机器学习复习题

一、单选题

1. 属于监督学习的机器学习算法是( )
   A. 贝叶斯分类器
   B. 主成分分析
   C. K-Means
   D. 高斯混合聚类

2. 属于无监督学习的机器学习算法是( )
   A. 支持向量机
   B. Logistic回归
   C. 层次聚类
   D. 决策树

3. 朴素贝叶斯分类器的特点是( )
   A. 假设样本服从正态分布
   B. 假设样本服从多项式分布
   C. 假设样本各维属性独立
   D. 假设样本各维属性存在依赖

4. 下列属于线性分类方法的是( )
   A. 决策树
   B. 感知机
   C. 最近邻
   D. 集成学习

5. 下列方法不受数据归一化影响的是( )
   A. SVM
   B. 神经网络
   C. Logistic回归
   D. 决策树

6. 关于线性鉴别分析的描述最准确的是，找到一个投影方向，使得( )
   A. 类内距离最大，类间距离最小
   B. 类内距离最小，类间距离最大
   C. 类内距离最大，类间距离最大
   D. 类内距离最小，类间距离最小

7. SVM的原理可简单描述为( )
   A. 最小均方误差分类

B. 最小距离分类
C. 最大间隔分类
D. 最近邻分类

8. SVM的算法性能取决于([redacted])
A. 核函数的选择
B. 核函数的参数
C. 软间隔参数C
D. 以上所有

9. 支持向量机的对偶问题是([redacted])
A. 线性优化问题
B. 二次优化
C. 凸二次优化
D. 有约束的线性优化

10. 以下对支持向量机中的支撑向量描述正确的是([redacted])
A. 最大特征向量
B. 最优投影向量
C. 最大间隔支撑面上的向量
D. 最速下降方向

11. 假定你使用阶数为2的线性核SVM，将模型应用到实际数据集上后，其训练准确率和测试准确率均为100%。现在增加模型复杂度（增加核函数的阶），会发生以下哪种情况([redacted])
A. 过拟合
B. 欠拟合
C. 什么都不会发生，因为模型准确率已经到达极限
D. 以上都不对

12. 关于决策树节点划分指标描述正确的是([redacted])
A. 类别非纯度越大越好
B. 信息增益越大越好
C. 信息增益越小越好
D. 基尼指数越大越好

13. 以下描述中，属于决策树策略的是([redacted])
A. 最优投影方向
B. 梯度下降方法
C. 最大特征值
D. 最大信息增益

14. 集成学习中基分类器如何选择，学习效率通常越好([redacted])
A. 分类器相似

B. 都为线性分类器
C. 都为非线性分类器
D. 分类器多样，差异大

15. 集成学习中，每个基分类器的正确率的最低要求( )
A. 50%以上
B. 60%以上
C. 70%以上
D. 80%以上

16. 下面属于Bagging方法的特点是( )
A. 构造训练集时采用Bootstraping的方式
B. 每一轮训练时样本权重不同
C. 分类器必须按顺序训练
D. 预测结果时，分类器的比重不同

17. 下面属于Boosting方法的特点是( )
A. 构造训练集时采用Bootstraping的方式
B. 每一轮训练时样本权重相同
C. 分类器可以并行训练
D. 预测结果时，分类器的比重不同

18. 随机森林方法属于( )
A. 梯度下降优化
B. Bagging方法
C. Boosting方法
D. 线性分类

19. 软间隔SVM的阈值趋于无穷，下面哪种说法正确( )
A. 只要最佳分类超平面存在，它就能将所有数据全部正确分类
B. 软间隔SVM分类器将正确分类数据
C. 会发生误分类现象
D. 以上都不对

20. 回归问题和分类问题的区别( )
A. 前者预测函数值为连续值，后者为离散值
B. 前者预测函数值为离散值，后者为连续值
C. 前者是无监督学习
D. 后者是无监督学习

21. 正则化的回归分析，可以避免( )
A. 线性化
B. 过拟合
C. 欠拟合

D. 连续值逼近

22. “啤酒-纸尿布”问题讲述的是，超市购物中，通过分析购物单发现，买了纸尿布的男士，往往又买了啤酒。这是一个什么问题( )
A. 关联分析
B. 回归
C. 聚类
D. 分类

23. 混合高斯聚类中，运用了以下哪种过程( )
A. EM算法
B. 集合运算
C. 密度可达
D. 样本与集合运算

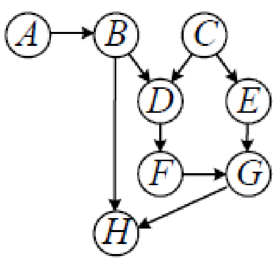
24. 主成分分析方法是一种什么方法( )
A. 分类方法
B. 回归方法
C. 降维方法
D. 参数估计方法

25. PCA在做降维处理时，优先选取哪些特征( )
A. 中心化样本的协方差矩阵的最大特征值对应特征向量
B. 最大间隔投影方向
C. 最小类内聚类
D. 最速梯度方向

26. 过拟合现象中( )
A. 训练样本的测试误差最小，测试样本的正确识别率却很低
B. 训练样本的测试误差最小，测试样本的正确识别率也很高
C. 模型的泛化能力很高
D. 通常为线性模型

27. 多层感知机方法中，可用作神经元的非线性激活函数( )
A. logistic 函数
B. p-范数
C. 线性内积
D. 加权求和
E.

28. 如右图所示有向图，以下陈述正确的有( )
A. B和G关于{C, F}条件独立
B. B和C关于F条件独立
C. B和G关于F条件独立
D. B和G关于{C, F, H}条件独立



29. 梯度下降算法的正确步骤是什么( )

(1)计算预测值和真实值之间的误差

(2)迭代更新，直到找到最佳权重

(3)把输入传入网络，得到输出值

(4)初始化随机权重和偏差

(5)对每一个产生误差的神经元，改变相应的（权重）值以减小误差

A. 1, 2, 3, 4, 5

B. 4, 3, 1, 5, 2

C. 3, 2, 1, 5, 4

D. 5, 4, 3, 2, 1

30. 假如使用一个较复杂的回归模型来拟合样本数据，使用岭回归，调试正则化参数λ，来降低模型复杂度。若λ较大时，关于偏差(bias)和方差(variance)，下列说法正确的是

A. 若λ较大时，偏差减小，方差减小

B. 若λ较大时，偏差减小，方差增大

C. 若λ较大时，偏差增大，方差减小

D. 若λ较大时，偏差增大，方差增大

31. 以下哪种方法会增加模型的欠拟合风险

A. 添加新特征

B. 增加模型复杂度

C. 减小正则化系数

D. 数据增强

32. 增加以下哪些超参数可能导致随机森林模型过拟合数据

(1). 决策树的数量；(2). 决策树的深度； (3). 学习率。

A. (1)

B. (2)

C. (3)

D. (2)(3)

33. 以下关于深度网络训练的说法正确的是

A. 训练过程需要用到梯度，梯度衡量了损失函数相对于模型参数的变化率

B. 损失函数衡量了模型预测结果与真实值之间的差异

C. 训练过程基于一种叫做反向传播的技术

D. 其他选项都正确

34. 关于CNN，以下结论正确的是

A. 在同样层数、每层神经元数量一样的情况下，CNN比全连接网络拥有更多的参数

B. CNN可以用于非监督学习，但是普通神经网络不行

C. Pooling层用于减少图片的空间分辨率

D. 接近输出层的filter主要用于提取图像的边缘信息

35. 关于k-means算法，正确的描述是

A. 能找到任意形状的聚类

B. 初始值不同，最终结果可能不同

C. 每次迭代的时间复杂度是$O(n^2)$，其中n是样本数量

D. 不能使用核函数

36. 下列关于过拟合现象的描述中，哪个是正确的

A. 训练误差小，测试误差大

B. 训练误差小，测试误差小

C. 模型的泛化能力高

D. 其余选项都不对

37. 下列哪个函数不可以做激活函数

A. y=tanh(x)

B. y=sin(x)

C. y=max(x,0)

D. y=2x

38. 在其他条件不变的前提下，以下哪种做法容易引起机器学习中的过拟合问题

A. 增加训练集量

B. 减少神经网络隐藏层节点数

C. 删除稀疏的特征

D. SVM算法中使用高斯核代替线性核

39. 下面方法中属于无监督学习算法的是

A. 线性回归

B. 支持向量机

C. 决策树

D. K-Means聚类

40. Bootstrap数据是什么意思

A. 有放回地从总共M个特征中抽样m个特征

B. 无放回地从总共M个特征中抽样m个特征

C. 有放回地从总共N个样本中抽样n个样本

D. 无放回地从总共N个样本中抽样n个样本

41. 下面关于Adaboost算法的描述中，错误的是

A. 是弱分类器的线性组合

B. 提升树是以分类树或者回归树为基本分类器的提升办法

C. 该算法实际上是前向分步算法的一个实现，在这个方法里，模型是加法模型，损失函数是指数损失，算法是前向分步算法。

D. 同时独立地学习多个弱分类器

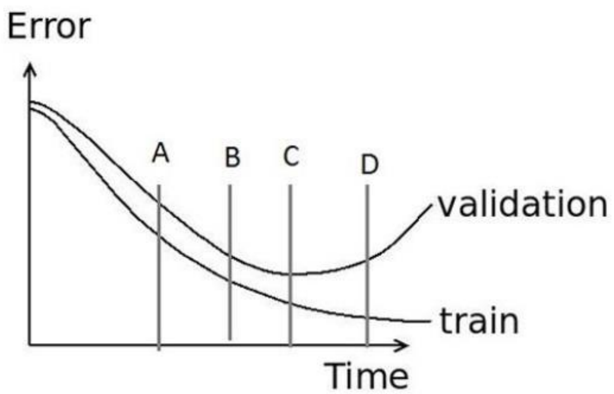
42. 在HMM中,如果已知观察序列和产生观察序列的状态序列,那么可用以下哪种方法直接进行参数估计
A. EM算法
B. 维特比算法
C. 前向后向算法
D. 极大似然估计

43. 以下哪种距离会侧重考虑向量的方向
A. 欧式距离
B. 海明距离
C. Jaccard距离
D. 余弦距离

44. 解决隐马模型中预测问题的算法是
A. 前向算法
B. 后向算法
C. Baum-Welch算法
D. 维特比算法

45. 梯度爆炸问题是指在训练深度神经网络的时候，梯度变得过大而损失函数变为无穷。在RNN中，下面哪种方法可以较好地处理梯度爆炸问题( )
A. 梯度裁剪　B. 所有方法都不行
C. Dropout　D. 加入正则项

46. 当训练一个神经网络来作图像识别任务时，通常会绘制一张训练集误差和验证集误差图来进行调试。在下图中，最好在哪个时间停止训练( )



A. A　B. B　C. C　D. D

47. 当不知道数据所带标签时，可以使用哪种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. 分类 B. 聚类 C. 关联分析 D. 隐马尔可夫链

48. 现在需要计算三个稠密矩阵A,B,C的乘积ABC，假设三个矩阵的尺寸分别为m

×n，n×p，p×q，且m <n <p <q，不考虑矩阵乘法的优化时，以下计算顺序效率最高的是

A. (AB)C
B. AC(B)
C. A(BC)
D. 效率都相同

49. 下列方法中没有考虑先验分布的是()
A. 最大后验估计
B. 贝叶斯分类器
C. 贝叶斯学习
D. 最大似然估计

50. 下列哪一项主要负责在神经网络中引入非线性？
A. 随机梯度下降
B. 修正线性单元（ReLU）
C. 输入的加权求和
D. 以上都不正确

51. 下列哪一种架构有反馈连接并常被用来处理序列数据？
A. 循环神经网络
B. 卷积神经网络
C. 全连接网络
D. 都不是

52. 在一个神经网络中，下面哪种方法可以用来处理过拟合？
A. Dropout
B. 分批归一化(Batch Normalization)
C. 正则化(regularization)
D. 都可以

53. 某小区人脸识别准入系统用来识别待进入人员的身份，此系统一共包括识别3种不同的人员：业主，物业人员，未收录人员。下面哪种学习方法最适合此种应用需求：。
A. 二分类
B. 多分类
C. 层次聚类
D. 线性回归

54. L1与L2范数在Logistic Regression 中，如果同时加入L1和L2范数，会产生什么效果()。
A. 可以做特征选择，并在一定程度上防止过拟合
B. 能解决维度灾难问题
C. 能加快计算速度

D. 能增加模型的拟合能力

55. 下列模型中属于生成式模型的是)
A. 线性分类器
B. 卷积神经网络
C. 线性判别分析
D. 朴素贝叶斯模型

56. 下列模型中属于判别式模型的是
A. 支持向量机
B. 隐马尔可夫模型
C. 朴素贝叶斯模型
D. 高斯混合模型

57. 下列属于无监督学习的是)
A. k-means
B. SVM
C. 最大熵
D. CRF

58. 关于“过拟合”现象的出现范围，下列说法哪个是正确的
A. 只在监督学习中出现
B. 只在非监督学习中出现
C. 在监督学习和非监督学习中都可能出现
D. 在任何类型的学习中都不会出现

59. 我们想在大数据集上训练决策树，为了使用较少时间，我们可以)
A. 增加树的深度
B. 增加学习率 (learning rate)
C. 减少树的深度
D. 减少树的数量

60. 对于k折交叉验证，以下对k的说法正确的是
A. k越大，不一定越好，选择大的k会加大评估时间
B. 选择更大的k，就会有更小的bias ，因为训练集更加接近总数据集
C. 在选择k时，要最小化数据集之间的方差
D. 以上所有

61. 以下不属于贝叶斯分类器参数估计的准则的是
A. 最大高斯后验
B. 最大beta后验
C. 最大间隔
D. 极大似然

62. 下列选项中属于机器学习可解决的问题的有

A. 分类

B. 聚类

C. 回归

D. 以上均可

63. 下列选项中，关于KNN算法说法不正确的是（

A. 能找出与待测样本相近的K个样本

B. 可以使用欧氏距离度量相似度

C. 实现过程相对简单

D. 效率很高

64. 关于特征预处理，下列说法中错误的是（ ）

A. 包含标准化和归一化

B. 标准化在任何场景下受异常值的影响都很小

C. 归一化利用了样本中的最大值和最小值

D. 标准化实际上是将数据在样本的标准差上做了等比例的缩放操作

65. 关于交叉验证，下列说法中错误的是 ）

A. 交叉验证能够直接提升模型的准确率

B. 交叉验证能够提供对模型泛化性能的更可靠估计

C. 交叉验证搭配网格搜索能够提升我们查找最优超参数组合的效率

D. 使用网格搜索时我们一般会提供超参数的可能取值字典

66. 请选择下面可以应用隐马尔可夫（HMM）模型的选项：

A. 基因序列数据集

B. 电影浏览数据集

C. 股票市场数据集

D. 所有以上

67. EM算法（Expectation Maximization Algorithm）是机器学习领域的一个经典算法，下面关于EM算法的说法中不正确的有：

A. EM算法属于一种分类算法

B. EM算法可用于隐马尔科夫模型的参数估计

C. EM算法可以分为E-step和M-step两步

D. EM算法可用于从不完整的数据中计算最大似然估计

68. 关于SVM的损失函数，下列说法中错误的是：

A. SVM适用于多种损失函数

B. 0/1损失函数的最终结果只有两个，0代表分类正确，1代表分类错误

C. 合页损失(Hinge loss)衡量了被误分类的样本离分割超平面的距离的大小程度

D. 分类SVM常用平方误差损失来衡量模型的好坏

69. 关于SVM核函数，下列说法中错误的是：
A. 核函数的引入提升了SVM在线性不可分场景下的模型的稳健性
B. 核函数就是一类具有将某一类输入映射为某一类输出的函数
C. 核函数把特征映射到的空间维度越高越好
D. 常见的核函数有线性核、高斯核、多项式核、sigmoid核

70. 下列关于Kmeans聚类算法的说法错误的是)
A. 对大数据集有较高的效率并且具有可伸缩性
B. 是一种无监督学习方法
C. 初始聚类中心随机选择
D. 初始聚类中心的选择对聚类结果影响不大

71. 关于朴素贝叶斯，下列说法错误的是：
A. 它是一个分类算法
B. 朴素的意义在于它基于假设：所有特征之间是相互独立的
C. 它实际上是将多条件下的条件概率转换成了单一条件下的条件概率，简化了计算
D. 以贝叶斯估计的角度来看朴素贝叶斯时，其没有估计联合概率

72. 避免直接的复杂非线性变换，采用线性手段实现非线性学习的方法是( )
A. 核函数方法
B. 集成学习
C. 线性鉴别分析
D. PCA

73. 下列选项中，关于逻辑斯蒂回归的说法不正确是：
A. 逻辑斯蒂回归是监督学习
B. 逻辑斯蒂回归是一个回归模型
C. 逻辑斯蒂回归是一个分类模型
D. 逻辑斯蒂回归使用sigmoid函数作为激活函数对回归的结果做了映射

74. 下列关于样本类别不均衡场景的描述正确的是 (
A. 样本类别不均衡会影响分类模型的最终结果
B. 样本类别不均衡场景下我们没有可行的解决办法
C. 欠采样是复制类别数较少的样本来进行样本集的扩充
D. 过采样会造成数据集部分信息的流失

75. 下列关于无监督学习描述错误的是
A. 无标签信息
B. 聚类是其中一个应用
C. 不能使用降维
D. 在现实生活中有广泛的应用

76. 将一个k分类问题分解成一对一分类问题时总共需要（ ）个分类器
A. k(k-1)/2
B. k(k-1)
C. k
D. k!

77. 下列关于聚类说法错误的是（ ）
A. 无需样本有标签
B. 可用于抽取一些特征
C. 可提取关于数据的结构信息
D. 同一个类内的样本之间差异较大

78. 下列关于k-means说法不正确的是（ ）
A. 算法有可能终止于局部最优解
B. 簇的数目需要事先给定
C. 对噪声和离群点敏感
D. 适合处理非凸型数据

79. 在有限支撑集上，下面分布中熵最大的是（ ）
A. 几何分布
B. 指数分布
C. 高斯分布
D. 均匀分布

80. 给定均值和方差的情况下，下面分布中熵最大的是（ ）
A. 几何分布
B. 指数分布
C. 高斯分布
D. 均匀分布

二、多选题

1. 可用于贝叶斯决策的函数（ ）
A. $\omega^* = \arg\max_{\omega_i} p(x|\omega_i)p(\omega_i)$

B. $g(x) = p(\omega_1|x) - p(\omega_2|x)$

C. $g(x) = \ln\frac{p(x|\omega_1)}{p(x|\omega_2)} + \ln\frac{p(\omega_1)}{p(\omega_2)}$

D. $p(\omega_1 | x)$

2. 对聚类问题描述不正确的（ ）
A. 监督学习
B. 无监督学习
C. 线性决策
D. 增量学习

3. 以下属于聚类方法的是（ ）
A. k-means
B. 层次聚类
C. Fisher鉴别
D. 密度聚类

4. 影响K-Means聚类算法结果的主要因素有（ ）
A. 样本顺序
B. 相似性度量
C. 初始聚类中心
D. 样本类别

5. 以下选项中属于聚类问题可用的相似性度量准则有（ ）
A. 样本-样本距离
B. 样本-集合距离
C. 集合-集合距离
D. 集合内样本间距

6. 以下选项中可用于实现层次聚类的方法有（ ）
A. 自左向右
B. 从右到左
C. 自底向上
D. 自顶向下

7. 以下选项中属于K均值聚类方法流程中步骤的有（ ）
A. 初始化类心
B. 利用标签将样本分类
C. 按当前类心对样本归类
D. 迭代类心

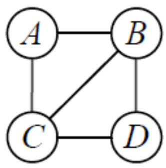
8. 以下可行的最近邻分类的加速方案（ ）
A. 分层搜索
B. 训练样本缩减
C. 样本增加
D. 非线性投影

9. Adaboost方法中，需要迭代调整的两个重要参数是(
A. 样本权重
B. 分类器权重
C. 梯度变化率
D. 梯度

10. 下面属于非线性模型的机器学习的方法()
A. 决策树
B. PCA
C. 多层感知机
D. M-P模型

11. 以下模型中属于贝叶斯网络的有()
A. 马尔可夫随机场
B. 隐马尔可夫模型
C. 条件随机场
D. 朴素贝叶斯分类器

12. 如右图所示无向图，它的团包括()
A. {A}
B. {A, B}
C. {A, B, C}
D. {A, B, C, D}



13. 同题13所示无向图，它的极大团包括()
A. {B, C, D}
B. {A, B}
C. {A, B, C}
D. {A, B, C, D}

14. 下面属于线性分类方法的是(
A. 线性回归
B. 决策树
C. 最近邻
D. Fisher鉴别

15. 下列哪些方法用于解决非线性可分问题（)
A. 线性判别器
B. 神经网络
C. 决策树
D. 最近邻分类器

16. 下面关于集成学习的描述，正确的是（
A. Bagging方法可以并行训练
B. Bagging方法基学习器的比重不同
C. Boosting方法可以并行训练
D. Boosting方法基学习器的比重不同

17. 下列选项中属于实现决策树分类方法时的常见组件有（ ）
A. 基分类器
B. 激活函数
C. 剪枝方法
D. 划分目标

18. 以下选项中属于决策树剪枝策略的有（ ）
A. 预剪枝
B. 后剪枝
C. 最大后验
D. 最小风险

19. 在机器学习领域，以下哪些算法可以通过神经网络的形式实现（
A. KNN
B. Logistic回归
C. K-means
D. 最小二乘估计

20. 下列算法属于深度学习的是（ ）
A. 卷积神经网络
B. 循环神经网络
C. 决策树
D. 受限玻尔兹曼机

21. 影响深度神经网络训练效果的因素有（ ）
A. 学习率
B. 训练集规模
C. 网络深度
D. 激活函数

22. 下列选项中属于特征降维带来的好处的是（ ）
A. 节省数据通信开销
B. 节省数据存储资源
C. 提升模型表示能力
D. 加快模型计算速度

23. 下面关于特征选择和特征提取的描述正确的是（ ）
A. Relief算法属于特征提取方法

B. 特征选择的目标是从原始的d个特征中选择k个特征
C. 特征提取的目标是根据原始的d个特征的组合形成k个新的特征
D. PCA属于特征选择方法

24. 给定两个特征向量，以下哪些方法可以计算这两个向量相似度（ ）
A. 欧式距离
B. 夹角余弦
C. 信息熵
D. 曼哈顿距离

25. 类别不平衡就是指分类问题中不同类别的训练样本相差悬殊的情况，例如正例有900个，而反例只有100个，这个时候我们就需要进行相应的处理来平衡这个问题,下列方法正确的是（ ）
A. 在训练样本较多的类别中进行欠采样
B. 在训练样本较多的类别中进行过采样
C. 直接基于原数据集进行学习，对预测值进行再缩放处理
D. 通过对反例中的数据进行插值，来产生额外的反例

26. 在机器学习中，下列关于各算法对应的损失函数正确的是（ ）
A. 最小二乘-Square loss
B. SVM-Hinge Loss
C. Logistic Regression-交叉熵损失函数
D. AdaBoost-指数损失函数

27. 以下关于正则化的描述正确的是（ ）
A. 正则化可以防止过拟合
B. L1正则化能得到稀疏解
C. L2正则化约束了解空间
D. Dropout也是一种正则化方法

28. 以下选项中可以用来降低过拟合的方法有（
A. 获取更多训练数据
B. 减少使用训练样本的量
C. 增加模型复杂度
D. 添加正则化方法

29. 以下选项中可以用来降低欠拟合的方法有（ ）
A. 获取更多训练数据
B. 添加有效的数据特征
C. 增加模型复杂度
D. 添加正则化方法

30. 以下哪些机器学习算法可以不对特征做归一化处理（
A. 随机森林　　B. 逻辑回归　　C. SVM　　D. 决策树

31. 以下哪些模型是分类模型（ ）
A.最近邻　B.K均值　C.朴素贝叶斯　D.逻辑回归

32. 在某神经网络的隐层输出中包含-1.5，那么该神经网络采用的激活函数不可能是（ ）
A. Sigmoid　B. Tanh　C. Relu　D. Leaky Relu

33. 关于集成学习正确的是（ ）
A. Bagging 降低偏差
B. Bagging 降低方差
C. Boosting 降低偏差
D. Boosting 降低方差

34. 以下选项中可以用来评价机器学习模型分类性能的指标有（ ）
A. 准确率
B. 召回率
C. F1值
D. 参数量大小

35. 最近邻分类中测度度量，经常采用范数距离，以下属于范数距离的是
A. $D(x,y)=\sum_i |x_i - y_i|$

B. $D(x,y)=\max_i |x_i - y_i|$

C. $D(x,y)=[(x-y)^T(x-y)]^{1/2}$

D. $D(x,y)=(x-y)^T\Sigma^{-1}(x-y)$

36. 在机器学习中需要划分数据集，常用的划分测试集和训练集的划分方法有哪些（ ）

A、留出法
B、交叉验证法
C、自助法
D、评分法

三、简答题

1. 试阐述LDA（线性鉴别分析）的分类思想。
答案：
通过最大化类别间距离和最小化类别内方差的方式，将数据投影到一个低维空间，

以实现有效的类别分离。

2. 请简要介绍SVM的设计思想。
答案：
寻找能够最大化类别间间隔的超平面，使得数据点距离超平面的最小距离（间隔）最大。

3. 试分析SVM 对噪声敏感的原因。
答案：
SVM追求边界的最大间隔，SVM 最优决策边界由支持向量决定。当增加噪声时，那么该噪声有可能成为一个支持向量，这意味着决策边界会被噪声影响。

4. 在数据处理时，为什么通常要进行标准化处理，请给出一种标准化处理的方法。
答案：
(1)使不同特征的数值范围一致，防止某些特征对模型的影响过大，
(2)减小训练难度，加速收敛，
(3)提高模型鲁棒性。

Z-score标准化，公式如下

$$z = \frac{x - \mu}{\sigma}$$

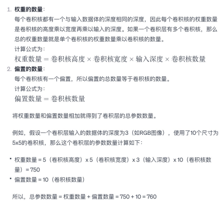
z是标准化后的值；x是原始数据的值；μ是原始数据的均值；σ是原始数据的标准差。

5. 试述将线性函数用作神经元激活函数的缺陷。
答案：
线性函数堆叠多层无法引入非线性，无论多少层的神经网络会退化成一个线性回归，不能处理非线性任务。

6. 试述学习率的取值对神经网络训练的影响。
答案：
学习率对神经网络训练的影响：
过小的学习率： 训练收敛缓慢，容易陷入局部最小值。
过大的学习率： 可能导致训练不稳定，振荡或无法收敛。
合适的学习率： 有助于快速且稳定地找到全局最小值。

7. 神经网络为什么会产生梯度消失。
答案：
传统神经网络的激活函数（sigmoid 或 tanh）是双线性饱和的，在梯度传递时，由于这类激活函数的梯度$\nabla f(x)$在[0,1]区间，梯度迭代过程中$[\nabla f(x)]^n \to 0$，使网络梯度消失。

8. 对3个32×32的特征图进行卷积层操作，卷积核10个5×5，Stride是1，pad为2，输出特征图的尺度是多少？卷积层的参数是多少？写出公式和结果。



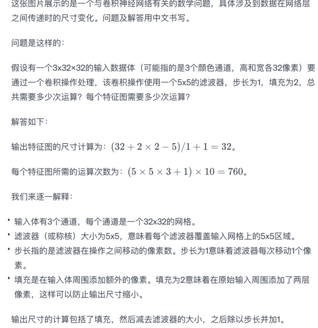
1. **权重的数量**：
每个卷积核都有一个与输入数据体的深度相同的深度，因此每个卷积核的权重数量是卷积核的高度乘以宽度再乘以输入的深度。如果一个卷积层有多个卷积核，那么总的权重数量就是单个卷积核的权重数量乘以卷积核的数量。
计算公式为：
$$权重数量 = 卷积核高度 \times 卷积核宽度 \times 输入深度 \times 卷积核数量$$
2. **偏置的数量**：
每个卷积核有一个偏置，所以偏置的总数量等于卷积核的数量。
计算公式为：
$$偏置数量 = 卷积核数量$$

将权重数量和偏置数量相加就得到了卷积层的总参数数量。

例如，假设一个卷积层输入的数据体的深度为3（如RGB图像），使用了10个尺寸为5x5的卷积核，那么这个卷积层的参数数量计算如下：

- 权重数量 = 5（卷积核高度）x 5（卷积核宽度）x 3（输入深度）x 10（卷积核数量）= 750
- 偏置数量 = 10（卷积核数量）

所以，总参数数量 = 权重数量 + 偏置数量 = 750 + 10 = 760



这张图片展示的是一个与卷积神经网络有关的数学问题，具体涉及到数据在网络层之间传递时的尺寸变化。问题及解答用中文书写。

问题是这样的：

假设有一个3x32×32的输入数据体（可能指的是3个颜色通道，高和宽各32像素）要通过一个卷积操作处理，该卷积操作使用一个5x5的滤波器，步长为1，填充为2，总共需要多少次运算？每个特征图需要多少次运算？

解答如下：

输出特征图的尺寸计算为：$(32+2\times2-5)/1+1=32$。

每个特征图所需的运算次数为：$(5\times5\times3+1)\times10=760$。

我们来逐一解释：

- 输入体有3个通道，每个通道是一个32x32的网格。
- 滤波器（或称核）大小为5x5，意味着每个滤波器覆盖输入网格上的5x5区域。
- 步长指的是滤波器在操作之间移动的像素数。步长为1意味着滤波器每次移动1个像素。
- 填充是在输入体周围添加额外的像素。填充为2意味着在原始输入周围添加了两层像素，这样可以防止输出尺寸缩小。

输出尺寸的计算包括了填充，然后减去滤波器的大小，之后除以步长并加1。

答案：
输出尺度（ 32+2×2-5）/1+1 = 32
卷积层的参数 （5×5×3+1）×10=760

9. 试析随机森林为何比决策树Bagging集成的训练速度更快。
答案：
Bagging在选择划分属性时需要考察结点的所有属性，而随机森林只需随机地考察一个属性子集，减少了每棵树的训练时所需考虑的特征数，因此更快。

10. 请给出L1范数和L2范数的计算方法及他们的使用场景。
答案：
L1范数:

$$\|\boldsymbol{x}\|_1 = \sum_i |x_i|$$

L2范数:

$$\|\boldsymbol{x}\|_2 = \sqrt{\sum_i x_i^2}$$

L1范数：推动部分特征的权重变为零，适用于特征稀疏化的场景。
L2范数：使得权重趋向均匀分布，适用于防止过拟合的场景。

11. 试述为什么基于L1范数可以进行特征选择。
答案：
L1范数惩罚项引入稀疏性,使得部分特征的权重变为零,从而实现自动特征选择。

12. 描述主成分分析的主要步骤。
答案：
（1）数据标准化。
（2）计算协方差矩阵，求协方差的特征值和特征向量。
（3）将特征值按照从大到小的顺序排序，选择其中最大的k个，然后将其对应的k个特征向量分别作为列向量组成特征向量矩阵。
（4）将样本点投影到选取的特征向量上。

13. 请描述机器学习中的分类任务。

答案：根据给定的训练集$T=\{(\mathbf{x}_1,y_1),\cdots,(\mathbf{x}_l,y_l)\}$，其中$\mathbf{x}_i \in C = R^n$ ，

$y_i \in \mathrm{Y} = \{1,2,\ldots,m\}$， $i=1,2,\ldots,l$，要求寻找C上的决策函数$g(\mathbf{x})$：$\mathrm{C} \to \mathrm{Y}$。

14. 请你从误差分解的角度，简述对泛化误差的理解。
答案：
泛化误差 = 偏差+方差+噪声；
偏差：度量了学习算法的期望预测与真实结果的偏离程度，刻画了学习算法本身

的拟合能力，
方差:度量了同样大小的训练集的变动所导致的学习性能的变化，即刻画了数据扰动所造成的影响;
噪声：表达了在当前任务上任何学习算法所能达到的期望泛化误差的下界，即刻画了学习问题本身的难度。

15.模型评估过程中，欠拟合和过拟合现象是什么。
答案:
过拟合：模型在训练集上表现良好，但在验证集上表现差，过度适应训练数据的噪声和特定样本。
欠拟合：模型未能学习训练数据的基本特征，导致训练和验证误差都较高。

16.说出几种降低过拟合和欠拟合的方法。
答案:
降低过拟合:
（1）数据增强
（2）降低模型复杂度
（3）正则化方法。
（4）集成学习方法。
降低欠拟合:
（1）进行特征工程，添加新特征。
（2）增加模型复杂度。
（3）减小正则化系数。

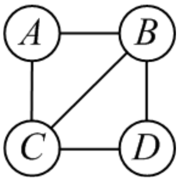
17.K均值算法的优缺点是什么，如何对其调优。
答案:
优点:
(1)对于大数据集有效
(2)简单、易于理解和实现，计算复杂度是 $O(NKt)$ 接近于线性，其中N是数据对象的数目，K是聚类的簇数，t是迭代的轮数。
缺点:
(1)对初始聚类中心的选择敏感。
(2)对异常值和噪声敏感。
(3)需要事先指定簇的数量K。
调优:
(1)使用更鲁棒的初始化方法合理选择K值。
(2)数据归一化
(3)离群点预处理。

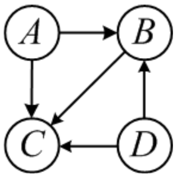
18.请简述relu激活函数的优缺点。
答案:
优点:
(1)形式简单，ReLU只需要一个阈值即可得到激活值
(2)计算高效

(3)缓解梯度消失，ReLU的非饱和性可以有效地解决梯度消失的问题。
缺点：
(1)可能存在神经元死亡问题（输出始终为零）
(2)不适用于所有情况

四、计算题

1. 试写出以下两个概率图模型联合分布的因子分解式。





答案：

左边：$P(A,B,C,D)=\frac{1}{Z}\psi_{ABC}(A,B,C)\psi_{BCD}(B,C,D)$

其中，$Z=\sum_{A,B,C,D}\psi_{ABC}(A,B,C)\psi_{BCD}(B,C,D)$

右边：$P(A,B,C,D)=P(A)P(D)P(B\mid A,D)P(C\mid A,B,D)$

分解形式：

$$P(\mathbf{x})=\frac{1}{Z}\prod_{C\in\mathbb{C}}\psi_C(\mathbf{x}_C)$$

其中，$\mathbf{x}=\{x_1,x_2,\dots,x_n\}$；$\mathbb{C}$ 为团集合，$\mathbf{x}_C$ 为团 C 对应的变量集合。

ψ_C 为定义在团 C 上的非负势函数（potential function）。

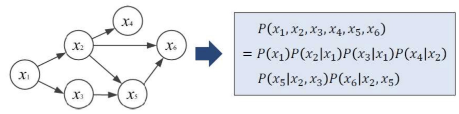
Z 是归一化因子：

$$Z=\sum_{\mathbf{x}}\prod_{C\in\mathbb{C}}\psi_C(\mathbf{x}_C)$$

分解形式：

$$P(\mathbf{x})=\prod_{i=1}^{n}P(x_i\mid\mathbf{x}_{\pi i})$$

其中，$\mathbf{x}=\{x_1,x_2,\dots,x_n\}$；$\mathbf{x}_{\pi i}$ 为 x_i 所有父结点构成的集合。



2. 回顾信封抽球问题的隐马尔可夫模型λ=($\mathbf{A}$, $\mathbf{B}$, $\boldsymbol{\pi}$)，其中

$$\pi=(0.5,0.5)\quad A=\begin{bmatrix}0 & 1\\ 0.5 & 0.5\end{bmatrix}\quad B=\begin{bmatrix}0.5 & 0.5\\ 0 & 1\end{bmatrix}$$

假设球的颜色序列为$\mathbf{x}$={x1=红, x2=黑, x3=黑, x4=黑, x5=红}，试利用前向算法和后向算法计算$P(\mathbf{x}\mid\lambda)$。

答案：

设第一个信封状态为0，第二个信封状态为1；红色状态为0，黑色为1。

（1） **前向算法**

$$\alpha_1(0)=\pi(0)\times B_{00}=0.25$$

$$\alpha_1(1)=\pi(1)\times B_{10}=0$$

$$\alpha_2(0)=B_{01}\times(\alpha_1(0)\times A_{00}+\alpha_1(1)\times A_{10})=0$$

$$\alpha_2(1)=B_{11}\times(\alpha_1(0)\times A_{01}+\alpha_1(1)\times A_{11})=0.25$$

$$\alpha_3(0)=B_{01}\times(\alpha_2(0)\times A_{00}+\alpha_2(1)\times A_{10})=0.0625$$

$$\alpha_3(1)=B_{11}\times(\alpha_2(0)\times A_{01}+\alpha_2(1)\times A_{11})=0.125$$

$$\alpha_4(0) = B_{01} \times (\alpha_3(0) \times A_{00} + \alpha_3(1) \times A_{10}) = 0.03125$$

$$\alpha_4(1) = B_{11} \times (\alpha_3(0) \times A_{01} + \alpha_3(1) \times A_{11}) = 0.125$$

$$\alpha_5(0) = B_{00} \times (\alpha_4(0) \times A_{00} + \alpha_4(1) \times A_{10}) = 0.03125$$

$$\alpha_5(1) = B_{10} \times (\alpha_4(0) \times A_{01} + \alpha_4(1) \times A_{11}) = 0$$

观测概率为：

$$\alpha_5(0) + \alpha_5(1) = 0.03125$$

（2） 后向算法

$$\beta_5(0) = 1$$

$$\beta_5(1) = 1$$

$$\beta_4(0) = A_{00} \times B_{00} \times \beta_5(0) + A_{01} \times B_{10} \times \beta_5(1) = 0$$

$$\beta_4(1) = A_{10} \times B_{00} \times \beta_5(0) + A_{11} \times B_{10} \times \beta_5(1) = 0.25$$

$$\beta_3(0) = A_{00} \times B_{01} \times \beta_4(0) + A_{01} \times B_{11} \times \beta_4(1) = 0.25$$

$$\beta_3(1) = A_{10} \times B_{01} \times \beta_4(0) + A_{11} \times B_{11} \times \beta_4(1) = 0.125$$

$$\beta_2(0) = A_{00} \times B_{01} \times \beta_3(0) + A_{01} \times B_{11} \times \beta_3(1) = 0.125$$

$$\beta_2(1) = A_{10} \times B_{01} \times \beta_3(0) + A_{11} \times B_{11} \times \beta_3(1) = 0.125$$

$$\beta_1(0) = A_{00} \times B_{01} \times \beta_2(0) + A_{01} \times B_{11} \times \beta_2(1) = 0.125$$

$$\beta_1(1) = A_{10} \times B_{01} \times \beta_2(0) + A_{11} \times B_{11} \times \beta_2(1) = 0.09375$$

观测概率为：

$$\pi(0) \times B_{00} \times \beta_1(0) + \pi(1) \times B_{10} \times \beta_1(1) = 0.03125$$

~~3~~. 在上述隐马尔可夫模型中，试用维特比算法确定最有可能的信封序列。

答案：

$$\delta_1(0) = \pi(0) \times B_{00} = 0.25$$

$$\delta_1(1) = \pi(1) \times B_{10} = 0$$

$$\delta_2(0) = \max(\delta_1(0) \times A_{00} \times B_{01}, \delta_1(1) \times A_{10} \times B_{01}) = 0$$

$$\phi_2(0) = 0,1$$

$\delta_2(1)=\max(\delta_1(0)\times A_{01}\times B_{11},\delta_1(1)\times A_{11}\times B_{11})=0.25$
$\phi_2(1)=0$

$\delta_3(0)=\max(\delta_2(0)\times A_{00}\times B_{01},\delta_2(1)\times A_{10}\times B_{01})=0.0625$
$\phi_3(0)=1$

$\delta_3(1)=\max(\delta_2(0)\times A_{01}\times B_{11},\delta_2(1)\times A_{11}\times B_{11})=0.125$
$\phi_3(1)=1$

$\delta_4(0)=\max(\delta_3(0)\times A_{00}\times B_{01},\delta_3(1)\times A_{10}\times B_{01})=0.03125$
$\phi_4(0)=1$

$\delta_4(1)=\max(\delta_3(0)\times A_{01}\times B_{11},\delta_3(1)\times A_{11}\times B_{11})=0.0625$
$\phi_4(1)=0,1$

$\delta_5(0)=\max(\delta_4(0)\times A_{00}\times B_{00},\delta_4(1)\times A_{10}\times B_{00})=0.015625$
$\phi_5(0)=1$

$\delta_5(1)=\max(\delta_4(0)\times A_{01}\times B_{10},\delta_4(1)\times A_{11}\times B_{10})=0$
$\phi_5(1)=0,1$

回溯最优路径：
0，1，0，1，0　或 0, 1，1, 1，0

4. 对于下表5个样本，分别使用ID3，C4.5，CART计算第一次节点划分时的最优特征。

表 3.1　5 个候选对象的属性以及女孩对应的主观意愿

	年龄	长相	工资	写代码	类别
小 A	老	帅	高	不会	不见
小 B	年轻	一般	中等	会	见
小 C	年轻	丑	高	不会	不见
小 D	年轻	一般	高	会	见
小 L	年轻	一般	低	不会	不见

ID3：

$$H(D)=-\frac{3}{5}\log_2\frac{3}{5}-\frac{2}{5}\log_2\frac{2}{5}=0.971,$$

$$H(D\mid 年龄)=\frac{1}{5}H(老)+\frac{4}{5}H(年轻)$$

$$=\frac{1}{5}(-0)+\frac{4}{5}\left(-\frac{2}{4}\log_2\frac{2}{4}-\frac{2}{4}\log_2\frac{2}{4}\right)=0.8,$$

$$H(D\mid 长相)=\frac{1}{5}H(帅)+\frac{3}{5}H(一般)+\frac{1}{5}H(丑)$$

$$=0+\frac{3}{5}\left(-\frac{2}{3}\log_2\frac{2}{3}-\frac{1}{3}\log_2\frac{1}{3}\right)+0=0.551,$$

$$H(D\mid 工资)=\frac{3}{5}H(高)+\frac{1}{5}H(中等)+\frac{1}{5}H(低)$$

$$=\frac{3}{5}\left(-\frac{2}{3}\log_2\frac{2}{3}-\frac{1}{3}\log_2\frac{1}{3}\right)+0+0=0.551,$$

$$H(D\mid 写代码)=\frac{3}{5}H(不会)+\frac{2}{5}H(会)$$

$$=\frac{3}{5}(0)+\frac{2}{5}(0)=0.$$

计算每个特征的信息增益：

$g(D,年龄)=0.171$, $g(D,长相)=0.42$,

$g(D,工资)=0.42$, $g(D,写代码)=0.971$.

写代码为最优划分特征。

C4.5：

$$H_{年龄}(D)=-\frac{1}{5}\log_2\frac{1}{5}-\frac{4}{5}\log_2\frac{4}{5}=0.722,$$

$$H_{长相}(D)=-\frac{1}{5}\log_2\frac{1}{5}-\frac{3}{5}\log_2\frac{3}{5}-\frac{1}{5}\log_2\frac{1}{5}=1.371,$$

$$H_{工资}(D)=-\frac{3}{5}\log_2\frac{3}{5}-\frac{1}{5}\log_2\frac{1}{5}-\frac{1}{5}\log_2\frac{1}{5}=1.371,$$

$$H_{写代码}(D)=-\frac{3}{5}\log_2\frac{3}{5}-\frac{2}{5}\log_2\frac{2}{5}=0.971.$$

信息增益比为：

$\mathbf{g_R}(\boldsymbol{D},年龄)=\frac{0.171}{0.722}=0.236$，$\mathbf{g_R}(\boldsymbol{D},长相)=\frac{0.42}{1.371}=0.306$

$\mathbf{g_R}(\boldsymbol{D},工资)=\frac{0.42}{1.371}=0.306$，$\mathbf{g_R}(\boldsymbol{D},写代码)=\frac{0.971}{0.971}=1$

CART：

$\mathrm{Gini}(D|$ 年龄 = 老 $)=0.4$，　$\mathrm{Gini}(D|$ 年龄 = 年轻 $)=0.4$，

$\mathrm{Gini}(D|$ 长相 = 帅 $)=0.4$，　$\mathrm{Gini}(D|$ 长相 = 丑 $)=0.4$，

$\mathrm{Gini}(D|$ 写代码 = 会 $)=0$，　$\mathrm{Gini}(D|$ 写代码 = 不会 $)=0$，

$\mathrm{Gini}(D|$ 工资 = 高 $)=0.47$，$\mathrm{Gini}(D|$ 工资 = 中等 $)=0.3$，

$\mathrm{Gini}(D|$ 工资 = 低 $)=0.4$.

会写代码或不会写代码为最优划分特征。

5. 抛一枚硬币问题，观察数据情况是：一枚硬币包括正反两面，共抛了30 次，其中12 次是正面，18 次是反面。采用Maximum Likelihood 方法，估计正面出现的概率和反面出现的概率。

答案：

设正面出现的概率为p，则反面出现的概率为1-p。

上述实验出现的概率为：

$$L(p)=C_{30}^{12}p^{12}(1-p)^{18}$$

对上式求偏导：

$$\frac{\partial L}{\partial p}=12C_{30}^{12}p^{11}(1-p)^{18}-18C_{30}^{12}p^{12}(1-p)^{17}$$

令偏导等于0，解得： $p=0.4$

所以，正面出现的概率为0.4，反面出现的概率为0.6。

6.给定两个数据集A和B；

$a_1=(1,2)^T,a_2=(2,3)^T,a_i\in A$

$b_1=(2,0)^T,b_2=(3,-1)^T,b_i\in B$

求A和B的Fisher最优鉴别矢量。

答案：

类内均值：

$$\mu_1=(\frac{3}{2},\frac{5}{2})^T \qquad \mu_2=(\frac{5}{2},-\frac{1}{2})^T$$

类内散度矩阵：

$$S_w=\sum_{x\in A}(x-\mu_1)(x-\mu_1)^T+\sum_{x\in B}(x-\mu_2)(x-\mu_2)^T$$

$$=\begin{bmatrix}1&0\\0&1\end{bmatrix}$$

$$S_w^{-1}=\begin{pmatrix}1&0\\0&1\end{pmatrix}$$

最优鉴别矢量：

第 k 类样本平均值（类心）：

$$\boldsymbol{u}_k=\frac{1}{|C_k|}\sum_{\boldsymbol{x}_i\in C_k}\boldsymbol{x}_i$$

两个类别的类心：

$$\boldsymbol{u}_1=\frac{1}{|C_1|}\sum_{\boldsymbol{x}_i\in C_1}\boldsymbol{x}_i,\qquad \boldsymbol{u}_2=\frac{1}{|C_2|}\sum_{\boldsymbol{x}_i\in C_2}\boldsymbol{x}_i$$

其中，类内散度矩阵：

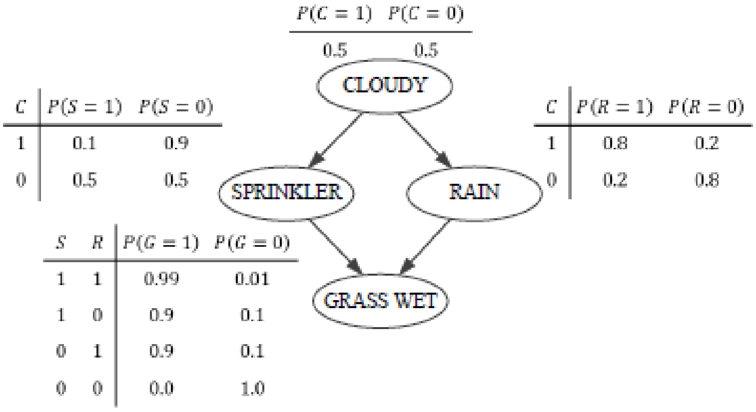
$$\boldsymbol{S}_w=\boldsymbol{X}_1\boldsymbol{X}_1^T+\boldsymbol{X}_2\boldsymbol{X}_2^T$$

若考虑先验可以定义： $\boldsymbol{S}_w=p(\omega_1)\boldsymbol{X}_1\boldsymbol{X}_1^T+p(\omega_2)\boldsymbol{X}_2\boldsymbol{X}_2^T$

$$\boldsymbol{w}=\boldsymbol{S}_W^{-1}(\boldsymbol{u}_1-\boldsymbol{u}_2)$$

$$\omega = S_w^{-1}(\mu_1 - \mu_2) = (-1,3)^T$$

7. 已知四个随机变量C、S、R、G，分别代表CLOUDY、SPRINKLER、RAIN和GRASS WET，它们之间构成的贝叶斯网络如图所示。



$P(C=1)$	$P(C=0)$
0.5	0.5

C	$P(S=1)$	$P(S=0)$
1	0.1	0.9
0	0.5	0.5

C	$P(R=1)$	$P(R=0)$
1	0.8	0.2
0	0.2	0.8

S	R	$P(G=1)$	$P(G=0)$
1	1	0.99	0.01
1	0	0.9	0.1
0	1	0.9	0.1
0	0	0.0	1.0

计算：1）在 G=1 的条件下，S=1 的概率；2）在 G=1 的条件下，R=1 的概率。

答案：

因子分解式：

$$P(C,S,R,G) = P(C)P(S|C)P(R|C)P(G|S,R)$$

（1）

$$P(S=1|G=1) = \frac{P(S=1,G=1)}{P(G=1)} = \frac{P(S=1,G=1)}{P(S=1,G=1)+P(S=0,G=1)}$$

$$P(S=1,G=1) = \sum_{C\in\{0,1\},R\in\{0,1\}} P(C,S=1,R,G=1)$$

$=0.5\times0.5\times0.8\times0.9+0.5\times0.5\times0.2\times0.99$

$+0.5\times0.1\times0.2\times0.9+0.5\times0.1\times0.8\times0.99$

$=0.2781$

$$P(B|A) = \frac{P(AB)}{P(A)}$$

$$P(S=0,G=1) = \sum_{C\in\{0,1\},R\in\{0,1\}} P(C,S=0,R,G=1)$$

$=0.5\times0.5\times0.8\times0+0.5\times0.5\times0.2\times0.9$

$+0.5\times0.9\times0.2\times0+0.5\times0.9\times0.8\times0.9$

$=0.369$

$$P(S=1|G=1) = \frac{0.2781}{0.2781+0.369} = 0.4298$$

（2）

$$P(R=1|G=1) = \frac{P(R=1,G=1)}{P(G=1)} = \frac{P(R=1,G=1)}{P(R=1,G=1)+P(R=0,G=1)}$$

$$P(R=1,G=1)=\sum_{C\in\{0,1\},S\in\{0,1\}} P(C,S,R=1,G=1)$$

$=0.5\times0.5\times0.2\times0.9+0.5\times0.5\times0.2\times0.99$

$+0.5\times0.9\times0.8\times0.9+0.5\times0.1\times0.8\times0.99$

$=0.4581$

$$P(R=0,G=1)=\sum_{C\in\{0,1\},S\in\{0,1\}} P(C,S,R=0,G=1)$$

$=0.5\times0.5\times0.8\times0+0.5\times0.5\times0.8\times0.9$

$+0.5\times0.9\times0.8\times0+0.5\times0.1\times0.2\times0.9$

$=0.189$

$$P(S=1|G=1)=\frac{0.4581}{0.4581+0.189}=0.7079$$

8. 已知 $P(\omega_1)=0.2$， $P(\omega_2)=0.8$，

$P(x=\text{阴天}|\omega_1)=0.6$， $P(x=\text{晴天}|\omega_1)=0.4$，

$P(x=\text{阴天}|\omega_2)=0.1$， $P(x=\text{晴天}|\omega_2)=0.9$

已知 $x=\text{阴天}$，求 x 所属类别。

答案：

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)}$$

$$P(\omega_1|x=\text{阴天})=\frac{p(x=\text{阴天}|\omega_1)P(\omega_1)}{p(x=\text{阴天})}$$

$$=\frac{p(x=\text{阴天}|\omega_1)P(\omega_1)}{p(x=\text{阴天}|\omega_1)P(\omega_1)+p(x=\text{阴天}|\omega_2)P(\omega_2)}$$

$$=\frac{0.6\times0.2}{0.6\times0.2+0.1\times0.8}=0.6$$

$$P(\omega_2|x=\text{阴天})=\frac{p(x=\text{阴天}|\omega_2)P(\omega_2)}{p(x=\text{阴天})}$$

$$=\frac{p(x=\text{阴天}|\omega_2)P(\omega_2)}{p(x=\text{阴天}|\omega_1)P(\omega_1)+p(x=\text{阴天}|\omega_2)P(\omega_2)}$$

$$=\frac{0.1\times0.8}{0.6\times0.2+0.1\times0.8}=0.4$$

$\therefore x\in\omega_1$

定理3(全概率公式)

定义：（完备事件组/样本空间的划分）

设B1，B2，…Bn是一组事件,若

（1）$\forall i\neq j\in\{1,2,\ldots,n\},B_inB_j=\varnothing$

（2）B1∪B2∪…∪Bn=Ω

则称B1，B2，…Bn样本空间Ω的一个划分，或称为样本空间Ω 的一个完备事件组。

定理（全概率公式）：

设事件组 $\{B_i\}$ 是样本空间Ω 的一个划分，且P（Bi）>0（i=1，2，…n）

则对任一事件A，有 $P(A)=\sum_{i=1}^{n}P(B_i)P(A|B_i)$

9. 有一种病，正常为 ω_1 ，不正常为 ω_2 ，已知：

$P(\omega_1)=0.9,P(\omega_2)=0.1$

现对某人进行检查，结果为 x，已知：

$P(x|\omega_1)=0.2,P(x|\omega_2)=0.4$

风险代价矩阵为：

$$L=\begin{bmatrix}L_{11} & L_{12}\\ L_{21} & L_{22}\end{bmatrix}=\begin{bmatrix}0 & 6\\ 1 & 0\end{bmatrix}$$

（1）用最小错误率贝叶斯决策进行判别。

（2）用最小风险贝叶斯决策进行判别。

答案

（1）

$$P(\omega_1 | x) \propto P(\omega_1)P(x | \omega_1)$$
$$P(\omega_2 | x) \propto P(\omega_2)P(x | \omega_2)$$

由于

$$\frac{P(\omega_1 | x)}{P(\omega_2 | x)} = \frac{P(\omega_1)P(x | \omega_1)}{P(\omega_2)P(x | \omega_2)} = \frac{9}{2}$$

根据贝叶斯最小错误率判决准则， $x \in \omega_1$ 。

（2）

将 X 判为第 j 类的风险为：

$$r_j(x) = \sum_{i=1}^{2} L_{ij} P(x | \omega_i) P(\omega_i), j = 1, 2$$

$$\begin{aligned} & r_1(x) - r_2(x) = L_{11}P(x | \omega_1)P(\omega_1) + L_{21}P(x | \omega_2)P(\omega_2) \\ & -L_{12}P(x | \omega_1)P(\omega_1) - L_{22}P(x | \omega_2)P(\omega_2) \\ & = P(x | \omega_1)P(\omega_1)(L_{11} - L_{12}) + P(x | \omega_2)P(\omega_2)(L_{21} - L_{22}) \end{aligned}$$

因为

$$\frac{P(x | \omega_2)P(\omega_2)(L_{21} - L_{22})}{P(x | \omega_1)P(\omega_1)(L_{12} - L_{11})} = \frac{1}{27} < 1$$

所以 $r_1(x) < r_2(x)$，根据贝叶斯最小风险决策可知 $x \in \omega_1$ 。

10. 以下为标注数据以及对应的特征，其中，A，B，C为两类特征，Y为类别标签，利用朴素贝叶斯分类器求A=0，B=1，C=1时，Y的分类标签。

A	1	0	0	1	0	1	0	0	1	1
B	0	1	1	0	1	0	0	1	0	1
C	0	0	1	0	1	1	0	1	0	0
Y	1	0	1	1	0	0	1	0	1	1

答案：

$P(A=0 | Y=0) = \frac{3}{4}$ ， $P(A=0 | Y=1) = \frac{1}{3}$

$P(B=1 | Y=0) = \frac{3}{4}$ ， $P(B=1 | Y=1) = \frac{1}{3}$

$P(C=1|Y=0)=\frac{3}{4}$， $P(C=1|Y=1)=\frac{1}{6}$

$P(Y=0)=\frac{2}{5}$ ， $P(Y=1)=\frac{3}{5}$

由贝叶斯公式得

$$P(Y=0|A=0,B=1,C=1)=\frac{P(A=0,B=1,C=1|Y=0)P(Y=0)}{P(A=0,B=1,C=1)}$$

$$=\frac{P(A=0|Y=0)P(B=1|Y=0)P(C=1|Y=0)P(Y=0)}{P(A=0,B=1,C=1)}$$

$$=\frac{\frac{3}{4}\times\frac{3}{4}\times\frac{3}{4}\times\frac{2}{5}}{P(A=0,B=1,C=1)}$$

$$=\frac{\frac{27}{160}}{P(A=0,B=1,C=1)}$$

同理

$$P(Y=1|A=0,B=1,C=1)=\frac{P(A=0,B=1,C=1|Y=1)P(Y=1)}{P(A=0,B=1,C=1)}$$

$$=\frac{P(A=0|Y=1)P(B=1|Y=1)P(C=1|Y=1)P(Y=1)}{P(A=0,B=1,C=1)}$$

$$=\frac{\frac{1}{3}\times\frac{1}{3}\times\frac{1}{6}\times\frac{3}{5}}{P(A=0,B=1,C=1)}$$

$$=\frac{\frac{1}{90}}{P(A=0,B=1,C=1)}$$

$\because P(Y=0|A=0,B=1,C=1)>P(Y=1|A=0,B=1,C=1)$

$\therefore Y=0$

11.根据下列样本，利用Fisher鉴别分析求投影方向。

序号	x_1	x_2	类别
1	5	7	1
2	4	3	2
3	7	8	2
4	8	6	2
5	3	6	1
6	2	5	1
7	6	6	1
8	9	6	2
9	5	4	2

答案：

第一类样本：$\{(5,7)^T,(3,6)^T,(2,5)^T,(6,6)^T\}$

第二类样本：$\{(4,3)^T,(7,8)^T,(8,6)^T,(9,6)^T,(5,4)^T\}$

$$\mu_1=(4,6)^T \quad \mu_2=(6.6,5.4)^T$$

$$S_w=\sum_{x\in C_1}(x-\mu_1)(x-\mu_1)^T+\sum_{x\in C_2}(x-\mu_2)(x-\mu_2)^T$$

$$=\begin{pmatrix}10 & 3\\ 3 & 2\end{pmatrix}+\begin{pmatrix}17.2 & 11.8\\ 11.8 & 15.2\end{pmatrix}=\begin{pmatrix}27.2 & 14.8\\ 14.8 & 17.2\end{pmatrix}$$

$$S_w^{-1}=\begin{pmatrix}0.069 & -0.059\\ -0.059 & 0.109\end{pmatrix}$$

$$\omega=S_w^{-1}(\mu_1-\mu_2)=(-0.215,0.220)^T$$

12.使用k-means算法，给出下列数据的聚类结果。

点	x_1	x_2
P1	0	1
P2	1	2
P3	2	2
P4	8	8
P5	9	10
P6	10	10

注：初始化聚类中心为P1和P2。

答案：

第一轮：

{P1}，{P2}

{P1}，{P2，P3}

{P1}，{P2，P3，P4}

{P1}，{P2，P3，P4，P5}

{P1}，{P2，P3，P4，P5，P6}

新的质心：(0，1)，(6，6.4)

第二轮：

{P1}，{}

{P1，P2}，{}

{P1，P2，P3}，{}

{P1，P2，P3}，{P4}

{P1，P2，P3}，{P4，P5}

{P1，P2，P3}，{P4，P5，P6}

新的质心：(1，5/3)，(9，28/3)

第三轮：

{P1}，{}

{P1, P2}, {}
{P1, P2, P3}, {}
{P1, P2, P3}, {P4}
{P1, P2, P3}, {P4, P5}
{P1, P2, P3}, {P4, P5, P6}
新的质心：(1, 5/3), (9, 28/3)
质心不再改变，得出最终的聚类结果：
{P1, P2, P3}, {P4, P5, P6}

13. 使用自底向上层次聚类，给出下列数据的聚类结果，簇之间的相似度采用簇质心的距离。

点	x_1	x_2
P1	0	1
P2	1	2
P3	2	2
P4	8	8
P5	9	10
P6	10	10

答案：
开始每一个点为一类：
{P1}, {P2}, {P3}, {P4}, {P5}, {P6}
对应的聚类质心坐标为：
（0,1），（1,2），（2,2），（8,8），（9,10），（10,10）
经过计算，{P2}与{P3}之间的距离最小，进行合并：
{P1}，{P2，P3}，{P4}，{P5}，{P6}
对应的聚类质心坐标为：
（0,1），（3/2,2），（8,8），（9,10），（10,10）
经过计算{P5}，{P6}之间的距离最小，进行合并：
{P1}，{P2，P3}，{P4}，{P5，P6}
对应的聚类质心坐标为：
（0,1），（3/2,2），（8,8），（19/2,10）
经过计算{P1}，{P2，P3}之间的距离最小，进行合并：
{P1，{P2，P3}}，{P4}，{P5，P6}
对应的聚类质心坐标为：
（1,5/3），（8,8），（19/2,10）
经过计算{P4}，{P5，P6}之间的距离最小，进行合并：
{P1，{P2，P3}}，{P4，{P5，P6}}
最后两个集合进行合并，得到最终的聚类结果：
{{P1，{P2，P3}}，{P4，{P5，P6}}}