

中国人民大学考试试卷

我郑重承诺：

在本次考试中，遵守考试纪律、自觉自爱、平等竞争、维护学校的荣誉和学生的尊严。

签字：

2022-2023 学年第 2 学期《机器学习基础》期末试题（A 卷）

2022.6.14（共 2 页）

题号	一	二	三	四	五	总分
题分	20	20	24	24	12	100
得分						

专业_____ 班级_____ 姓名_____ 学号_____

一、判断题（共 4 小题，每小题 5 分，共 20 分）

答题须知：本题答案必须写在如下表格中，否则不给分。

小题	1	2	3	4
答案				

- 对于 $\mathbf{A} \in \mathbb{R}^{D \times L}$ 和 $\mathbf{B} \in \mathbb{R}^{L \times D}$ ， $\text{trace}(\mathbf{AB}) = \text{trace}(\mathbf{BA})$ 吗？（“是”或“否”）_____。
- 对于一个酉矩阵 $\mathbf{U} \in \mathbb{R}^{D \times D}$ ， $\forall \mathbf{X} \in \mathbb{R}^{D \times L}$ ， $\|\mathbf{UX}\|_F = \|\mathbf{X}\|_F$ 恒成立吗？（“是”或“否”）_____。
- 对于任意两个概率分布 p 和 q ，KL 散度 $KL(p||q) = KL(q||p)$ 吗？（“是”或“否”）_____。
- 两个随机变量相关则一定不相互独立，不相关则一定相互独立，对吗？（“是”或“否”）_____。

二、不定项选择题（共 4 小题，每小题 5 分，共 20 分）

答题须知：本题答案必须写在如下表格中，否则不给分。

小题	1	2	3	4
答案				

- 逻辑回归（Logistic Regression）是……………（ ）
(A) 分类模型 (B) 回归模型 (C) 判别型模型 (D) 生成型模型
- 给定 $f(x) = \frac{1}{1+\exp(-x)}$ ，下列哪些函数表示 $\frac{df}{dx}$ ？……………（ ）
(A) $\frac{\exp(-x)}{(1+\exp(-x))^2}$ (B) $f(x)(1-f(x))$ (C) $\frac{1}{(1+\exp(-x))(1+\exp(x))}$ (D) $\frac{1}{1-\exp(-x)}$
- 下列关于 Nadaraya-Watson（NW）估计器的说法哪些是正确的？……………（ ）
(A) NW 估计器是参数化模型 (B) NW 估计器使用的核函数必须是非负的
(C) NW 估计器使用的核函数必须是平滑的 (D) NW 估计器是非参数化模型

- 下列关于主成分分析的说法哪些是正确的？……………（ ）

(A) 主成分向量的个数可以大于数据矩阵的秩。
(B) 不同的主成分向量一定是相互正交的。
(C) 主成分向量可以通过对数据的协方差矩阵做特征值分解进行求解。
(D) 对应服从多元高斯分布的数据，可以使用主成分分析实现数据白化。

三、算法设计题（共 2 小题，每小题 12 分，共 24 分）

- 给定一个均匀分布 $\text{Uniform}([0, 1])$ 的样本生成器。（1）设计一个伯努利分布 $\text{Bernoulli}(0.5)$ 的样本生成器。（2）设计一个指数分布 $p(X) = \exp(-X)$ ， $X \geq 0$ 的样本生成器。（3）基于上述两个样本生成器，设计一个 D 维 Laplace 分布 $\text{Laplace}(\mathbf{b}, \sigma \mathbf{I}_D)$ 的样本生成器。其中 $\mathbf{b} \in \mathbb{R}^D$ 和 $\sigma > 0$ 为给定的参数。（知识点 Lab 1、Lecture 10）

- 假设观测到的数据 $\{\mathbf{x}_n \in \mathbb{R}^D, y_n \in \mathbb{R}\}_{n=1}^N$ 是由一个线性模型产生的，且这个线性模型的参数的先验分布是一个混合分布：

$$\text{Model: } y = \mathbf{x}^T \mathbf{w} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

$$\text{Prior: } p(\mathbf{w}) = c \underbrace{\frac{1}{2^D} \exp(-\|\mathbf{w}\|_1)}_{\text{Laplace}(\mathbf{0}, \mathbf{I}_D)} + (1-c) \underbrace{\frac{1}{(2\pi)^{D/2}} \exp\left(-\frac{1}{2}\|\mathbf{w}\|_2^2\right)}_{\mathcal{N}(\mathbf{0}, \mathbf{I}_D)} \quad (1)$$

其中控制混合比例的参数 c 是已知的。请从 Bayesian 统计的角度，设计一个后验概率最大化（MAP）算法对模型参数进行学习。要求写出（1）目标函数的形式；（2）迭代算法中的关键步骤。（知识点：Lecture 4、9）

四、证明题（共 2 小题，每小题 12 分，共 24 分）

- 流形学习的常见形式是基于数据矩阵 $\mathbf{X} \in \mathbb{R}^{N \times D}$ 构造一个半正定矩阵 $\Phi \in \mathbb{R}^{N \times N}$ ，再求解如下的优化问题：

$$\min_{\mathbf{Z} \in \mathbb{R}^{N \times L}} \text{trace}(\mathbf{Z}^T \Phi \mathbf{Z}), \quad s.t. \mathbf{Z}^T \mathbf{Z} = \mathbf{I}_L. \quad (2)$$

其中 $L < D$ 。证明（1）对于局部线性嵌入（LLE）和 Laplace 特征映射（LE），它们构造的矩阵 Φ 满足 $\Phi \mathbf{1}_N = \mathbf{0}_N$ 。（2）证明当 Φ 是正定矩阵的情况下 $\mathbf{Z} = \mathbf{U}_L$ ，其中 $\mathbf{U}_L = [\mathbf{u}_1, \dots, \mathbf{u}_L]$ 表示最小的 L 个特征值对应的特征向量构成的矩阵。（知识点 Lecture 7、8、Lab 6）

- 给定一组数据 $\mathbf{x} = \{\mathbf{x}_n\}_{n=1}^N$ ，Mean-shift 算法采用下述迭代的算法计算均值 $\mathbf{m} = \{\mathbf{m}(x_n)\}_{n=1}^N$ ：
步骤 1：初始化 $\mathbf{m}^{(0)}(x_n) = x_n$ for $n = 1, \dots, N$ 。
步骤 2：For $t = 1, \dots, T$ ：

$$\mathbf{m}^{(t)}(x_n) = \frac{\sum_{i=1}^N K(\mathbf{m}^{(t-1)}(x_i), \mathbf{m}^{(t-1)}(x_n)) \mathbf{m}^{(t-1)}(x_i)}{\sum_{j=1}^N K(\mathbf{m}^{(t-1)}(x_j), \mathbf{m}^{(t-1)}(x_n))}. \quad (3)$$

小明同学在基于 RBF 核函数实现上述 Mean-shift 算法的时候，错误地将“步骤 2”中的核函数固定成了 $\mathbf{K} = [K(x_n, x_i)]$, $i, n = 1, \dots, N$ 。(1) 试证明，这种情况下算法依然收敛，即存在 $\lim_{t \rightarrow \infty} \mathbf{m}^{(t)}$ 。(2) 收敛点是什么? (知识点: Lecture 11、14、Lab 8)

五、简答题（共 1 小题，每小题 12 分，共 12 分）

1. 小明同学收集到一批工作简历数据，其中每一份简历都包括了简历拥有者的“性别、年龄、毕业学校、工作经历”（可选的特征），和该简历拥有者最终“是否被雇佣”（标签）。假设小明同学需要基于这批数据设计并学习一个分类器。请（1）分别论述一种生成型分类模型和一种判别型分类模型的设计方案和学习方案，尽量保证最终得到的模型不会对“性别”和“年龄”具有歧视性。（2）你认为哪种分类模型更难学习、公平性更难实现?