

2023 年上海市高等学校信息技术水平考试试卷

四级 人工智能——语音识别（模拟卷）

（本试卷考试时间 150 分钟）

一、单选题（本大题 18 道小题，每小题 1 分，共 18 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. 两位同事从上海出发前往深圳出差，他们在不同时间出发，搭乘的交通工具也不同，能准确描述两者“上海到深圳”距离差别的是_____。

- A. 欧式距离
- B. 余弦距离
- C. 曼哈顿距离
- D. 切比雪夫距离

2. SVM（支持向量机）与 LR（逻辑回归）在数学本质上的区别是_____。

- A. 损失函数
- B. 是否有核技巧
- C. 是否支持多分类
- D. 其余选项皆错

3. _____不能防止过拟合。

- A. 交叉验证
- B. 低维嵌入
- C. 剪枝
- D. 集成学习

4. 一个计算机程序从经验 E 中学习任务 T，并用 P 来衡量表现。并且，T 的表现 P 随着经验 E 的增加而提高。假设我们给一个学习算法输入了很多历史天气的数据，让它学会预测天气。_____是 P 的合理选择。

- A. 计算大量历史气象数据的过程
- B. 天气预报任务
- C. 正确预测未来日期天气的概率
- D. 其余选项皆不是

5. 关于 PCA（主成分分析）算法和 KPCA（核主成分分析）算法的说法中，正确的是_____。

- A. PCA 所做的是对坐标轴线性变换，即变换后的新基是一条曲线
- B. KPCA 对坐标轴做了线性变换，数据所映射的新基是一条直线
- C. PCA 与 KPCA 都可以对数据进行降维
- D. KPCA 算法应用到了核函数，所以在计算量上有一定减小

6. 以下关于 VC 维（Vapnik-Chervonenkis Dimension）的描述，正确的是_____。

- A. 在一个假设空间 H 中，如果 $d_{vc}(H)$ 是无穷大，则存在 VC bound
- B. 在一个假设空间 H 中，如果 $d_{vc}(H)$ 是有限的，则不存在 VC bound

- C. 模型较为复杂时, d_{vc} 通常较小
D. VC 维反应了模型的学习能力, VC 维越大, 则模型的容量越大
7. 以下关于梯度下降法的描述, 错误的是_____。
- A. 随机梯度下降算法每次基于一个样本来更新梯度值, 这导致更新过程波动较大, 收到噪声影响较大
B. 批量梯度下降算法每次基于全部样本来更新梯度值, 这使得其更新速度较慢, 但是更新过程较为平稳
C. 小批量梯度下降算法每次选取一个 Batch Size 的数据样本进行梯度值更新, 它综合了批量梯度下降与随机梯度下降的优缺点, 是常用的梯度下降方法
D. 批量梯度下降方法在大数据集上的收敛速度最快
8. 某工程师在构建机器学习模型时发现模型在训练集和测试集上表现都很差, 可能的原因是_____。
- A. 模型过于复杂, 但是训练集比较小
B. 正则化参数中惩罚项系数取值太小
C. 测试集样本数量太少
D. 模型训练时间过长
9. 在进行分类模型建模时, 如果样本类别不均衡, 处理方式错误的是_____。
- A. 过采样小样本, 欠采样大样本
B. 丢弃小样本类别
C. 通过集成方法解决
D. 采集新的数据
10. A 工厂生产配件的合格率是 97%, B 工厂生产配件的合格率是 96%, C 公司从 A 工厂采购了 60%的配件, 从 B 工厂采购了 40%的配件, 现抽检到一个不合格的配件, 该配件是 A 工厂生产的概率是_____。
- A. 52.9%
B. 43.9%
C. 67%
D. 47.3%
11. 常用的语音合成前端模型有_____。
- A. BERT
B. Tacotron
C. VQVAE
D. HiFiGAN
12. 通常人耳可感受到的声音频率范围为_____。
- A. 0-20Hz
B. 10-20KHz
C. 200Hz-20KHz
D. 20Hz-40KHz

13. _____不能够提升解码的速度。
- A. 使用裁剪的轻量化声学模型
 - B. 使用多线程解码
 - C. 使用二遍解码
 - D. 设计良好的解码算法
14. DNN-HMM (Deep Neural Network-Hidden Markov Model) 声学模型的训练步骤如下, 其中顺序正确的是_____。
- a. 基于 Viterbi 算法的强制对齐方法给每个语音帧打上一个 HMM 状态标签
 - b. 训练一个 GMM-HMM 声学模型
 - c. 用 DNN 模型替换 HMM 模型中计算观察概率的 GMM 部分, 但保留转移概率和初始概率等部分
 - d. 训练一个基于 DNN 训练算法的 DNN 模型
- A. abcd
 - B. cdab
 - C. dacb
 - D. badc
15. 麦克风阵列相对于单麦克风的优点, 主要是能利用_____。
- A. 语音信息
 - B. 噪声信息
 - C. 空间信息
 - D. 时间信息
16. _____不属于端到端语音识别模型。
- A. CTC
 - B. LAS
 - C. RNN-T
 - D. HMM
17. 语音信号的倒谱分析就是求取语音倒谱特征参数的过程, 它可以通过_____来实现。
- A. 线性预测
 - B. 维纳滤波
 - C. 同态信号处理
 - D. 卡尔曼滤波
18. 在实际应用中如果出现新的词汇, 为了提高识别效果, _____模块必须进行修改。
- A. 声学模型
 - B. 发音词典
 - C. 解码
 - D. HMM 模型

二、多选题（本大题 10 道小题，每小题 2 分，共 20 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。

- 关于神经网络，说法正确的是_____。
 - 增加网络层数，可能会增加测试集分类错误率
 - 增加网络层数，一定会增加训练集分类错误率
 - 减少网络层数，可能会减少测试集分类错误率
 - 减少网络层数，一定会减少训练集分类错误率
- 属于稀疏表示应用领域的是_____。
 - 图像去噪
 - 压缩感知
 - 人脸识别
 - 目标跟踪
- 关于集成学习的说法中，正确的是_____。
 - Bagging 是通过结合几个模型降低泛化误差的技术。主要思想是分别训练几个不同的模型，然后让所有模型表决测试样例的输出
 - Boosting 的代表算法是随机森林
 - Boosting 是一种框架算法，主要是通过对样本集的操作获得样本子集，然后用弱分类算法在样本子集上训练生成一系列的基分类器
 - Stacking 是将各个弱学习器的学习成果，并行结合起来，形成以预测值（标签）为数据的训练集，用来训练下一层学习器
- 关于决策树算法的描述，正确的是_____。
 - 构建决策树时，可以选择信息熵与 GINI 系数作为分割指标
 - 构建决策树时，数据集分割前与分割后的“纯度”差异越小，决策树越好
 - 决策树既可以处理分类问题，也可以处理回归问题
 - 不同的决策树算法差异点主要在特征选择过程中
- 关于不同的模型评价指标，正确的是_____。
 - 对于样本类别分布均衡的分类问题，准确率是有意义的
 - 回归问题可以采用 MSE 作为评价指标
 - F1 值综合了准确率和召回率
 - 召回率越高意味着模型性能越好
- 属于声学特征的有_____。
 - Mel 频率倒谱系数 (MFCC)
 - 感知线性预测系数 (PLP)
 - 滤波器组 (FBank)
 - 语谱图
- 语音识别系统中最常用的评价指标是词错误率 (Word Error Rate, WER)。在中文里，通常使用字错误率 (Character Error Rate, CER) 来表示。WER 的计算方法是，对于一段音频，已知其标注文本和语音识别的结果，将识别结果中错误词的累计个数除以标注中总的

词数，结果表示为一个百分数。对于错误词常用的操作有_____。

- A. 插入错误
- B. 删除错误
- C. 重复错误
- D. 替换错误

8. 语音信号处理也可以简称为语音处理，它是利用数字信号处理技术对语音信号进行处理的一门学科，属于语音处理范畴的是_____。

- A. 语音编码
- B. 语音合成
- C. 语音识别
- D. 说话人识别

9. 关于 N-gram 语言模型和神经网络 (NN) 语言模型的优缺点说法正确的是_____。

- A. NN 语言模型具有长距离依赖，具有更强的约束性
- B. N-gram 语言模型采用了极大似然估计
- C. NN 语言模型无法避免数据稀疏所带来的集外词问题
- D. N-gram 语言模型较 NN 语言模型有更好的可解释性

10. 关于共振峰语音合成方法，说法正确的是_____。

- A. 靠前的共振峰往往容易受到噪声等环境干扰，因此越靠后的共振峰越重要
- B. 共振峰合成方法具有很强的韵律修改能力
- C. 通过共振峰频率的移动，可以修改合成语音中与说话人特征相关的部分
- D. 无论基音频率如何改变，共振峰的特性几乎不受影响

三、是非题（本大题 23 道小题，每小题 1 分，共 23 分）。

1. xgboost 是一种优秀的集成算法，其优点包括速度快，对异常值不敏感，支持自定义损失函数等等。

2. 线性回归的自变量和残差不一定保持相互独立。

3. L1 正则和 L2 正则的共同点是都会让数据集中的特征数量减少。

4. 过拟合是有监督学习的挑战，而不是无监督学习。

5. 对于一个 SVM（支持向量机），去除不支持的向量后仍然能分类。

6. 给定 n 个数据点，如果其中一半用于训练，一半用于测试，则训练误差和测试误差之间的差别会随着 n 的增加而减少。

7. 随机梯度下降 (Stochastic Gradient Descent) 算法是用小规模样本近似估计梯度的方法，适合在大规模数据上训练深度神经网络，但在逻辑回归、SVM 等算法中的作用很有限。

8. 对于某个神经网络，其激活函数是 ReLU。若使用线性激活函数代替 ReLU，那么该神经网络仍然能表征 XNOR 函数。
9. ID3 决策树学习算法以信息增益为准则来选择划分属性；C4.5 决策树算法使用增益率来选择最优划分属性；CART 决策树使用 Gini（基尼）指数来选择划分属性。
10. 梯度下降法可能陷于局部极小值，EM 算法则不会。
11. 深度学习与机器学习算法之间的区别在于，后者无需进行特征提取。
12. 每次使用 K-means 聚类算法得到的聚类结果可能会不一样。
13. 在某神经网络的隐藏层输出中包含-1.5，那么该神经网络采用的激活函数不可能是 sigmoid、tanh、relu。
14. 在线商品 A 的用户点击率为 1%，用某模型进行点击预测，得到了 99% 的预测准确率，则该模型的预测准确率很高，可以投入使用。
15. 强化学习理论受到行为主义心理学启发，侧重在线学习并试图在探索-利用（exploration-exploitation）间保持平衡。不同于监督学习和非监督学习，强化学习不要求预先给定任何数据，而是通过接收环境对动作的奖励（反馈）获得学习信息并更新模型参数。
16. 掩蔽效应分为同时掩蔽和异时掩蔽。同时掩蔽又成为时域掩蔽，异时掩蔽又称为频域掩蔽。
17. 语音识别技术正在不断发展中，语音识别技术可以实现一种语种的语音转文本的转写，但是不能实现多种语种的混合交谈转写。
18. PPL（Perplexity）越小，一般表明语言模型性能越差。
19. 基于双向 LSTM（Long-Short Term Memory）声学模型的语音识别系统同时利用未来和历史信息对当前时刻声学特征建模，可以应用于要求实时响应的在线语音识别任务中。
20. 使用 GMM 对 HMM 每个状态的语音特征分布进行建模，只要混合的高斯分布数目足够多，GMM 可以拟合任意精度的概率分布。
21. 语音识别解码器利用加权有限状态转换器（WFST），把发音词典、声学模型和语言模型编译成动态解码网络，可以大大提高语音识别系统的准确率，但是会一定程度上降低识别速度。
22. 可以采用模型状态绑定的方法，减少三音子建模时 HMM 的状态参数。
23. 基本声学单元（简称基元）的选择是声学模型建模中一个基本而重要的问题。基元的选择

可以基于语音学只是选择包括词、音节、半音节、声韵母和音素等，也可以基于数据驱动的方式产生。

四、操作题

以下案例应用题题目请在文件“C:\KS\人工智能-语音识别-答题纸.docx”中作答！

案例应用题（20分）

以往人们学习英语，模仿单词朗读只能靠反复的听音辨音；后来有了语音匹配技术，可以对学生单词和句子朗读的标准程度打分，提升了英语口语学习的效率。某在线语言教学公司为了进一步提升对学习者的口语发音打分的准确率，改善用户体验，计划推出结合视频和语音的打分系统。该系统在有外教示范发音视频(含语音)的条件下，录制学习者的发音视频(含语音)并对口语发音水平进行评价。

您所就职的人工智能公司接到了该在线语言教学公司的联合开发该打分系统的需求，为其设计和提供结合声音和脸部动作的多维度多模态的英语口语发音水平的打分算法。假设贵公司已经有了一套效果很好的英语语音识别系统和面部关键点检测模块供你使用，且当前仅需要对单词朗读打分，不需对整句朗读打分。

问题：

- 1) 需要对外教的示范视频提前做哪些处理？(6分)

请在答题纸作答！此处答题一律无效！

- 2) 学习者在开始朗读前，系统需要做哪些检查工作确保学习者给予较好的配合？(4分)

请在答题纸作答！此处答题一律无效！

- 3) 从用户开始朗读到给出打分有哪些关键算法流程？(8分)

请在答题纸作答！此处答题一律无效！

- 4) 在这个系统中，除了给出朗读得分外，还可以输出哪些信息帮助学习者提高其口语发音？(2分)

请在答题纸作答！此处答题一律无效！