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)
   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(C)
   A. 支持向量机
   B. Logistic回归
   C. 层次聚类
   D. 决策树

3. 二项式分布的共轭分布是(C)
   A. 正态分布
   B. Dirichlet分布
   C. Beta分布
   D. 指数分布

4. 多项式分布的共轭分布是(B)
   A. 正态分布
   B. Dirichlet分布
   C. Beta分布
   D. 指数分布

5. 朴素贝叶斯分类器的特点是(C)
   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

6. 对于正态密度的贝叶斯分类器，各类协方差矩阵相同时，决策函数为(A)
   A. 线性决策函数
   B. 非线性决策函数
   C. 最小距离分类器
   D. 以上都有可能

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

8. 下列属于非线性分类方法的是(D)
   A. 最小距离分类器
   B. 线性鉴别分析
   C. 感知机
   D. 核SVM

9. 下列分类方法中不会用到梯度下降法的是(C)
   A. 感知机
   B. 最小二乘分类器
   C. 最小距离分类器
   D. Logistic回归

10. 下列方法使用最大似然估计的是(C)
   A. 线性鉴别分析
   B. 感知机
   C. Logistic回归
   D. SVM

11. 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(B)
   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

12. SVM的原理的简单描述，可概括为(C)
   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

13. SVM的算法性能取决于(D)
   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

14. 支持向量机的对偶问题是(C)
   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

15. 以下对支持向量机中的支撑向量描述正确的是(C)

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

16. 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(A)
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

17. 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)
A. 核函数方法
B. 集成学习
C. 线性鉴别分析
D. Logistic回归

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

19. 以下描述中，属于决策树策略的是(D)
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

20. 集成学习中基分类器的选择如何，学习效率通常越好(D)
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

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

22. 下面属于Bagging方法的特点是(A)

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

23. 下面属于Boosting方法的特点是(D)
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

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

25. 假定有一个数据集S，但该数据集有很多误差，采用软间隔SVM训练，阈值为C，如果C的值很小，以下那种说法正确(A)
A. 会发生误分类现象
B. 数据将被正确分类
C. 不确定
D. 以上都不对

26. 软间隔SVM的阈值趋于无穷，下面哪种说法正确(A)
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

27. 一般，K-NN最近邻方法在什么情况下效果好(B)
A. 样本较多但典型性不好
B. 样本较少但典型性较好
C. 样本呈团状分布
D. 样本呈链状分布
注：最近邻属于分类算法，样本多而且典型性不好容易造成分类错误（尤其是在分类边界上的样本点）。样本分布对聚类算法的影响较大。

28. 回归问题和分类问题的区别(A)
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

29. 最小二乘回归方法的等效回归方法(D)
   A. Logistic回归
   B. 多项式回归
   C. 非线性基函数回归
   D. 线性均值和正态误差的最大似然回归

30. 正则化的回归分析，可以避免(B)
   A. 线性化
   B. 过拟合
   C. 欠拟合
   D. 连续值逼近

31. “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)
   A. 关联分析
   B. 回归
   C. 聚类
   D. 分类

32. KL散度是根据什么构造的可分性判据(C)
   A. 最小损失准则
   B. 后验概率
   C. 类概率密度
   D. 几何距离

33. 密度聚类方法充分考虑了样本间的什么关系(C)
   A. 范数距离
   B. 集合运算
   C. 密度可达
   D. 样本与集合运算

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

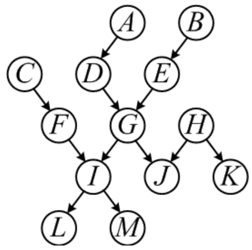
35. 主成分分析方法是一种什么方法(C)
   A. 分类方法
   B. 回归方法
   C. 降维方法
   D. 参数估计方法

36. PCA在做降维处理时，优先选取哪些特征(A)

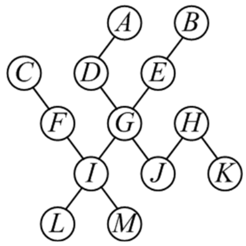
A. 中心化样本的协方差矩阵的最大特征值对应特征向量
B. 最大间隔投影方向
C. 最小类内聚类
D. 最速梯度方向

37. 过拟合现象中(A)
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

38. 如右图所示有向图，节点G的马尔可夫毯为(D)
A. {D,E}
B. {I,J}
C. {D,E,I,J}
D. {D,E,F,H,I,J}



39. 如右图所示无向图，节点G 的马尔可夫毯为(C)
A. {D,E}
B. {I,J}
C. {D,E,I,J}
D. {D,E,F,H,I,J}



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

41. 在有限支撑集上，下面分布的熵最大(D)
A. 几何分布
B. 指数分布
C. 高斯分布
D. 均匀分布

42. 已知均值和方差，下面哪种分布的熵最大(C)
A. 几何分布
B. 指数分布
C. 高斯分布
D. 均匀分布

43. 以下模型中属于概率图模型的是(D)

A. 决策树
B. 感知机
C. 支持向量机
D. 受限玻尔兹曼机

44. 在标准化公式$z_{norm}^{(i)}=\frac{z^{(i)}-\mu}{\sqrt{\sigma^2+\varepsilon}}$中，使用$\varepsilon$的目的是(D)

A. 为了加速收敛
B. 如果μ过小
C. 使结果更准确
D. 防止分母为零

45. 梯度下降算法的正确步骤是什么(B)
(1)计算预测值和真实值之间的误差
(2)迭代跟新，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

46. 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（C）
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

49. 以下哪种方法会增加模型的欠拟合风险（D）
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

50. 以下说法正确的是（C）
A. 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
B. 梯度下降有时会陷于局部极小值，但EM算法不会
C. 除了EM算法，梯度下降也可求混合高斯模型的参数

D. 基于最小二乘的线性回归问题中，增加L2正则项，总能降低在测试集上的MSE误差

51. 在训练神经网络时，如果出现训练error过高，下列哪种方法不能大幅度降低训练error（D）
A. 增加一个隐藏层
B. 在隐藏层中增加更多神经元
C. 对训练数据进行标准化
D. 增加训练数据

52. 以下哪种激活函数可以导致梯度消失（B）
A. ReLU
B. Tanh
C. Leaky ReLU
D. 其他都不是

53. 增加以下哪些超参数可能导致随机森林模型过拟合数据（B）
(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)

54. 以下关于深度网络训练的说法正确的是（D）
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

55. 以下哪一项在神经网络中引入了非线性（B）
A．Dropout
B．ReLU
C．卷积函数
D．随机梯度下降

56. 在线性回归中使用正则项，你发现解的不少coefficient都是0，则这个正则项可能是（A）
(1). L0-norm；(2). L1-norm；(3). L2-norm。
A. (1)(2)
B. (2)(3)
C. (2)
D. (3)

57. 关于CNN，以下结论正确的是（C）

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

58. 关于k-means算法，正确的描述是（B）
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

59. 下列关于过拟合现象的描述中，哪个是正确的（A）
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

60. 以下关于卷积神经网络，说法正确的是（C）
A. 卷积神经网络只能有一个卷积核
B. 卷积神经网络可以有多个卷积核，但是必须同大小
C. 卷积神经网络可以有多个卷积核，可以不同大小
D. 卷积神经网络不能使用在文本这种序列数据中

61. LR模型的损失函数是（A）
A. 交叉熵
B. 均方误差
C. Hinge loss
D. 分类准确率

62. GRU和LSTM的说法正确的是（D）
A. GRU通过output gate控制memory;
B. LSTM对memory不做控制，直接传递给下一个unit
C. GRU不对上一时刻的信息做任何控制;
D. GRU的参数比LSTM的参数少;

63. 以下方法不可以用于特征降维的有（D）
A. Linear Discriminant Analysis
B. Principal Component Analysis
C. Singular Value Decomposition
D. Monte Carlo method

64. 下列哪个函数不可以做激活函数（D）
A. y=tanh(x)

B. y=sin(x)
C. y=max(x,0)
D. y=2x

65. 有两个样本点，第一个点为正样本,它的特征向量是(0,-1);第二个点为负样本,它的特征向量是(2,3),从这两个样本点组成的训练集构建一个线性SVM分类器的分类面方程是（C）
A. 2x+y=4
B. x+2y=5
C. x+2y=3
D. 以上都不对

66. 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（D）
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

67. 下方法中属于无监督学习算法的是（D）
A. 线性回归
B. 支持向量机
C. 决策树
D. K-Means聚类

68. Bootstrap数据是什么意思（C）
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

69. 下面关于贝叶斯分类器描述错误的是（B）
A. 以贝叶斯定理为基础
B. 是基于后验概率，推导出先验概率
C. 可以解决有监督学习的问题
D. 可以用极大似然估计法解贝叶斯分类器

70. 下面关于Adaboost算法的描述中，错误的是（D）
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

71. 以下机器学习中，在数据预处理时，不需要考虑归一化处理的是（C）
A. 逻辑回归
B. 支持向量机
C. 树形模型
D. 神经网络

72. 二分类任务中，有三个分类器h1,h2,h3，三个测试样本x1,x2,x3。假设1表示分类结果正确，0表示错误，h1在x1,x2,x3的结果分别(1,1,0)，h2,h3分别为(0,1,1)，(1,0,1)，按投票法集成三个分类器，下列说法正确的是（A）
A. 集成提高了性能
B. 集成没有效果
C. 集成降低了性能
D. 集成效果不能确定

73. 有关机器学习分类算法的Precision和Recall，以下定义中正确的是（假定tp = true positive, tn = true negative, fp = false positive, fn = false negative）（A）
A. Precision= tp / (tp + fp), Recall = tp / (tp + fn)
B. Precision = tp / (tn + fp), Recall = tp /(tp + fn)
C. Precision = tp / (tn + fn), Recall = tp /(tp + fp)
D. Precision = tp / (tp + fp), Recall = tp /(tn + fn)

74. 下列哪个不属于常用的文本分类的特征选择算法（D）
A. 卡方检验值
B. 互信息
C. 信息增益
D. 主成分分析

75. 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（D）
A．EM算法
B. 维特比算法
C. 前向后向算法
D. 极大似然估计

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

77. 解决隐马模型中预测问题的算法是（D）
A. 前向算法

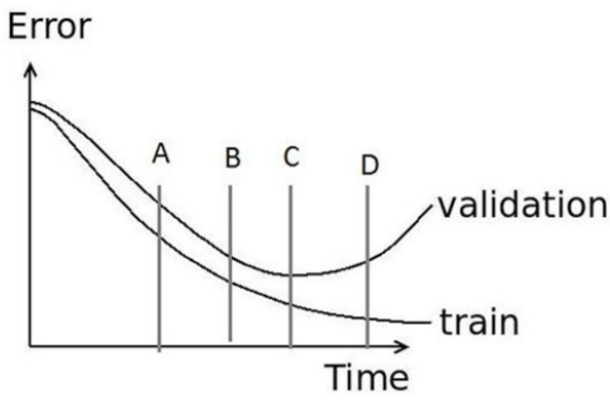
B. 后向算法
C. Baum-Welch算法
D. 维特比算法

78. 在Logistic Regression 中,如果同时加入L1和L2范数,会产生什么效果（A）
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. 可以获得更准确的结果

79. 普通反向传播算法和随时间的反向传播算法（BPTT）有什么技术上的不同 (B)
A. 与普通反向传播不同的是，BPTT会在每个时间步长内减去所有对应权重的梯度
B. 与普通反向传播不同的是，BPTT会在每个时间步长内叠加所有对应权重的梯度
C. BPTT使用的是二阶梯度
D. 没有差别

80. 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)
A. 梯度裁剪　　B. 所有方法都不行
C. Dropout　　D. 加入正则项

81. 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(C)



A. A　B. B　C. C　D. D

二、多选题

1. 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(ABC)

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

2. 对单层感知机判别分类，描述正确的是(ABC)
   A. 线性分类
   B. 监督学习
   C. 错误误差最小
   D. 错误误差最大

3. 可用于贝叶斯决策的函数(ABC)

   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. $p(\omega_1|x)$

4. 特征选择是重要的特征降维手段之一，以下描述中属于特征选择情况是(ABC)
   A. 提取向量偶数位特征，重新表示向量
   B. 前向序贯搜索子集，进行子集评价，选择最优子集
   C. 采用可分性度量，度量每个特征，进行选择
   D. 主成分分析降维

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

6. 以下属于聚类方法的是(ABD)
   A. k-means

B. 层次聚类
C. Fisher鉴别
D. 密度聚类

7. 以下可用于聚类性能测量的评估方法(ABCD)
A. Jaccard系数
B. FM指数
C. Rand指数
D. DB指数

8. 以下可行的最近邻分类的加速方案(AB)
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(AB)
A. 样本权重
B. 分类器权重
C. 梯度变化率
D. 梯度

10. 以下对层次聚类描述正确的(BD)
A. 监督学习
B. 自顶向下寻找最优划分
C. 集成学习
D. 自底向上寻找最优合并

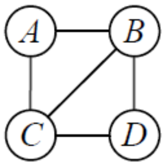
11. 支持向量机可能解决的问题(ABC)
A. 线性分类
B. 非线性分类
C. 回归分析
D. BP算法

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

13. 信息熵是信息论中重要的信息度量，以下不正确的是(BD)
A. 可度量不确定程度
B. 是运算中的商
C. 可度量信息量
D. 是向量的模

14. 以下模型中属于贝叶斯网络的有(BD)
   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

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



16. 同题15所示无向图，它的极大团包括(AC)
   A. {B,C,D}
   B. {A,B}
   C. {A,B,C}
   D. {A,B,C,D}

17. 若A为假命题，B为真命题，以下命题中假命题有(B)
   A. $\neg A$
   B. $A \cap B$
   C. $A \cup B$
   D. $B \leftarrow A$

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

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

20. 下面关于集成学习的描述，正确的是(AD)
   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

21. 下面关于最大熵模型的描述，正确的是(ABC)
A. 思想是在满足一定约束条件下，概率分布的熵最大
B. 是一种信息论模型
C. 在已知均值和方差的条件下，最大熵分布是高斯分布
D. 在已知均值和方差的条件下，最大熵分布是指数分布

22. 如果SVM模型欠拟合, 以下方法哪些可以改进模型(AD)
A. 增大惩罚参数C的值
B. 减小惩罚参数C的值
C. 减小核系数(gamma参数)
D. 增大核系数(gamma参数)

23. 在大数据集上训练决策树, 为了使用较少时间, 我们可以(C)
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

24. 以下哪些算法, 可以用神经网络去构造(BD)
A. KNN
B. Logistic回归
C. 决策树
D. 最小二乘估计

25. 对于PCA说法正确的是(ABD)
A. 必须在使用PCA前规范化数据
B. 应该选择使得模型有最大variance的主成分
C. 应该选择使得模型有最小variance的主成分
D. 可以使用PCA在低维度上做数据可视化

26. 下列算法属于深度学习的是(ABD)
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

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

28. 下面关于特征选择和特征提取的描述正确的是(BC)

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

29. 假设我们有一个使用ReLU激活函数(ReLU activation function)的神经网络，假如我们把ReLU激活替换为线性激活，那么这个神经网络能够模拟出同或函数(XNOR function)吗(B)

A. 可以

B. 不能

C. 不好说

D. 不一定

30. 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（ABD）

A. 欧式距离

B. 夹角余弦

C. 信息熵

D. 曼哈顿距离

31. 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（ACD）

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

32. 在机器学习中，下列关于各算法对应的损失函数正确的是（ABCD）

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

33. 以下关于正则化的描述正确的是（ABCD）

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

34. 以下可以有效解决过拟合的方法是（ABD）

A. 增加样本数量

B. 通过特征选择减少特征数量

C. 训练更多的迭代次数

D.　采用正则化方法

35.　在分类问题中,我们经常会遇到正负样本数据量不等的情况,比如正样本为10w条数据,负样本只有1w条数据,以下最合适的处理方法是（ACD）
A.　将负样本重复10次，生成10w样本量，乱顺序参与分类
B.　直接进行分类，可以最大限度利用数据
C.　从10w正样本中随机抽取1w参与分类
D.　将负样本每个权重设置为10，正样本权重为1，参与训练过程

三、简答题
1.试阐述LDA（线性鉴别分析）的分类思想。
答案：给定训练样例集，设法将样例投影到一条直线上，使得同类样例的投影点尽可能接近，异类样例的投影点尽可能远离；在对新样本进行分类时，将其投影到同样的这条直线上，再根据投影点的位置来判断新样本的类别。

2.试分析SVM 对噪声敏感的原因。
答案：给定训练集，SVM 最优决策边界由支持向量决定。当增加噪声时，那么该噪声有极高的可能是含噪声训练集的一个支持向量，这意味着决策边界需要变。

3.假设数据挖掘的任务是将8 个点聚类成3 个簇，A1(2,10),A2(2,5),A3(8,4),B1(5,8),B2(7,5),B3(6,4),C1(1,2),C2(4,9),距离函数是欧几里得距离。假设初始选择A1，B1，C1 分别作为每个聚类的中心，用k—均值算法来给出：
（1）第一次循环执行后的三个聚类中心；
（2）最后的三个簇。
答案：
（1）第一轮
A1(2,10)
B1（5,8）,A3（8,4）,B2（7,5）,B3（6,4）,C2（4,9）
C1（1,2）,A2（2,5）
对应中心分别是（2,10),（6,6）,(1.5,3.5)
（2）最后三个簇 {A1(2,10),B1（5,8）,C2（4,9）} {A3（8,4）,B2（7,5）,B3（6,4）} {C1（1,2）,A2（2,5）}

4.写出距离函数的四个基本性质。
答案：

（1）非负性：$dist(x_i,x_j)\geq 0$

（2）同一性：$dist(x_i,x_j)=0$ 当且仅当 $x_i=x_j$

（3）对称性：$dist(x_i,x_j)=dist(x_j,x_i)$

（4）直递性：$dist(x_i,x_j)\leq dist(x_i,x_k)+dist(x_k,x_j)$

5.在数据处理时,为什么通常要进行标准化处理。

答案：在实际问题中，我们使用的样本通常是多维数据，每一维对应一个特征，这些特征的量纲和数量级都是不一样的，这时需要对数据进行标准化处理，是所有的特征具有同样的尺度。

6.随机变量X的支撑集(也就是非零值域)定义为[a,b],没有别的限制加在X上,该随机变量的最大熵分布是什么。

答案：最大熵分布为在[a,b]区间上的均匀分布。根据最大熵模型，推导出X 概率密度函数是一个常函数，所以最大熵分布为均匀分布。

7. 随机变量X的给定均值和方差限制在X上,该随机变量的最大熵分布是什么。

答案：根据最大熵模型推导出X 概率密度函数是一个高斯分布。

8.试述将线性函数用作神经元激活函数的缺陷。

答案：如果单用线性函数作为激活函数，无论多少层的神经网络会退化成一个线性回归，不能处理非线性分类任务。

9.试述学习率的取值对神经网络训练的影响。

答案：如果学习率太低，每次下降的很慢，使得迭代次数非常多。如果学习率太高，在后面迭代时会出现震荡现在，在最小值附近来回波动。

10.神经网络为什么会产生梯度消失，有什么解决方案。

答案：前面层上的梯度是来自于后面层上梯度的乘积。当存在过多的层次时，且激活函数的梯度小于1 时，就会使前面层的梯度变得很小，更新速度过慢，导致梯度消失。

一种解决方案是使用Relu 激活函数替换sigmoid，relu 函数的梯度不会随着x的增大而变小，sigmoid 在x 取值较大时梯度趋近于0。

11.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。

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

12.试析随机森林为何比决策树Bagging集成的训练速度更快。

答案：随机森林是Bagging算法的一个扩展变体，以决策树为基学习器构建Bagging集成，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，所以随机森林比决策树Bagging训练速度更快，泛化能力越强。

13. 试述为什么基于L1范数可以进行特征选择。

答案：基于L1范数的特征选择：不能直接设置最终选择特征的个数k；通过设置正则化系数λ来隐式控制k；λ值越大，模型越关注稀疏性，得到的非零系数个数越少；反之，非零稀疏个数越多；可以设置一个选择特征个数的上限，通过设置不同λ值，得到满足要求的特征。

从有条件极值问题的角度来看，L1范数相当于将模型界空间限制在了L1-ball上，目标函数的等高线有很大的概率与坐标轴和边相交，这样的解具有稀疏性。

14. 请指出数据聚类存在哪些挑战性问题。

答案：（1）能够处理高维数据：在高维空间聚类更具挑战性，随着维数的增加，具有相同距离的两个样本其相似程度可以相差很远。对于高维稀疏数据，这一点更突出。

（2）对噪声鲁棒：在实际中，绝大多数样本集都包含噪声、空缺、部分未知属性、孤立点、甚至错误数据。

（3）具有约束的聚类：在实际应用中，通常需要在某种约束条件下进行聚类，既满足约束条件，以希望有高聚类精度，是一个挑战性问题。

（4）对初始输入参数鲁棒：具有自适应的簇数判定能力，对初始聚类中心鲁棒。

（5）能够解决用户的问题：聚类结果能被用户所理解，并能带来经济效益，特别是在数据挖掘领域。

15.写出AdaBoost算法流程。

答案：

输入： 训练集 $D=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\ldots,(\boldsymbol{x}_m,y_m)\}$;
基学习算法 $\mathfrak{L}$;
训练轮数 T.

过程：

$\mathcal{D}_1(\boldsymbol{x})=1/m.$

for $t=1,2,\ldots,T$ **do**

$h_t=\mathfrak{L}(D,\mathcal{D}_t)$;

$\epsilon_t=P_{\boldsymbol{x}\sim\mathcal{D}_t}(h_t(\boldsymbol{x})\neq f(\boldsymbol{x}))$;

if $\epsilon_t>0.5$ **then break**

$\alpha_t=\frac{1}{2}\ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right)$;

$$\mathcal{D}_{t+1}(\boldsymbol{x})=\frac{\mathcal{D}_t(\boldsymbol{x})}{Z_t}\times\begin{cases}\exp(-\alpha_t), & \text{if } h_t(\boldsymbol{x})=f(\boldsymbol{x})\\ \exp(\alpha_t), & \text{if } h_t(\boldsymbol{x})\neq f(\boldsymbol{x})\end{cases}$$
$$=\frac{\mathcal{D}_t(\boldsymbol{x})\exp(-\alpha_t f(\boldsymbol{x})h_t(\boldsymbol{x}))}{Z_t}$$

end for

输出： $H(\boldsymbol{x})=\operatorname{sign}\left(\sum_{t=1}^{T}\alpha_t h_t(\boldsymbol{x})\right)$

16.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。

17.描述一下分类问题。

答案：根据给定的训练集$T=\{(\mathbf{x}_1,y_1),\cdots,(\mathbf{x}_l,y_l)\}$，其中$\mathbf{x}_i\in C=R^n$ ，$y_i\in \mathrm{Y}=\{1,2,...,m\}$，$i=1,2,...,l$，要求寻找C上的决策函数$g(\mathbf{x})$：$\mathrm{C}\to\mathrm{Y}$。

18. 阐述一下对泛化误差的理解。

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。

四、计算题

1. 设随机变量$X=(X_1,X_2)$的协方差矩阵$\Sigma=\begin{bmatrix}2&1\\1&2\end{bmatrix}$，试求特征值和特征向量，并写出投影矩阵。

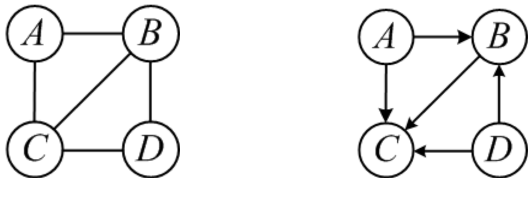
答案：求得特征根为$\lambda_1=3,\lambda_2=1$ 。

λ_1对应的特征向量为$v_1=\left(\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2}\right)^T$ ，λ_2对应的特征向量为$v_2=\left(\frac{\sqrt{2}}{2},-\frac{\sqrt{2}}{2}\right)^T$。

进行PCA的投影矩阵为$[v_1,v_2]$即

$$W=\begin{bmatrix}\frac{\sqrt{2}}{2}&\frac{\sqrt{2}}{2}\\\frac{\sqrt{2}}{2}&-\frac{\sqrt{2}}{2}\end{bmatrix}$$

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

3. 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$P(\mathbf{x}|\lambda)$。

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：

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

4. 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

5.给定一类样本x_1,x_2,x_3，写出类别均值、类内散度矩阵。

$$x_1=(1,3,5,8,10,12)^T$$
$$x_2=(0,1,5,9,2,1)^T$$
$$x_3=(4,6,8,2,3,4)^T$$

答案：

类别均值： $\bar{x}=\frac{1}{3}\sum_{i=1}^{3}x_i=(\frac{5}{3},\frac{10}{3},6,\frac{19}{3},5,\frac{17}{3})^T$

类内散度矩阵：

$$S_w=\sum_{i=1}^{3}(x_i-\bar{x})(x_i-\bar{x})^T$$

$$=\begin{bmatrix} \frac{26}{3} & \frac{31}{3} & 7 & -\frac{47}{3} & -3 & -\frac{1}{3} \\ \frac{31}{3} & \frac{38}{3} & 8 & -\frac{55}{3} & 0 & \frac{13}{3} \\ 7 & 8 & 6 & -13 & -6 & -5 \\ -\frac{47}{3} & -\frac{55}{3} & -13 & \frac{86}{3} & 9 & \frac{16}{3} \\ -3 & 0 & -6 & 9 & 38 & 49 \\ -\frac{1}{3} & \frac{13}{3} & -5 & \frac{16}{3} & 49 & \frac{194}{3} \end{bmatrix}$$

6. 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。

7. 给定两类数据集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：

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

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

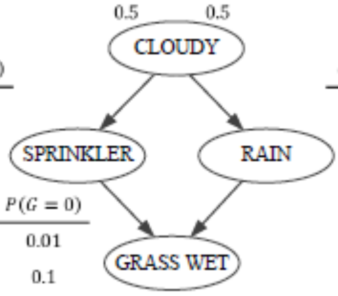
$$S_w^{-1}=\begin{pmatrix}1&0\\0&1\end{pmatrix}$$

最优鉴别矢量：

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

8. 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：

$$P(C,S,R,G)=P(C)P(S\mid C)P(R\mid C)P(G\mid S,R)$$

（1）

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

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

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

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

$$=0.2781$$

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

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

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

$$=0.369$$

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

（2）

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

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

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

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

$$=0.4581$$

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

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

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

$$=0.189$$

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

9．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：

$$
\begin{aligned}
P(\omega_1 \mid x=\text{阴天}) &= \frac{p(x=\text{阴天} \mid \omega_1)P(\omega_1)}{p(x=\text{阴天})} \\
&= \frac{p(x=\text{阴天} \mid \omega_1)P(\omega_1)}{p(x=\text{阴天} \mid \omega_1)P(\omega_1)+p(x=\text{阴天} \mid \omega_2)P(\omega_2)} \\
&= \frac{0.6\times 0.2}{0.6\times 0.2+0.1\times 0.8} = 0.6 \\
P(\omega_2 \mid x=\text{阴天}) &= \frac{p(x=\text{阴天} \mid \omega_2)P(\omega_2)}{p(x=\text{阴天})} \\
&= \frac{p(x=\text{阴天} \mid \omega_2)P(\omega_2)}{p(x=\text{阴天} \mid \omega_1)P(\omega_1)+p(x=\text{阴天} \mid \omega_2)P(\omega_2)} \\
&= \frac{0.1\times 0.8}{0.6\times 0.2+0.1\times 0.8} = 0.4
\end{aligned}
$$

$\therefore x \in \omega_1$

10. 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：

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

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）

$$
\begin{aligned}
P(\omega_1 \mid x) \propto P(\omega_1)P(x \mid \omega_1) \\
P(\omega_2 \mid x) \propto P(\omega_2)P(x \mid \omega_2)
\end{aligned}
$$

由于

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

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：

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

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

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$x\in\omega_1$。

11. 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

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

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

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

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

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

$\therefore Y=0$

12.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：$\{(4,3)^T,(7,8)^T,(8,6)^T,(9,6)^T,(5,4)^T\}$

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

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

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

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

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

13.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}

14. 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}}}

15.分别使用感知机算法、最小距离分离器、最小平方准则、Fisher鉴别分析，求决策函数。

数据：正类4个样本{(1,1)，(1,2)，(2,2)，(1,0)}；

负类4个样本{(4,5)，(5,6)，(6,7)，(6,6)}；

答案：见Chapter 3 Linear Classifiers习题答案。