避免“龙格现象”，也就是防止了过拟合的出现。

题目 4

首先我们要说明 K-Means 算法的目标函数：使得所有数据的离差平方和最小化。也就是说

$$f(x^\dagger) = \sum_i (x_i - x^\dagger)^T (x_i - x^\dagger)$$

其中， x_i 是表达一个点位置的向量； x^\dagger 是 x_i 点所处的一类的聚类中心。那么 K-Means 算法由两步组成：

- 计算每一类的中心位置，并且更新中心点。
- 我们要说明，经过这一步之后，目标函数的值不增，那么我们只需说明一个引理：设 y_1, y_2, \dots, y_n 是欧氏空间中的 n 个向量，则函数

$$g(y) = \sum_i (y_i - y)^T (y_i - y)$$

取到最小值，当且仅当 $y = 1/n \sum_i y_i$ 。这个引理是容易证明的，因为：

$$g(y)' = -2 \sum_i (y_i - y) = 0$$

使得 $g(y) = 0$ ，引理得证。

那么，在进行完第一步之后，各类的中心点从 x^\dagger 变成 x^* ，我们有：

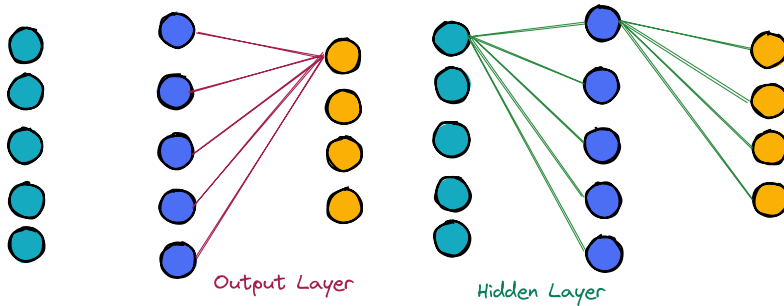
$$\sum_i (x_i - x^*)^T (x_i - x^*) \leq \sum_i (x_i - x^\dagger)^T (x_i - x^\dagger)$$

- 更新每一个样本所属的中心点

由于每一个样本都被重新划分到离它最近的中心点，因此在更新后，目标函数 $\sum_i (x_i - x^*)^T (x_i - x^*)$ 的值必然比更新前更小。

因此，K-Means 算法运行过程中，目标函数值组成单调递减、下界为 0 的序列，显然序列的极限存在。

题目 5



对于输出层：使用 e 表示损失函数， ω_{ji} 表示我们要优化的权重

$$\begin{aligned} \frac{\partial e}{\partial \omega_{ji}} &= \frac{\partial e}{\partial y_i} \frac{\partial y_i}{\partial \omega_{ji}} \\ &= \frac{\partial e}{\partial y_i} \frac{\partial y_i}{\partial \sigma} \frac{\partial \sigma}{\partial \omega_{ji}} \end{aligned}$$

如果使用的损失函数是 MSE Loss，那么：

$$\frac{\partial y_i}{\partial \sigma_i} = 1 - y_i^2 \quad \frac{\partial e_i}{\partial y_i} = y_i - d_i \quad \frac{\partial \sigma_i}{\partial \omega_{ij}} = x_{ji}$$

因此：

$$\frac{\partial e_i}{\partial \omega_{ji}} = (y_i - d_i)(1 - y_i^2)x_{ji}$$

对于隐藏层，我们将要优化的权重记为 ω_{kj} ，那么

$$\frac{\partial e}{\partial \omega_{kj}} = \sum_i \frac{\partial e}{\partial y_i} \frac{\partial y_i}{\partial \sigma} \frac{\partial \sigma_i}{\partial y_j} \frac{\partial y_j}{\partial \sigma} \frac{\partial \sigma}{\partial \omega_{kj}}$$

代入计算结果，得到：

$$\frac{\partial e}{\partial \omega_{kj}} = \sum_i (y_i - d_i)(1 - y_i^2)\omega_{ji}(1 - y_j^2)x_{kj}$$

题目 6

课程中提到的多层感知机的发展历史如下：

- 1957 年，Rosenblatt 提出了单层感知机，这个感知机使用非 0 即 1 的阈值激活函数，只能完成线性可分的任务
- 之后，多层感知机被提出，多层感知机中采用了非线性的激活函数，使之可以处理非线性的分类和拟合问题。更多的层数也增强了模型的拟合能力。
- 在 1986 年，反向传播算法被提出，损失函数关于多层感知机中各个参数的梯度可以被方便地计算。

- 在 1998 年左右，卷积网络被提出，这使得神经网络被用作图像处理
- 在 2006 年，深度信念网络被提出。与一般的多层感知机不同的是，深度信念网络是一个生成模型，此外，它使用逐层训练限制玻尔兹曼机的方式克服了传统神经网络中梯度消失的问题。其训练方式包括自底向上的与训练和自顶向下的微调。
- 1982 年左右，Hopfield 网络被提出，这是一种带有反馈的神经网络。该网络中定义了一种“能量函数”，在神经元的激活状态达到稳定时，能量函数也达到最小值。通过合理地指定网络中的权重，Hopfield 网络可以完成联想记忆、组合优化等任务。
- 后来，循环神经网络被提出，用于解决时间序列预测、机器翻译等序列中有自相关性的问题。LSTM 网络解决了循环神经网络中梯度消失的问题。
- 其他的神经网络包括自组织特征映射网络等。这种网络的目的是将高维的数据映射到低维空间中，并保证数据间的相对位置不变。

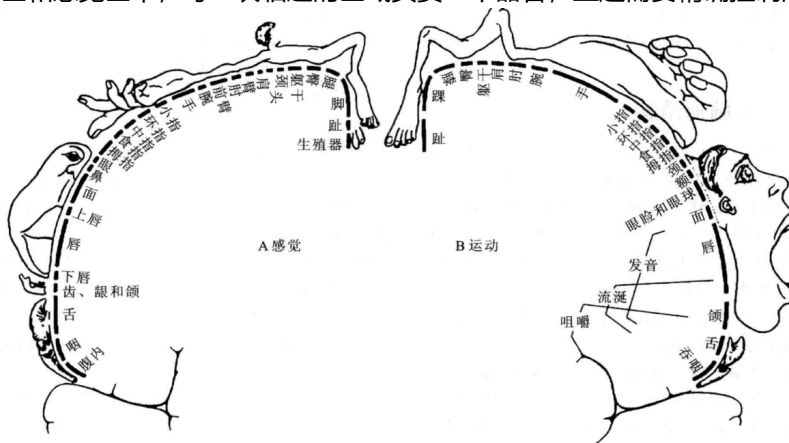
题目 7

使用 Hopfield 网络解决组合优化问题主要包括以下几个步骤：

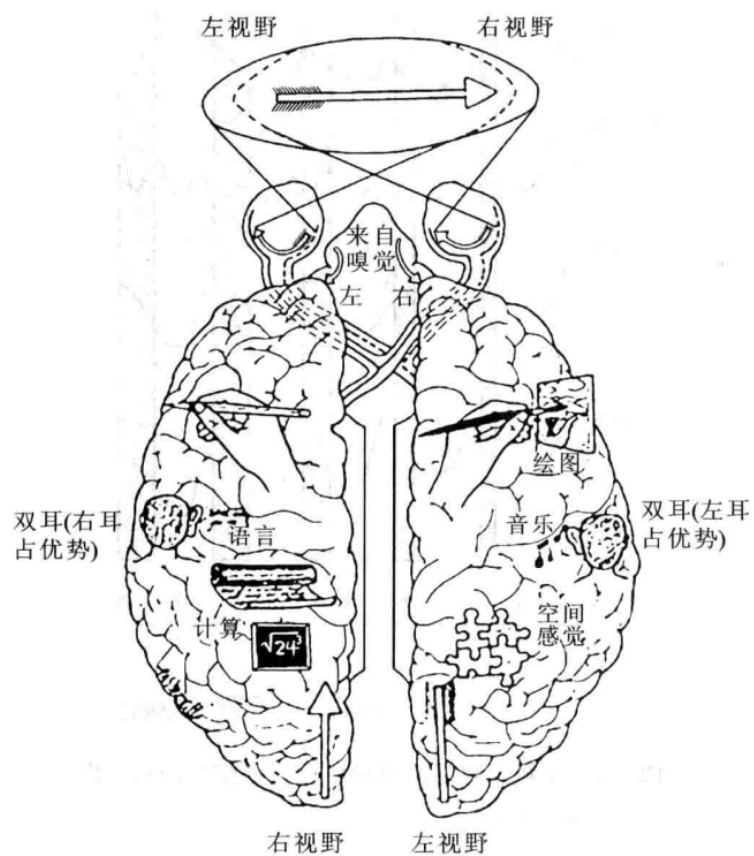
- 使用神经元的激活状态表达决策变量的值。例如在旅行商问题中，我们可以将神经元排列成 $n \times n$ 的矩阵，其中第 i 行第 j 列的神经元代表着第 i 个城市是在第 j 顺位被访问的
- 写出问题的优化目标和约束条件，并使用拉格朗日乘子法原问题转化成无约束优化问题。这是因为 Hopfield 网络中无法直接表达约束条件，因此必须将约束条件放入目标函数中。例如，在旅行商问题中，我们的约束条件包括每座城市必须被访问一次、同一时间只能访问一座城市，等等。
- 将目标函数转换成 Hopfield 能量函数的形式，并且写出网络各边的权重。例如，对于无输入的 Hopfield 网络，目标函数的形式必须是 $E = -\frac{1}{2} S^T w S$
- 生成网络的初始状态，之后网络的状态将随机地变化，不断接受能量更小的状态，直到能量不再变化为止。

题目 8

自组织特征映射网络的生物学证据可能是：人脑中，实现某一功能/控制某一器官的区域是临近的。例如，在人脑运动区和感觉区中，每一块临近的区域负责一个器官，且越需要精确控制的部分，占据的脑区面积越大。



另一个典型的例子是，两个大脑半球的功能是特化的。左半球几乎专门执行运算和复杂推理，右半球则执行一些与空间想象和艺术思维相关的任务。



注：图片引自陈阅增《普通生物学》第四版。