

2020—2021 学年第一学期期末试卷

学号_____ 姓名_____ 成绩_____

考试日期： 2021 年 01 月 08 日，晚上 19:00 - 21:00

考试科目：《 机器学习 》（A 卷）

注意事项：1、请大家仔细审题

2、不能违反考场纪律

一. 2077 年，人类发射宇宙飞船，观测了 800 颗地外行星，根据其大小和轨道的远近、以及是否适宜人类居住，列出下面的表格：

表 1

计数	尺寸	轨道	是否宜居？
20	大	近	是（+）
170	大	远	是（+）
139	小	近	是（+）
45	小	远	是（+）
130	大	近	否（-）
30	大	远	否（-）
11	小	近	否（-）
255	小	远	否（-）

请使用 ID3 算法根据以上数据，建立一个决策树，使得人们能够根据其尺寸大小和轨道远近，尽可能准确地预测出一个星球是否宜居。要求详细地写出每个计算步骤，画出决策树的各层，并指出每个节点的正类和负类的数量以及叶子节点所代表的类别。（10 分）

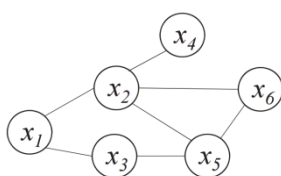
二. 简述支持向量机（SVM）的基本思想，写出模型表达式。说明支持向量机如何处理噪声和离群点以及如何处理非线性可分的数据。（10 分）

三. 简述集成学习（Ensemble Learning）的概念，并以串行化方法 Boosting 和并行化方法 Bagging 为例简单解释其原理。（10 分）

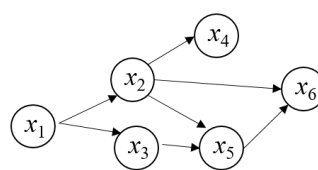
四. 说明 K 均值算法和混合高斯模型的异同点。简述 EM 算法的步骤。(10 分)

五. (1) 列出图 1 (a)中所有的极大团;(4 分)

(2) 令 $X=\{x_1, x_2, \dots, x_6\}$, 分别写出图 1 (a)和(b)中变量的联合概率分布(用 $\psi_Q(X_Q)$ 表示定义在团 X_Q 上的势函数)。(6 分)



(a)马尔可夫随机场



(b)贝叶斯网络

图 1

六. 假设损失函数为 E , 根据残差模块图 2 (a)进行残差神经网络图 2 (b)的推导, 求 $\frac{\partial E}{\partial x^l}$ 。(10 分)

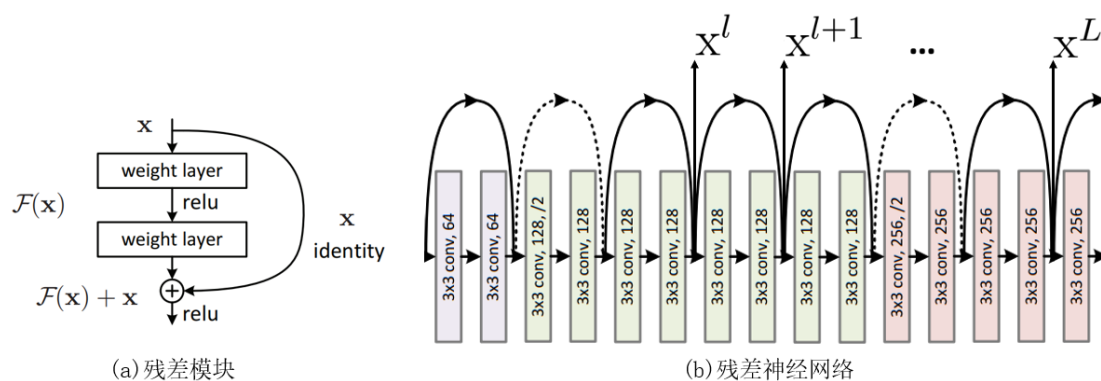


图 2

七. 将 D 维数据集 $\{x_n\}, n=1,2,\dots,N$ 降为 M 维 ($M < D$), 请从最大方差思想角度推导主成分分析 (Principal Component Analysis) 的计算过程。(10 分)

八. 通过含有一层隐藏层的神经网络推导 BP 反传算法。(10 分)

其中:

给定训练集 $D = \{(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), \dots, (\mathbf{x}_N, \mathbf{y}_N)\}$, $\mathbf{x}_i \in \mathbb{R}^4, \mathbf{y}_i \in \mathbb{R}^3$ 。

激活函数为 Sigmoid 函数 $f(x) = \frac{1}{1+e^{-x}}$ 。

准则函数为 $E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y(\mathbf{x}_n, \mathbf{w}) - y_n\}^2$, $y(\mathbf{x}_n, \mathbf{w})$ 表示第 n 个样本 \mathbf{x}_n 对应的神经网络输出向量。

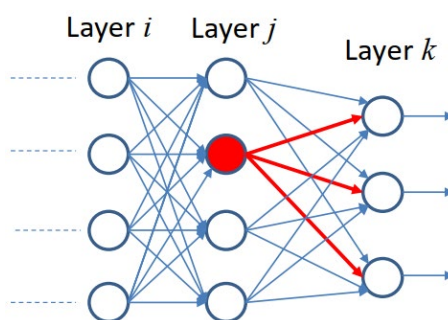


图 3

九. 根据本课程的知识以及你对深度学习的认识, 简述深度学习算法目前存在的问题, 并对深度学习下一步的发展方向给出自己的观点。(20 分)