

以下关于经典的 **k-means** 聚类的说法哪个是错误的？

正确答案: A

k-means 聚类算法是全局收敛的

k-means 的聚类结果和初始聚类中心点的选取有关

k-means 聚类算法的时间复杂度是 $O(NKT)$ ，其中 N 为样本个数， K 为聚类团数量， T 为迭代次数

k-means 聚类算法无法自动确定聚类团数量

以下两种描述分别对应哪两种对分类算法的评价标准？

① 美团推荐的餐厅有多少是用户喜欢的

② 用户喜欢的餐厅有多少被美团推荐了

正确答案: A

Precision, Recall

Recall, Precision

Precision, AUC

Recall, AUC

假定一个二维数组的定义语句为“`int a[3][4]={{3,9},{2,8,5}};`”，则元素 `a[1][2]` 的值为

正确答案: A

5

4

2

8

快速排序的最坏时间复杂度是？

正确答案: D

A $O(\lg N)$

$O(N)$

C $O(N \lg N)$

D $O(N^2)$

HMM 的无监督训练方法是？

正确答案: A

EM 算法

viterbi 算法

前向-后向算法

梯度下降算法

下面那个不是序列标注模型？

正确答案: B

HMM

ME (最大熵)

MEMM

CRF

下列说法正确的是?

正确答案: C

回归函数 A 和 B, 如果 A 比 B 更简单, 则 A 几乎一定会比 B 在测试集上表现更好
梯度下降有时会陷于局部极小值, 但 EM 算法不会
在 AdaBoost 算法中, 所有被错分的样本的权重更新比例相同
当训练数据较少时更难发生过拟合

逻辑回归的损失函数是?

正确答案: A

log 对数损失函数

平方损失函数

指数损失函数

Hinge 损失函数

0-1 损失函数

1. 如下表是用户是否使用某产品的调查结果。

UID 性别 地区 学历 收入 用户是否使用调查产品

1 男 北方 博士 低 是

2 女 北方 本科 中 否

3 男 南方 本科 高 否

4 女 北方 研究生 中 是

请计算性别、地区、学历、收入中对用户是否使用调查产品信息增益最大的属性。

($\log_2 3 \approx 0.63$)

正确答案: C

性别

B. 地区

学历

收入

美团点评是大家吃喝玩乐的生活好帮手: 比如大家搜索“7 天连锁酒店”, 会返回一系列相关的结果。如果要识别这句话的商家核心词-“7 天”, 以下哪种方法识别商家核心词准确率最低, 不能直接使用?

正确答案: C

通过分析用户搜索日志，人工制定模板或规则来识别
标注一些数据，抽取特征，用 CRF 训练出模型来识别
使用 word2vec 训练词向量，对词聚类，根据词属于的类别来识别
人工整理商家核心词词表，按是否命中词表里的词来识别

以下哪些主要用于无监督的深度学习网络

正确答案: A B C

Restricted Boltzmann Machines

AutoEncoder

Deep Belief Networks

Recurrent Neural Network

同事去美团上班，自己开车，滴滴拼车，滴滴快车，骑滑板车的概率分别是 0.3,0.2,0.1,0.4，
各种方式迟到的概率分别 1/4,1/3,1/12,0，下列语句中正确的是

正确答案: C D

如果他准点，那么骑滑板车的概率大于等于 0.5

单人出行 (自己开车或者滴滴快车) 准点机会比滴滴拼车要低

如果他迟到，自己开车的概率是 0.5

如果他准点，使用滴滴的概率等于自己开车的概率

设有一个栈，元素一次进栈的顺序是 A,B,C,D,E。下列可能的出栈顺序有？

正确答案: A C D

ABCED

EABCD

BCDAE

EDCBA

机器学习中能够用于指导特征选择的指标

正确答案: A B C D

信息增益

信息增益率

基尼系数

信息熵