

# 《人工智能导论》2024

## 智能科学与技术学院

### 第三次作业

2024 年 11 月 22 日

1. 在介绍支持向量机(SVM)的时候，我们提到，SVM 的基本思想是当存在多个线性分类超平面时，找到一个最大化分类间隔 (margin)的超平面，请尝试回答，如果存在多个具有相同 margin 的分类超平面，应当如何处理？尝试基于你的回答设计一个新的学习算法。
2. 核函数是统计机器学习时代处理线性不可分问题的一个重要技巧，一个函数能够作为核函数需要满足对称性、正定性。而神经网络具有万有逼近性，当我们面临一个新任务时，是否可以结合该任务的特性，设计一个神经网络的优化目标，使得通过该目标优化后的神经网络可以近似作为该任务上的核函数。尝试描述你的算法。
3. Sigmoid 函数 $\text{sigmoid}(z) = \frac{1}{1+e^{-z}}$ 可以将 $z$ 转换到 0 到 1 的区间内。Softmax 是另一种常用的激活函数，其定义为： $\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^n e^{z_j}}$ ，它可以将输入向量 $z = [z_1, z_2, \dots, z_n]$ 转换为一个概率分布（所有输出之和为 1）。
  - a) 试证明，对于一个二分类问题（即 $n = 2$ ），Softmax 函数与 Sigmoid 函数等价。

- b) 试推导 Softmax 函数的梯度公式，即对输入 $z_i$ 求偏导。
  - c) 在实际的神经网络训练中，Softmax 函数通常与交叉熵损失 (cross-entropy loss) 一起使用，交叉熵损失的表达式为： $L = -\sum_{i=1}^n y_i \log \hat{y}_i$ ，其中 $y_i$ 为真实标签的 one-hot 编码， $\hat{y}_i$ 为 Softmax 的预测概率分布，请计算结合 Softmax 与交叉熵损失的梯度。
4. 对于 KNN 算法，距离度量函数的选择对性能有很大的影响，是否可以设计一个学习算法，能够基于目标任务的训练数据自动学习得到一个距离度量函数。
- a) 为了完成该目标，请首先设计一个带有可学习参数的距离度量。（提示：可以令该可学习参数为一个矩阵 $W$ ）
  - b) 当有了一个带有可学习参数的距离度量 $W$ 之后，请继续设计一个损失函数。（提示 1：可以结合第三题中的 softmax 函数与交叉熵损失；提示 2：softmax 函数计算的是样本属于每个类别的概率，可以尝试思考在 KNN 算法中一个样本属于每个类别的概率是如何计算的？这两者如何结合？）
  - c) 接下来请设计优化算法，使其能够基于损失函数的值对 $W$ 进行更新。