

南开大学软件学院本科生 2019-2020 学年第 2 学期《机器学习导论》课程期末考试试卷（A 卷）

专业： 年级： 学号： 姓名： 成绩：

草 稿 区

得 分

一、名词解释（本题共 15 分，每小题 3 分）

1. SVM
2. LSTM
3. 统计学习理论
4. 聚类(Clustering):
5. PCA

得 分

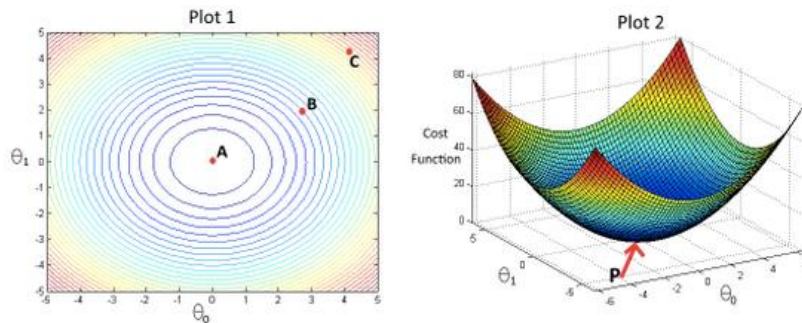
二、是非题（本题共 10 分，每小题 2 分）

1. () Logistic 回归即可以解决回归问题，又可以解决分类问题。
2. () 深度神经网络不能仅靠增加神经网络的层数实现，因不能很好地解决梯度不稳定（消失或爆炸）。
3. () 判决式模型和生成式的区别在于是否需要推断生成训练数据的分布。
4. () 非负矩阵分解的损失函数可以使用误差平方和或者 KL 散度来表示。
5. () 在偏差-方差均衡理论里，过拟合的表现是训练错误较低，但在验证集或测试集上的错误率高，且不稳定。

得分

三、选择题（本题共 20 分，每小题 4 分）

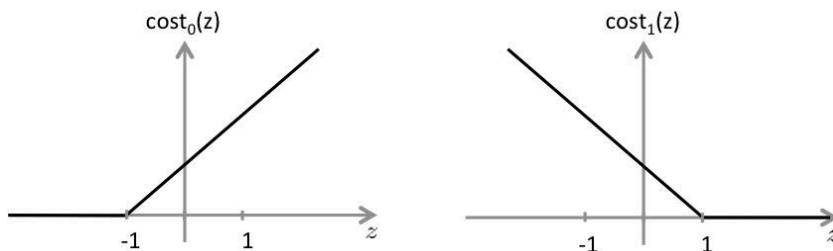
1. () 机器学习指一段计算机程序，它可以从任务 **T** 的经验 **E** 中学习其统计规律，使用性能 **P** 进行衡量并进行优化。假设我们为一个学习算法提供很多历史销售数据，让学习出来的模型能够预测天气。那么任务 **T** 是指什么？
- A) 可视化所有历史经验数据 **E**
 B) 能够正确预测将来某一天天气的概率
 C) 预测明天海潮来的时间
 D) 预测明天是否有日全食
2. () 根据下面的代价函数 $J(\theta_0, \theta_1)$ 的等高线图和三维图像，选择正确的答案

Plots for Cost Function $J(\theta_0, \theta_1)$ 

- A. 如果从点 **B** 开始进行梯度下降，一个好的学习步长将会到达最大值 **A** 附近。
- B. 如果从点 **B** 开始进行梯度下降，一个好的学习步长将一定会到达最小值 **A**。
- C. 如果从点 **B** 开始进行梯度下降，一个好的学习步长将一定会到达最小值 **C**。
- D. 图 2 中的点 **P**（全局最大值）对应图 1 中的点 **C**。

3. () SVM 的损失函数分为训练错误惩罚(经验风险)和模型复杂度惩罚两部分, 其中训练错误惩罚部分

为 $\min_{\theta} C \sum_{i=1}^m [y^{(i)} cost_1(\theta^T x^{(i)}) + (1 - y^{(i)}) cost_0(\theta^T x^{(i)})]$ 。请问下列哪个选项完全正确。

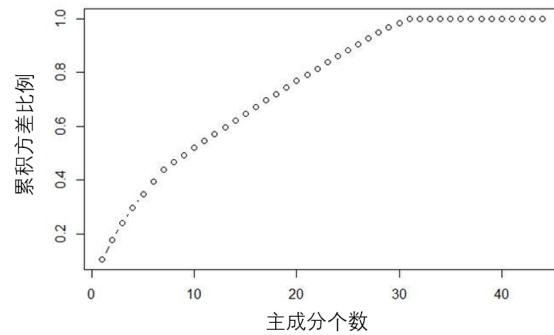


- A. 对于所有 $y^{(i)}=1$ 的样本, 尽可能满足 $\theta^T x^{(i)} \geq 0$; 对于所有 $y^{(i)}=0$ 的样本, 尽可能满足 $\theta^T x^{(i)} \leq 0$;
- B. 对于所有 $y^{(i)}=0$ 的样本, 尽可能满足 $\theta^T x^{(i)} \leq -1$; 对于所有 $y^{(i)}=1$ 的样本, 尽可能满足 $\theta^T x^{(i)} \geq 1$ 。
- C. 该损失函数主要用于线性可分问题, 因为存在将两类样本分隔开的超平面, 因此它主要用于寻找支持向量
- D. 该损失函数主要用于线性不可分问题, 因为该函数容易相对于模型参数进行求导, 所以被用来惩罚落入到间隔中或对方类别中的样本。

4. () 给定贝叶斯公式 $P(c_j|\mathbf{x}) = P(\mathbf{x}|c_j)P(c_j) / P(\mathbf{x})$, 问公式中 $P(\mathbf{x} | c_j)$ 为那一项。

- A. 类先验概率
- B. 后验概率
- C. 类条件概率
- D. 联合概率

5. () 在使用 PCA 分解后, 如果期望尽可能保持最大可分性, 请问最少选择多少个主成分合适。



A. 7 B. 30 C. 35 D. 全部

得分

四、简答题（本题共 15 分，每小题 5 分）

1. u 和 v 为三维向量，其中， $u = \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix}$ ， $v = \begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix}$ 。问 $u^T v$ ， $u \cdot v$ 分别得多少？

2. 给出线性回归的损失函数、SVM 的 Hinge 损失和 AdaBoost 的指数损失。

3. 请简介共轭梯度法和随机梯度下降法。

得分

五、综合题（本题共 40 分，每小题 10 分）

1. 给出神经网络的两种损失函数（基于交叉熵的和基于平方误差的），以及使用 BP 算法中所需的分解到每层节点的误差，以及相应的对于模型参数的梯度，其中，反向传递的误差和神经元连接权重的梯度需要有主要的求解步骤。

2. 给出带重启随机游走算法，给出其损失函数，并从统计学习角度解释损失函数的组成部分。

3. 给出主成分分析的①主要思路，②算法伪代码，③以主成分分析为例讲解维度压缩的优点。

4. 近年来，社交网络被广泛用于各类推荐应用中，以书目推荐为例，使用用户-书籍打分矩阵和用户-用户关注关系两组数据，基于非负矩阵分解和随机游走构建一个融合社交网络的书籍推荐系统（假设每个用户都给别的书籍打过分。）。

	B1	B2	B3	...	Bn
U1	2	5			5
U2	4	1	1		
U3		5	2		1
...					
Um	2	1	5		5

	U1	U2	U3	...	Um
U1		1	1		
U2			1		1
U3	1	1			
...					
Um		1	1		

草 稿 区