机器学习第一阶段测试题

一、选择题

1. 以下带佩亚诺余项的泰勒展开式错误的一项是（D）

A. $e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + o(x^3)$　　B. $arcsin\ x = x + \frac{1}{2} * \frac{x^3}{3} + o(x^3)$

C. $sin\ x = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 + o(x^5)$　　D. $cos\ x = 1 + \frac{1}{2!}x^2 - \frac{1}{4!}x^4 + o(x^4)$

分析：$cos\ x = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 + o(x^4)$

2. 以下关于凸优化的说法错误的一项是（C）

A. 集合 C 任意两点间线段均在集合 C 内，则 C 为凸集

B. 集合 C 的凸包是能够包含 C 的最小凸集

C. 多面体不一定是凸集

D. 线性变换能保持原集合的凸性

分析：多面体是指有限半空间和超平面的交集，多面体一定是凸集

3. 以下说法错误的一项是（C）

A. 当目标函数是凸函数时，梯度下降法的解是全局最优解

B. 进行 PCA 降维时需要计算协方差矩阵

C. 沿负梯度下降的方向一定是最优的方向

D. 利用拉格朗日函数能解带约束的优化问题

分析：沿负梯度方向是函数值下降最快的方向但不一定是最优方向

4. K-means 无法聚以下哪种形状样本？（ ）

A. 圆形分布　　B. 螺旋分布

C. 带状分布　　D. 凸多边形分布

分析：基于距离的聚类算法不能聚非凸形状的样本，因此选 B

5. 若 $X_1,X_2,...X_n$ 独立同分布于(μ,σ^2)，以下说法错误的是（C）

A. 若前 n 个随机变量 $Y_n=\frac{1}{n}\sum_{i=1}^{n}X_i$ 的均值，对于任意整数ε，有：$\lim_{n\to\infty}P\{|Y_n-\mu|<\varepsilon\}=1$

B. 随机变量 $Y_n=\frac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{n}\sigma}$ 的收敛到标准正态分布

C. 随机变量 $Y_n=\sum_{i=1}^{n}X_i$ 收敛到正态分布 $N(\mu,\sigma^2)$

D.样本方差 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})$ 其中样本均值 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$

分析：A：大数定理概念；B、C：中心极限定理概念；C 错，应该收敛到正态分布 $N(n\mu,n\sigma^2)$

D：样本的统计量公式

二、公式推理题

1. 请写出标准正态分布的概率密度函数、期望、以及方差

分析：概率密度函数：$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$；期望：$E(x)=0$；方差：$D(x)=1$

2. 请根据表中的分类结果混淆矩阵给出查准率（准确率）P 和查全率（召回率）R 的计算公式

真实情况	预测结果	
	正例	反例
正例	TP（真正例）	FN（假反例）
反例	FP（假正例）	TN（真反例）

分析：$P=\frac{TP}{TP+FP}$，$R=\frac{TP}{TP+FN}$

三、简答题

1. 求函数 $f(x,y)=x^2+3lny$ 的梯度向量

分析：$\nabla f(x,y)=(\frac{\partial f(x,y)}{\partial x},\frac{\partial f(x,y)}{\partial y})$，所以答案为（2x, 3/y）

2. 列举你知道的无约束最优化方法（至少三个），并选一种方法进行详细介绍

分析：梯度下降法，牛顿法，拟牛顿法，共轭梯度法...（介绍略）

3. 请简要叙述正则化项中的 L1 和 L2 方法

分析：1 正则化和 L2 正则化可以看做是损失函数的惩罚项。L1 正则化是指权值向量 w 中各个元素的绝对值之和，通常表示为||w||1。L2 正则化是指权值向量 w 中各个元素的平方和然后再求平方根（可以看到 Ridge 回归的 L2 正则化项有平方符号），通常表示为||w||2。
L1 正则化可以产生稀疏权值矩阵，即产生一个稀疏模型，可以用于特征选择
L2 正则化可以防止模型过拟合；一定程度上，L1 也可以防止过拟合

4. 简述 k-means 的主要优缺点及针对缺点的优化方案

分析：优点：经典、简单、快速、对密集簇效果较好
缺点：对 K 值敏感，且只适用于能求距离均值的应用，不适合非凸簇或大小差别很大的簇
改进：二分 k-means，k-means++...

5. 简述 UserCF 的主要步骤

分析：一、找到和目标用户兴趣相似的用户集合——计算两个用户的兴趣相似度
二、找到这个集合中的用户喜欢的，且目标用户没有听说过的物品推荐给目标用户——找出物品推荐