《机器学习导论》期末考试

命题人：易凯
人工智能与机器人研究所
西安交通大学社会心理学研究所
西安交通大学软件学院
williamyi96@gmail.com

1 选择题 (3×8)

1. 机器学习按照一定的规则可以进行分类，以下不属于该类别的一项是：

A) 监督学习

B) 弱监督学习

C) 无监督学习

D) 强化学习

答案：B

2. 以下不属于监督学习方法的是：

A) K-means 算法

B) SVM 算法

C) 决策树算法

D) 逻辑回归算法

答案：A

3. 以下关于欠拟合以及过拟合的说法不正确的是：

A) 已标注的训练数据样本过少会导致欠拟合

B) 正则化是有效缓解数据过拟合的一种有效机制

C) 模型在训练集上对数据点的拟合越好，则在测试集上表现的性能会越好

D) 无标注的训练数据也可以有效解决模型欠拟合问题

该论文作者是西安交通大学 15 级本科生，学号为 2151601053。目前在西安交通大学人工智能与机器人研究所从事基于脑认知系统构建的人工智能的研究，指导老师为郑南宁老师；在西安交通大学社会心理学研究所从事计算心理学相关领域的研究，指导老师为赵靓老师与喻丰老师。整体研究方向为认知科学，机器视觉，人工智能，混沌分形，变分推理，计算心理学。

答案：C

4. 以下关于信息熵的说法错误的是：

A) 信息的信息量等于不确定性的多少

B) 信息熵可以作为数据纯净度或者混乱度的衡量指标

C) 信息熵具有连续性和对称性，但是其不具有极值性

D) 互信息可以作为变量间相互关系的度量标准

答案：C

5. 关于 C4.5 对 ID3 的改进不包括：

A) 用信息增益率代替信息增益来选择属性

B) 能够完成对连续值属性的离散化处理

C) 能够减轻训练数据的不均匀概率分布导致的决策树失衡问题

D) 其在决策树构建完成之后会对决策树进行剪枝

答案：C

6. 以下关于损失函数 (loss function) 的设计说法错误的是：

A) 损失函数是有监督学习中调控模型训练过程的重要指标

B) 损失函数的设计将直接影响到模型是否能够收敛

C) 损失函数的正则化项是一种有效的防止训练过拟合的手段

D) 当前，仅有均方差损失函数是可取的，也因此其被广泛应用

答案：D

7. 以下关于 SVM (支持向量机) 的说法，错误的是：

A) 训练好的模型的算法复杂度是由支持向量的个数决定的，而不是由数据的维度决定的。所以 SVM 不太容易产生 Overfitting

B) SVM 训练出来的模型完全依赖于支持向量，即使训练集里面所有的非支持向量的点都被去除，重复训练过程，结果仍然会得到完全一样的模型

C) 一个 SVM 如果训练得出的支持向量个数比较少，SVM 训练出的模型比较容易进行泛华

D) 训练得当的 SVM 可以被直接应用于求解高维空间的非线性可分问题

答案：D

8. 关于深度神经网络的以下说法错误的是：

A) 深度神经网络从本质上而言是一种复杂的非线性分类器

B) 神经网络的卷积核大小一般公认为 1×1 与 3×3 的结合

C) 随机梯度下降是常用的进行模型正向推理的算法

D) 神经网络的解卷积可以增强网络的表征能力

答案：C

2 判断题 (2×15)

1. GINI 指标被用于度量数据划分或者数据集的不纯度，CART 算法选用 GINI 指标来选择分裂属性 (T)

2. KNN 算法是一种基于距离尺度的无监督方法 (T)

3. 长方形公路上不同型号的车辆聚类，KNN 是其中一种行之有效的方法 (F)

4. 逻辑回归问题通过设定边界转换值可以转变为离散状态回归问题 (T)

5. Apriori 算法在基于循环产生候选集时，没有排除不应该参与组合的元素 (T)

6. FP-Growth 算法需要对数据库进行三次扫描，第一次扫描构建 FP 树，第二次扫描删除无用数据，第三次扫描从精简 FP 树中挖掘频繁项集 (F)

7. 在其他条件不变的情况下，使用高斯核的线性回归，如果降低核的大小，那么训练误差则会随之增大 (F)

8. 回归问题和分类问题本质区别在于回归问题变量是连续的，而分类问题变量是离散的 (T)

9. 机器学习任务处理的各个阶段中，最为重要的是特征提取 (T)

10. SVM 本质上是一种高维空间的非线性分类器 (F)

11. 在核回归中，最影响回归的过拟合性和欠拟合之间平衡的参数是核函数的宽度 (T)

12. 梯度下降有时会陷入局部极小值，但是 EM 算法不会 (F)

13. 神经网络中的池化过程就是特征提炼的过程 (T)

14. 神经网络中 $1*1$ 卷积会通过共享权值的方式减小网络计算量，但是其并不能改变 feature map 的深度 (F)

15. AlexNet 中将网络划归为部分等价的两个部分使用到了分而治之的思想 (F)

3 填空题 (1×10)

1. 一般认为随机误差由 $\underline{(1)}$, $\underline{(2)}$, $\underline{(3)}$ 三个部分构成。(偏差、测量误差、随机噪声)

2. 神经网络中的池化有 $\underline{(4)}$, $\underline{(5)}$, $\underline{(6)}$。(最大池化，平均池化，加权池化，部分池化 – 选取三个)

3. 在 MNIST 数据集上使用卷积神经网络提高泛化能力的方法有 (7), (8), (9), (10)。(正则化，更优权值初始化策略，更好的损失函数，增加网络层数，多数据集训练 – 选取四个)

4 解答题 (5×3)

4.1 交叉验证与 K-折交叉验证

1. 简述交叉验证 (CV) 与 K-折交叉验证 (K-CV) 的含义及其异同

解答：交叉验证 (Cross Validation) 是用来验证分类器的性能的一种统计分析方法，基本思想是将所有的数据样本分为训练集以及验证集这两个部分。使用训练集对模型进行训练，使用验证集对模型的性能以及泛化能力进行测试。

K-折交叉验证是将初始数据均分为 K 组，然后任意一次将其中的一组设置为验证集，其他 K-1 作为测试集，将分类准确率的平均数作为最后的衡量指标。这种方法得到的准确率较为准确，但是其往往针对大量数据样本有着更大的计算量。

4.2 聚类算法对数据的拓扑要求

2. KNN（最紧邻算法）是一种典型的聚类方法，相关的聚类方法对数据样本的拓扑结构有着一定的限制。试分析，KNN 算法是否可以用于长方形街道上车辆的聚类? 如果是请说明理由，如果否则在说明理由的基础之上找出适合的聚类方法，并分析。

解答：KNN 不能应用于长方形街道上车辆的聚类检测。因为 KNN 要求聚类的数据满足椭球形分布，而长方形街道显然不满足条件。而我们可以使用 Chameleon 聚类等聚类方法。

4.3 强化学习

3. 请说明什么是强化学习 (包括机制，马尔科夫过程，reward 函数设计)

解答：强化学习属于机器学习中极为重要的一个分支，其主要是 agent 与 environment 之间的交互，其中 agent 能够对部分观测到的 environment 的状态做出 action，environment 将对这个 action 针对当前的状态做出判断，分析去 reward。通过这样的一种交互实现 agent 的主动式自学习。马尔科夫过程指的是虽然当前状态 S_t 与之前的 $S_,...,S_{t-1}$ 状态均有关，但是建模仅仅认为其与 S_{t-1} 状态有关。强化学习的 reward function 的设计是基于 Bellman 方程的递归序列。$Q*(s,a) = E_{s'}[r + \lambda \max_{a'} Q*(s',a')|s,a]$。

5 材料题 (21×1)

5.1 贝叶斯变分推理

1. 通过之前的学习我们已经掌握了朴素贝叶斯算法，现材料将给定贝叶斯变分推理相关的材料 [1]，请基于该材料完成基于高斯分布的贝叶斯变分推理模型推导。

假设我们有一个纯粹的贝叶斯模型，其中每个参数都有　个先验概率分布。这个模型也可以有潜在变量以及参数，我们会把所有潜在变量和参数组成的集合记作 Z。类似地，我们会把所有观测变量的集合记作 X。例如，我们可能有 N 个独　同分布的数据，其中 $X = x_1, ..., x_n$ 且 $Z = z_1, ..., z_n$。我们的概率模型确定了联合概率分布 p(X; Z)，我们的　标是找到对后验概率分布 p(Z j X) 以及模型证据 p(X) 的近似。与我们关于 EM 的讨论相同，我们可以将对数边缘概率分解，即

$$\ln p(X) = \zeta(q) + KL(q||p) \tag{1}$$

其中我们定义了:

$$\zeta(q) = \int q(Z) \ln \left\{ \frac{p(X, Z)}{q(Z)} \right\} dz \tag{2}$$

$$KL(q||p) = -\int q(Z) \ln \left\{ \frac{p(Z|X)}{q(Z)} \right\} dz \tag{3}$$

通过关于概率分布 q(Z) 的最优化来使下界 $\zeta(q)$ 达到最大值，这等于最小化 KL 散度。

使用分解概率分布，我们可以得到:

$$\begin{aligned} \zeta(q) &= \int \prod_i q_i \left\{ \ln p(X, Z) - \sum_i \ln q_i \right\} dZ \\ &= \int q_i \left\{ \int \ln p(X, Z) \prod_{i \neq j} q_i \, dZ_i \right\} dZ_j - \int q_j \ln q_j dZ_j + C \\ &= \int q_j \ln \tilde{p}(X, Z) dZ_j - \int q_j \ln q_j dZ_j + C \end{aligned} \tag{4}$$

其中：

$$\ln \tilde{p}(X, Z_j) = \mathbb{E}_{i \neq j}[\ln p(X, Z)] + C \tag{5}$$

则我们得到最优解 $q_j^*(Z_j)$ 的一般表达式形式为：

$$\ln q_j^* = \mathbb{E}_{i \neq j}[\ln p(X, Z)] + C, \quad \mathbb{E}_{i \neq j}[\ln p(X, Z)] = \int \ln p(X, Z) \prod_{i \neq j} q_i dZ_i \tag{6}$$

因此，我们可以根据迭代学习的方式对 $q_j(Z_j)$ 进行优化直到最优解，通过最优解对之后的对应输入时间进行预测。

请根据以上内容完成一元高斯分布的推导。

解答：如假设使用一元高斯分布，则有似然函数为：

$$p(\mathcal{D}|\mu,\tau) = (\frac{\tau}{2\pi})^{\frac{N}{2}} \exp\left\{-\frac{\tau}{2}\sum_{n=1}^{N}(x_n-\mu)^2\right\} \tag{7}$$

其中，引入 μ 和 τ 的共轭先验分布，形式为：

$$p(\mu|\tau) = \mathcal{N}(\mu|\mu_0,(\lambda_0\tau)^{-1}), \qquad p(\tau) = Gam(\tau|a_0,b_0) \tag{8}$$

通过化归得到：

$$\begin{aligned} \ln q_\mu^*(\mu) &= \mathbb{E}_\tau[\ln p(\mathcal{D}|\mu,\tau) + \ln p(\mu|\tau)] + C \\ &= -\frac{\mathbb{E}[\tau]}{2}\left\{\lambda_0(\mu-\mu_0)^2 + \sum_{n=1}^{N}(x_n-\mu)^2\right\} + C \end{aligned} \tag{9}$$

参考文献

[1] David M. Blei, Alp Kucukelbir, and Jon D. Mcauliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 2016.