**2020年上海市高等学校信息技术水平考试试卷**

**四级 人工智能（视觉技术方向 模拟卷）**

（本试卷考试时间 150 分钟）

**一、单选题 ( 本大题 16 道小题 ，每小题 1 分，共 16 分），从下面题目给出的A、B、C、D四个可供选择的答案中选择一个正确答案。**

1.在回归模型中，下列\_\_\_\_在权衡欠拟合和过拟合中影响最大。

A.多项式阶数

B.更新权重 w 时，使用的是矩阵求逆还是梯度下降

C.使用常数项

D.增加数据量

2.A和B分别代表两个事件，如果P(A, B)降低，同时P(A)上升，\_\_\_\_是正确的。

A.P(B|A)降低

B.P(A|B)降低

C.P(B)降低

D.P(B)上升

3.癌症检查数据样本有10000个，其中10个数据祥本是有癌症，其它是无癌症。假设分类模型在无癌症数据9990中预测正确了9980个，在10个癌症数据中预测正确了9个，此时真阳=9，真阴=9980，假阳=10，假阴=1 。则该分类模型的F1-score为\_\_\_\_。

A.62.07%

B.99.89%

C.47.36%

D.76. 27%

4.在测试一假设h时，发现在一包含n=1000个随机抽取样例的样本s上，它出现r=300个错误, 计算errors(h)的标准差为\_\_\_\_。

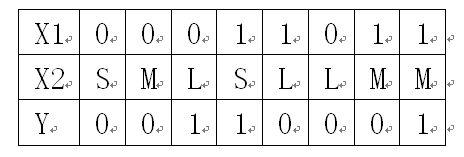
A.0.0145

B.0.145

C.1.45

D.14.5

5. 下表为某训练集数据，其中X1，X2为特征，Y为分类标记，则使用该训练集学习到的朴素贝叶斯分类器对x = (1，M) 的分类结果为\_\_\_\_。



A.0

B.1

C.不确定

D.0和1都有可能

6.关于主成分分析算法，以下步骤\_\_\_\_是错误的。

A.对所有样本进行去中心化

B.计算样本的协方差矩阵

C.对协方差矩阵做特征值分解

D.取最大的低维空间维数特征值所对应的特征向量输出投影矩阵

7.四个点坐标为(1,1)，(1,0)，(-1,-1)，(-1,0)，用 SVM 分类的决策边界是\_\_\_\_。

A.x = 0

B.y = x

C.y = -x

D.y = 0

8.在大数据集上训练决策树，为减少训练时间，可使用以下\_\_\_\_\_\_\_\_方法。

A.减少树的深度

B.增加树的深度

C.增加学习率

D.减少树的数量

9.关于偏差和方差，以下说法不正确的是\_\_\_\_。

A.如果能保证或验证一批训练集来自同一个分布，算法在这批训练集上的学习结果会是一致的

B.偏差-方差分解试图对学习算法的期望泛化错误率进行拆解

C.泛化误差可以分解为偏差、方差和噪声之和

D.方差与偏差通常是有冲突的，其中方差刻画数据扰动造成的影响，偏差刻画的是学习算法本身的拟合能力

10.下列哪个神经网络结构会发生权重共享\_\_\_\_\_\_\_\_。

A.卷积神经网络和循环神经网络

B.卷积神经网络

C.循环神经网络

D.全连接神经网络

11.\_\_\_\_\_\_不是双目立体视觉中的重要步骤。

A.图像滤波

B.立体匹配

C.相机标定

D.图像校正

12.图像的边缘以及其他灰度突变与\_\_\_\_\_\_相关。

A.高频分量

B.低频分量

C.具有高频和其他突变的边缘，灰度具有低频分量

D.低频边缘和其他灰度突变的高频分量

13.BatchNorm层对于input batch会统计出mean和 variance 用于计算 EMA。如果input batch 的 shape 为(B, C, H, W)，统计出的 mean 和 variance 的 shape 为\_\_\_\_\_\_。

A.1 \* C \* 1 \* 1

B.B \* 1 \* 1 \* 1

C.B \* C \* 1 \* 1

D.1 \* 1 \* 1 \* 1

14.梯度下降算法的正确步骤是\_\_\_\_\_\_。

a.计算预测值和真实值之间的误差

b.重复迭代，直至得到网络权重的最佳值

c.把输入传入网络，得到输出值

d.用随机值初始化权重和偏差

e.对每一个产生误差的神经元，调整相应的（权重）值以减小误差

A.dcaeb

B.daceb

C.dceab

D.daecb

15.在训练一个卷积神经网络时，不同最优化方式，如SGD，ADAM下列说法中正确的是\_\_\_\_\_\_。

A.相同超参数数量情况下，比起自适应的学习率调整方式，SGD加手动调节通常会取得更好效果

B.在实际场景下，应尽量使用ADAM，避免使用SGD

C.同样的初始学习率情况下，ADAM的收敛速度总是快于SGD方法

D.同样的初始学习率情况下，ADAM比SGD容易过拟合

16.在训练一个卷积神经网络时，不同最优化方式，如SGD，ADAM下列说法中正确的是\_\_\_\_\_\_。

A.相同超参数数量情况下，比起自适应的学习率调整方式，SGD加手动调节通常会取得更好效果

B.在实际场景下，应尽量使用ADAM，避免使用SGD

C.同样的初始学习率情况下，ADAM的收敛速度总是快于SGD方法

D.同样的初始学习率情况下，ADAM比SGD容易过拟合

**二、多选题 ( 本大题 6 道小题 ，每小题 2 分，共 12 分），从下面题目给出的A、B、C、D四个可供选择的答案中选择所有正确答案。**

1.关于朴素贝叶斯分类方法描述正确的有\_\_\_\_。

A.需要计算先验概率

B.对缺失数据敏感

C.对小规模的数据表现很好

D.算法成立的前提是假设各属性之间互相独立

2.正则化能处理过拟合的原因是\_\_\_\_\_\_\_\_。

A.惩罚了模型的复杂度，避免模型过度学习训练集，提高泛化能力

B.正则项降低了每一次系数w更新的步伐，使参数更小，模型更简单

C.正则化使得训练集和测试集数据分布更为接近，因而避免了在训练集上过拟合

D.贝叶斯学派的观点，认为加入了先验分布（l1拉普拉斯分布，l2高斯分布），减少参数的选择空间

3.通过监督学习进行二分类模型训练过程中，可能会遇到正负样本分布不均的情况(比如正样本有50万但是负样本有100万)，下列\_\_\_\_方法可以进行恰当处理。

A.将所有的数据加入训练集，充分利用所有数据

B.从100万负样本中随机抽取50万

C.正样本权重设置为2，负样本权重设置为1

D.复制两份正样本参与到训练中去

4.\_\_\_\_\_\_是Lucas-Kanade(LK)光流法的基本假设。

A.空间一致

B.恒定亮度

C.时间连续

D.剧烈运动

5.RANSAC算法经常用于计算机视觉中，下列关于RANSAC算法的说法正确的是\_\_\_\_\_\_。

A.RANSAC是"RANdom SAmple Consensus（随机抽样一致）"的缩写

B.RANSAC 算法是一种不确定性算法，有可能得不到合理的结果

C.RANSAC算法中假设的inliers是符合模型的

D.RANSAC算法中假设的outliers不是符合模型的

6.以下深度学习常用的激活函数中，保留副半轴激活值的有\_\_\_\_\_\_。

A.Tanh

B.PReLU

C.Swish

D.ELU

**三、是非题 ( 本大题 15 道小题 ，每小题 1 分，共 15 分），从下面题目给出的两个可供选择的答案中选择一个正确答案。**

1.如果决策树对训练集拟合不足，通过缩放输入特征可以有效改善。

A.正确

B.错误

2.Boosting和Bagging都是组合多个分类器投票的方法，二者都是根据单个分类器的正确率决定其权重。

A.正确

B.错误

3.如果两个变量的 Pearson 相关性系数为零，则它们不相关。

A.正确

B.错误

4.当不知道数据所带标签时，可以使用分类技术促使带同类标签的数据与带其他标签的数据相分离。

A.正确

B.错误

5.回归问题和分类问题都有可能发生过拟合。

A.正确

B.错误

6.向量x=[1,2,3,4,-9,0]的L1范数是1。

A.正确

B.错误

7.神经网络中激活函数引入了非线性。

A.正确

B.错误

8.参数化的方法可以使得类条件概率估计简化，但是估计结果的准确性严重依赖于所假设的概率分布形式是否符合真实数据分布。

A.正确

B.正确

9.假设我们有三个簇中心μ1=[1;2]，μ2=[-3 ;0];，μ3=[4 ;2]。此外，我们还有一个训练示例x(i)=[-2;1]。则在一个集群分配步骤之后，c(i)将会是2。

A.正确

B.错误

10.评价规则优劣的标准应该优先考虑规则准确率，同时考虑覆盖样例数和属性次序。

A.正确

B.错误

11.一个跨度比较大的灰度图像具有较低的动态范围，而暗淡的灰度图具有较高的动态范围。

A.正确

B.错误

12.直方图均衡是一种全局运算，图像的二值化是一种局部运算。

A.正确

B.错误

13.在小波变换的步骤中，图像被连续地分割成更小的区域。

A.正确

B.错误

14.MobileNet采用了深度可分离卷积卷积结构代替传统的卷积方式。

A.正确

B.错误

15.一个常见的提高生成对抗网络（GAN）训练速度的方法是，每训练一次判别器中的参数, 训练k(k ≥ 1)次生成器。

A.正确

B.错误

**四、操作题**

**素材、样张、KS目录均在zip文件中，可双击此图标打开**

**以下第（一）题（案例应用题）题目请在文件"C:\KS\人工智能-计算机视觉-答题纸.docx"中作答！**

(一)、案例应用题（共20分）

高空抛物具有非常大的危险性，由高空抛物引发的恶性案件近年来居高不下，因此国家和地方政府纷纷针对高空抛物行为进行专项立法。如何获取高空抛物的相关证据是目前该类型案件中的难点。基于视频监控系统的解决方案是目前较为可靠的证据来源和预防手段。因此，融入AI技术的智能视频分析系统越来越受到关注。

1.请简述高空抛物智能识别统主要由那几个主要模块组成，每个模块的功能是什么？（满分6分）

|  |
| --- |
| **请在答题纸作答！此处答题一律无效！** |

2.请列举高空抛物场景中常见的目标检测算法有哪些？（每项1分，满分6分）

|  |
| --- |
| **请在答题纸作答！此处答题一律无效！** |

3.请简述利用智能视频分析系统进行高空抛物识别可能存在哪些技术问题。（每项1分，满分4分）

|  |
| --- |
| **请在答题纸作答！此处答题一律无效！** |

4.请列举融入AI技术的智能视频分析系统在其它领域的应用。（每项1分，满分4分）

|  |
| --- |
| **请在答题纸作答！此处答题一律无效！** |