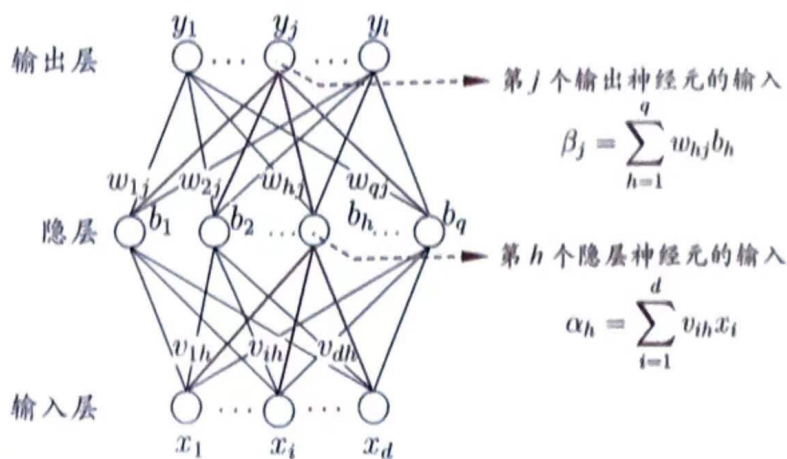


1. 对于训练数据集，其正实例点是 $x_1 = (3,3)^T$, $x_2 = (4,3)^T$ ，负实例点是 $x_3 = (1,1)^T$, $x_4 = (2,1)^T$ 。请利用随机梯度下降法求感知机模型 $f(x) = \text{sign}(\omega \cdot x + b)$ 。要求清晰地描述数据、模型、策略、算法四个部分，对于算法求解过程给出清楚的描述，其中参数初始化设置如下： $\omega_1 = 1$, $\omega_2 = 0$, $b = -1.5$, $\eta = 0.5$ 。

2. 给定含有6个样本的集合 $X = \{(3,0), (1,0), (1,1), (6,0), (6,2), (5,1)\}$ ，请用k均值聚类算法将样本聚到2个类中。

3. 给定训练集数据集 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ ，其中 $x_i \in \mathbb{R}^d$, $y_i \in \mathbb{R}^l$ ，即输入示例由 d 个属性描述，输出 l 维实值向量。下图给出了一个拥有 d 个输入神经元、 l 个输出神经元、 q 个隐层神经元的多层前馈网络结构，其中输出层第 j 个神经元的阈值用 θ_j 表示，隐层第 h 个神经元的阈值用 γ_h 表示，输出层和隐层的激活函数都是Sigmoid函数，其他变量如图所示。在均方损失下，给定学习率 η ，请推导出误差逆传播算法对权值和阈值的迭代更新公式。



1. 列举机器学习的一般分类，并分别简述其概念；解释什么是模型的泛化能力。
2. 请列举4个常见的损失函数并写出其表达式。什么是经验损失和期望损失，它们的区别是什么？正则化项在损失函数中有什么作用？
3. 在评估模型的过程中，训练误差和测试误差分别反应了什么信息？在实际情况中经常面临数据集不足的情况，为了更好的评估模型，可以采样哪几种划分数据集的方法？请分别简述其概念。

1. 对于线性可分数据集 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ ，其中 $x_i = (x_{i1}; x_{i2}; \dots; x_{id})$ ， $y_i \in \{-1, 1\}$ ，请用支持向量机和硬间隔最大化，推导给出划分超平面参数求解过程中的原始问题、拉格朗日函数、对偶问题以及KKT条件。（提示：请说明所使用的公式符号，并对推导过程给出必要的解释和说明）

2. 对于数据集 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ ，其中 $x_i = (x_{i1}; x_{i2}; \dots; x_{id})$ ， $y_i \in \{-1, 1\}$ ，给定基学习算法 \mathcal{L} ，请用AdaBoost算法训练一个 T 个学习器的集成，要求给出算法学习过程的关键步骤（或者伪代码），并清楚写出各个基分类器的权重计算、训练数据的权值分布更新的表达式。（提示：请说明所使用的公式符号，并对推导过程给出必要的解释和说明）

表 5.1 贷款申请样本数据表

| ID | 年龄 | 有工作 | 有自己的房子 | 信贷情况 | 类别 |
|----|----|-----|--------|------|----|
| 1 | 青年 | 否 | 否 | 一般 | 否 |
| 2 | 青年 | 否 | 否 | 好 | 否 |
| 3 | 青年 | 是 | 否 | 好 | 是 |
| 4 | 青年 | 是 | 是 | 一般 | 是 |
| 5 | 青年 | 否 | 否 | 一般 | 否 |
| 6 | 中年 | 否 | 否 | 一般 | 否 |
| 7 | 中年 | 否 | 否 | 好 | 否 |
| 8 | 中年 | 是 | 是 | 好 | 是 |
| 9 | 中年 | 否 | 是 | 非常好 | 是 |
| 10 | 中年 | 否 | 是 | 非常好 | 是 |
| 11 | 老年 | 否 | 是 | 非常好 | 是 |
| 12 | 老年 | 否 | 是 | 好 | 是 |
| 13 | 老年 | 是 | 否 | 好 | 是 |
| 14 | 老年 | 是 | 否 | 非常好 | 是 |
| 15 | 老年 | 否 | 否 | 一般 | 否 |

使用 ID3, C4.5, CART 其中一种算法构建决策树, 写出计算过程, 画出决策树

Part2