

下列选项中，用来求解加权有向图的最短路径的算法是：

正确答案: A 你的答案: 空 (错误)

Ford-Bellman 算法

KM 算法

深度优先检索算法

广度优先检索算法

下列机器学习算法中，属于无监督学习的有：

正确答案: C 你的答案: 空 (错误)

PU-learning

KNN

EM 算法

Logistic Regression

机器学习模型 error 和模型 bias 和 variance 之间的关系：

正确答案: B 你的答案: 空 (错误)

没有关系

$\text{error} = \text{bias}^2 + \text{variance}$

$\text{error} = \text{bias} + \text{variance}^{0.5}$

$\text{error} = \text{bias} + \text{variance}$

在 Logistic Regression 中,如果同时加入 L1 和 L2 范数,会产生什么效果:

正确答案: D 你的答案: 空 (错误)

能解决维度灾难问题

L1 和 L2 范数不能同时使用

可以获得更准确的结果

可以做特征选择,并在一定程度上防止过拟合

高斯混合模型(GMM)的极大似然估计(MLE)

正确答案: A 你的答案: 空 (错误)

只能得到非退化(协方差矩阵正定)的局部最优解

存在非退化的全局最优解

解情况依赖于数据分布

机器下列四种排序中()的空间复杂度最大??

正确答案: A 你的答案: 空 (错误)

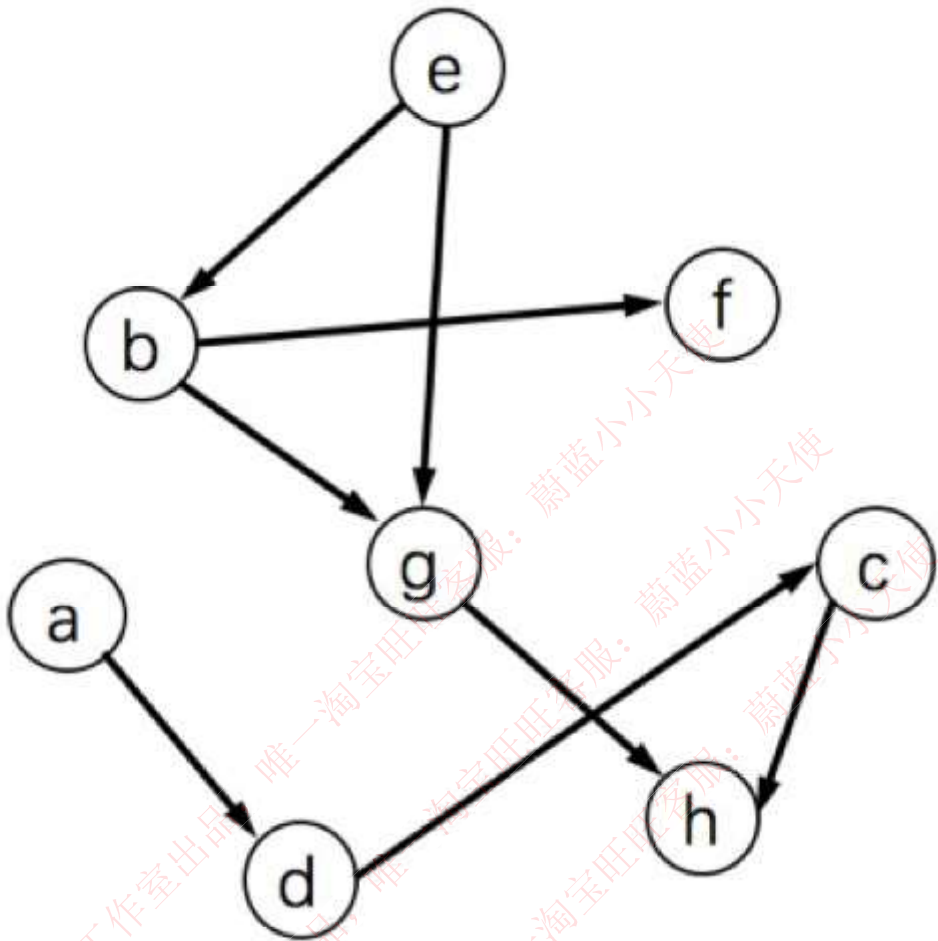
快速排序

冒泡排序

希尔排序

堆排序

下面哪个序列不是下图的一个拓扑排序：



正确答案: B 你的答案: 空 (错误)

- ebfgadch
- adchebfg
- aebdgfch
- aedbfgh

独立随机变量 x 和 y ，概率密度分别为 $p(x)$ 和 $p(y)$ ，那么 $z = x + y$ 的概率密度 $p(z)$ 为：

正确答案: C 你的答案: 空 (错误)

- $P(x) + p(y)$
- $P(x) * p(y)$
- $P(x)$ 和 $p(y)$ 的卷积

给定 $1 \sim n (n > 17)$ 范围内的所有数字，每个人每次从中取出一个数 X ，然后把 X 的约数以及 X 的倍数都从这 n 个数中取走，依次轮流，谁最后没有数可以取，谁就输了。请问以下结论正确的是：

正确答案: A 你的答案: 空 (错误)

先手必赢

先手必输

先手有可能赢, 也有可能输, 受 n 的奇偶性影响

先手有可能赢, 也有可能输, 受 n 是否是质数影响

x, y, z 在 $[0, 1]$ 之间均匀取值且互相独立, 那么 $x^2 + y^2 + z^2 < 1$ 的概率为:

正确答案: D 你的答案: 空 (错误)

$$\frac{\pi}{3}$$

$$\frac{\pi}{4}$$

$$\frac{\pi}{12}$$

以上都不是

在统计模式识分类问题中, 当先验概率未知时, 可以使用?

正确答案: A 你的答案: 空 (错误)

最小最大损失准则

最小误判概率准则

最小损失准则

最大损失准则

$P = \begin{pmatrix} a & b \\ b & c \end{pmatrix}, a > 0, b > 0, c > 0, x \in R^2, \|x\| = 1$, 那么 $x^T P x$ 的最大值是

正确答案: C 你的答案: 空 (错误)

$$\frac{a + c + \sqrt{a^2 + c^2 - 2ac + 4b^2}}{4}$$

$$\frac{a + c - \sqrt{a^2 + c^2 + 2ac + 4b^2}}{2}$$

$$\frac{a + c + \sqrt{a^2 + c^2 - 2ac + 4b^2}}{2}$$

$$\frac{a + c + \sqrt{a^2 + c^2 + 2ac + 4b^2}}{4}$$

A 为 $m \times n$ 的实矩阵, 那么矩阵 AA' 和 $A'A$ 具有

正确答案: A 你的答案: 空 (错误)

相同的非零特征值
相同的特征向量集
相同的核空间

电话号码表 `phone` 中含有 100 万条数据, 其中号码字段 `phone_num` 上创建了唯一索引, 且电话号码全部由数字组成, 要统计号码为 188 的电话号码的数量, 下面写法执行速度最慢的是____

正确答案: C 你的答案: 空 (错误)

```
select count(*) from phone where phone_num >= '188' and phone_num < '188a'
```

```
select count(*) from phone where phone_num like '188%'
```

```
select count(*) from phone where substr(phone_num, 1, 3) = '188'
```

都一样快

一位滴滴实习生开发出了一套简易作弊检测系统, 此系统存在一定误差。如果一个用户确实存在作弊行为, 但是此系统没有检测出的概率为 5%; 正常用户, 但是误检为作弊的概率为 1%; 我们已经知道, 一个用户作弊的概率为 0.1%。目前一个人被此方法检测出有作弊, 那此人确实有作弊的概率接近多少?

正确答案: A 你的答案: 空 (错误)

90%

70%

30%

10%

$$f(x) = \frac{1}{1+e^{-x}}$$

Sigmoid 函数是: 请问对其求导得到什么?

正确答案: D 你的答案: 空 (错误)

$$f(x)(1+f(x))$$

$$\frac{f(x)}{1-f(x)}$$

$$\frac{f(x)}{1+f(x)}$$

$$f(x)(1-f(x))$$

卷积神经网络 (CNN) 中 Dropout 层的作用是:

正确答案: B 你的答案: 空 (错误)

加快收敛速度

防止过拟合

丰富训练样本

增强正样本

为了区分各种不同的应用程序, 传输层使用 () 来进行标识

正确答案: B 你的答案: 空 (错误)

IP 地址

端口号

服务接入点 (SAP)

协议号

矩阵 $\begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 2 \\ 1 & 0 & 3 \end{pmatrix}$ 的秩是

正确答案: B 你的答案: 空 (错误)

1

2

3

其他都不是

红黑树的插入算法复杂度最坏情况是

正确答案: B 你的答案: 空 (错误)

$O(n)$

$O(\log(n))$

$O(n\log(n))$

其他都不对

从 10 个红色球、20 个黄色球、30 个蓝色球中, 每次随机取出一个球 (不放回), 则红色球最先被取完的概率是()

正确答案: A 你的答案: 空 (错误)

7/12

5/6

1/3

4/24/33

总体 X 服从指数分布, 其概率密度函数为:

$$f(x; \theta) = \begin{cases} \theta e^{-\theta x}, & x > 0 \\ 0, & x \leq 0 \end{cases}, \theta \geq 0.$$

则 θ 的最大似然估计为 ()

计为 ()

正确答案: C 你的答案: 空 (错误)

$$\frac{1}{\bar{X}+1}$$

$$\frac{1}{\bar{X}} + 1$$

$$\frac{1}{\bar{X}}$$

\bar{x}

下列说法哪个是正确的?

正确答案: C 你的答案: 空 (错误)

在使用逻辑回归模型的时候, 对样本的拟合程度越高越好。

在使用 SVM 模型的时候, 使用的核方法越复杂越好。

在使用深度学习模型的时候, 可以使用 dropout 层来防止过拟合。

在使用 k means 聚类的时候, k 值的选取可以是任意的。

高斯混合模型(GMM)的极大似然估计(MLE)

正确答案: A 你的答案: 空 (错误)

只能得到非退化(协方差矩阵正定)的局部最优解

存在非退化的全局最优解

解情况依赖于数据分布

主成分分析 PCA 和数据矩阵的奇异值分解 SVD

正确答案: B 你的答案: 空 (错误)

完全等价, 可以得到相同的特征空间

数据均值归零后, SVD 与 PCA 等价, 可以得到相同的特征空间

无法得到不等价形式

在一次模型预测实验里, 测试集合的 Label 是[1,1,0,0], 某个模型的输出值是[0.7,0.9,0.8,0.1], 那么这个模型在该测试集合上的 AUC 是:

正确答案: C 你的答案: 空 (错误)

0.5

0.6

0.75

0.8

位势函数法的积累势函数 $K(x)$ 的作用相当于 Bayes 判决中的?

正确答案: A 你的答案: 空 (错误)

后验概率

先验概率

类概率密度

似然估计

下列哪些属于常见的组合模型:

正确答案: A B D 你的答案: 空 (错误)

bagging

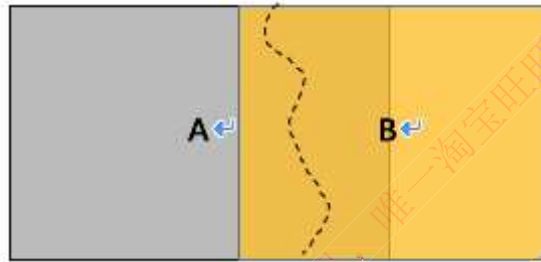
boosting
LR
RandomForest

在模型训练过程中，下列哪些方法可以防止模型过拟合（overfitting）：

正确答案: A B C D 你的答案: 空 (错误)

增大数据量
减少 feature 个数
正则化
交叉验证

假设有两个高度相同的矩阵 **A** 和 **B**（如下图所示），其中填满了非负实数值。将两个矩阵如下重叠放置，假设重叠区域的宽度是固定且已知的，请设计一个算法在重叠区域寻找一条“缝”（如下图虚线所示），使得这条“缝”经过的两个矩阵中的对应位置的元素差值的绝对值之和最小，该“缝”必须 4 邻域连续。请先使用数学语言定义该问题，然后设计算法并描述其思路，必要时可以使用伪代码。



参考答案

该题目是由经典的 image composition 问题变化而来，可以使用动态规划算法来解决，具体可以参考 paper: Efros and Freeman, "Image Quilting for Texture Synthesis and Transfer", SIGGRAPH 2001。主要考查同学的数学建模和算法基础，其实不需要计算机视觉或图形学的相关背景。数学语言描述该问题：求一条路径 L ，使得 $\sum_{(x \in L)} |A(x) - B(x)|$ 的值最小。算法是，首先计算重叠区域两个矩阵相应元素的差值绝对值，即求重叠区域的 $C = |A - B|$ ，然后转化为在 C 矩阵中用动态规划算法寻找最小花费路径。

对下面的损失函数

$$J(\theta) = \frac{1}{2m} \left[\sum_{i=1}^m (h_{\theta}(x_i) - y_i)^2 + \lambda \sum_{j=1}^n \theta_j^2 \right]$$

- 1) 请描述损失函数第 2 项别名和作用，并再列举一个第 2 项的其他形式。并说明两者的区别
- 2) 请根据题目求解参数 θ 的迭代优化公式，写出求解过程。
- 3) 假如现在采用输入一个样本，更新一次参数的方式，参数更新采用 $\theta = \theta + \alpha * \delta(\theta)$ 的形式。请问你能想到什么样的加速收敛的方式。

参考答案

1) L2 正则化项, 作用是倾向于得到平均的参数解, 避免出现绝对值很大的参数。还有 L1 正则化项, 其可以得到稀疏解。

2) 迭代公式:

$$\frac{\alpha}{m} \sum_{i=1}^m (h_{\theta}(x_i) - y_i)$$

3) SGD, L-BFGS, 拟牛顿

在一个 $n \times m$ 的格子迷宫内, 有部分格子可以通过, 有部分格子是禁止通行的, 允许上下左右 4 个方向移动。假定小强在迷宫中的坐标 (i, j) 的位置上, 从迷宫的任意边界走出即为走出迷宫。请给出小强以最少步骤走出迷宫的算法。

参考答案

1、暴力不优方案: 穷尽遍历

2、广度遍历

选择小强周边可通行的格子做一个第一层遍历节点, 并标记为已遍历。

继续选择第一层节点周边未遍历节点作为下一层遍历节点, 并标记为已遍历;

依次遍历下去, 最先出现迷宫边界节点时, 广度遍历的层次路径即为最少路径

职场精英工作室出品, 唯一淘宝旺旺客服: 蔚蓝小小天使
职场精英工作室出品, 唯一淘宝旺旺客服: 蔚蓝小小天使
职场精英工作室出品, 唯一淘宝旺旺客服: 蔚蓝小小天使