

机器学习与深度学习自测题

一、单选题 (简单共 122 题 中等共 98 题 困难共 86 题 一共 306 题)

简单

1. 关于 Logit 回归和 SVM 不正确的是 (A)。

A. Logit 回归目标函数是最小化后验概率

B. Logit 回归可以用于预测事件发生概率的大小

C. SVM 目标是结构风险最小化

D. SVM 可以有效避免模型过拟合

2. 是否能用神经网络算法设计逻辑回归算法? A

A、是

B、否

3. 关于支持向量机 SVM,下列说法错误的是 (C)

A.L2 正则项,作用是最大化分类间隔,使得分类器拥有更强的泛化能力

B.Hinge 损失函数,作用是最小化经验分类错误

C.分类间隔为, $\|w\|$ 代表向量的模

D.当参数 C 越小时,分类间隔越大,分类错误越多,趋于欠学习

4. 决策树的父节点和子节点的熵的大小关系是什么? B

- A. 决策树的父节点更大
- B. 子节点的熵更大
- C. 两者相等
- D. 根据具体情况而定

5. 以下哪种方法属于判别式模型(discriminative model) (D)

- A、隐马模型(HMM)
- B、朴素贝叶斯
- C、LDA
- D、支持向量机

6. 下面关于 ID3 算法中说法错误的是 (D)

- A、ID3 算法要求特征必须离散化
- B、信息增益可以用熵，而不是 GINI 系数来计算
- C、选取信息增益最大的特征，作为树的根节点
- D、ID3 算法是一个二叉树模型

7. 以下属于欧式距离特性的有 (A)

- A、旋转不变性
- B、尺度缩放不变性
- C、不受量纲影响的特性

8. 如果以特征向量的相关系数作为模式相似性测度，则影响聚类算法结果的主要因素有

(B)

A、已知类别样本质量

B、分类准则

C、量纲

9. 一监狱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 4 种不同的人员：狱警，小偷，送餐员，其他。下面哪种学习方法最适合此种应用需求： B

A、二分类问题

B、多分类问题

C、层次聚类问题

D、k-中心点聚类问题

E、回归问题

F、结构分析问题

10. 以下说法中错误的是 (C)

A、SVM 对噪声（如来自其他分部的噪声样本）具备鲁棒性

B、在 adaboost 算法中，所有被分错样本的权重更新比例不相同

C、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确定其权重

D、给定 n 个数据点，如果其中一半用于训练，一半用户测试，则训练误差和测试误差之间的差别会随着 n 的增加而减少的

11. 位势函数法的积累势函数 $K(x)$ 的作用相当于 Bayes 判决中的 (A)

- A、后验概率
- B、先验概率
- C、类概率密度
- D、类概率密度与先验概率的和

12. 下列哪些不特别适合用来对高维数据进行降维 C

- A、LASSO
- B、主成分分析法
- C、聚类分析
- D、小波分析法
- E、线性判别法
- F、拉普拉斯特征映射

13. 对数几率回归 (logistics regression) 和一般回归分析有什么区别? D

- A、对数几率回归是设计用来预测事件可能性的
- B、对数几率回归可以用来度量模型拟合程度
- C、对数几率回归可以用来估计回归系数
- D、以上所有

14. 在一个 n 维的空间中, 最好的检测 outlier(离群点)的方法是 (C)

A、作正态分布概率图

B、作盒形图

C、马氏距离

D、作散点图

15. 一般, k-NN 最近邻方法在 (B) 的情况下效果较好

A、样本较多但典型性不好

B、样本较少但典型性好

C、样本呈团状分布

D、样本呈链状分布

16. 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 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

17. 符号集 a、b、c、d, 它们相互独立, 相应概率为 $1/2$ 、 $1/4$ 、 $1/8$ 、 $1/16$,

其中包含信息量最小的符号是 (A)

A、a

B、b

C、c

D、d

18. 在数据清理中，下面哪个不是处理缺失值的方法？ D

A、估算

B、整例删除

C、变量删除

D、成对删除

19. "过拟合是有监督学习的挑战，而不是无监督学习"以上说法是否正确： B

A、正确

B、错误

20. 逻辑回归与多元回归分析有哪些不同？ D

A、逻辑回归预测某事件发生的概率

B、逻辑回归有较高的拟合效果

C、逻辑回归回归系数的评估

D、以上全选

21. 支持向量是那些最接近决策平面的数据点 A

A、对

B、错

22. SVM 的效率依赖于 D

- A、核函数的选择
- B、核参数
- C、软间隔参数
- D、以上所有

23. 训练 SVM 的最小时间复杂度为 $O(n^2)$ ，那么一下哪种数据集不适合用 SVM? A

- A、大数据集
- B、小数据集
- C、中等大小数据集
- D、和数据集大小无关

24. 以下关于硬间隔 hard margin 描述正确的是 A

- A、SVM 允许分类存在微小误差
- B、SVM 允许分类是有大量误差

25. 关于 SVM 泛化误差描述正确的是 B

- A、超平面与支持向量之间距离
- B、SVM 对未知数据的预测能力
- C、SVM 的误差阈值

26. 如果我使用数据集的全部特征并且能够达到 100%的准确率，但在测试集上仅能达到

70%左右, 这说明: C

- A、欠拟合
- B、模型很棒
- C、过拟合

27. SVM 中的代价参数表示: C

- A、交叉验证的次数
- B、使用的核
- C、误分类与模型复杂性之间的平衡
- D、以上均不是

28. SVM 中使用高斯核函数之前通常会进行特征归一化, 以下关于特征归一化描述不正确的是? C

- A、经过特征正则化得到的新特征优于旧特征
- B、特征归一化无法处理类别变量
- C、SVM 中使用高斯核函数时, 特征归一化总是有用的

29. 假设你训练 SVM 后, 得到一个线性决策边界, 你认为该模型欠拟合。假如你想修改 SVM 的参数, 同样达到模型不会欠拟合的效果, 应该怎么做? A

- A、增大参数 C
- B、减小参数 C
- C、改变 C 并不起作用

D、以上均不正确

30. 假设你训练 SVM 后，得到一个线性决策边界，你认为该模型欠拟合。在下次迭代训练模型时，应该考虑： C

A、增加训练数据

B、减少训练数据

C、计算更多变量

D、减少特征

31. 下面哪个属于 SVM 应用 D

A、文本和超文本分类

B、图像分类

C、新文章聚类

D、以上均是

32. 判断：没有必要有一个用于应用维数降低算法的目标变量。 A

A、真

B、假

33. 想象一下，机器学习中有 1000 个输入特征和 1 个目标特征，必须根据输入特征和目标特征之间的关系选择 100 个最重要的特征。你认为这是减少维数的例子吗？ A

A、是

B、不是

34. 假设现在只有两个类，这种情况下 SVM 需要训练几次？ A

A、 1

B、 2

C、 3

D、 4

35. 判断：PCA 可用于在较小维度上投影和可视化数据。 A

A、 真

B、 假

36. 以下哪种算法不能用于降低数据的维数？ D

A、 t-SNE

B、 PCA

C、 LDA

D、 都不是

37. 判断：降维算法是减少构建模型所需计算时间的方法之一。 A

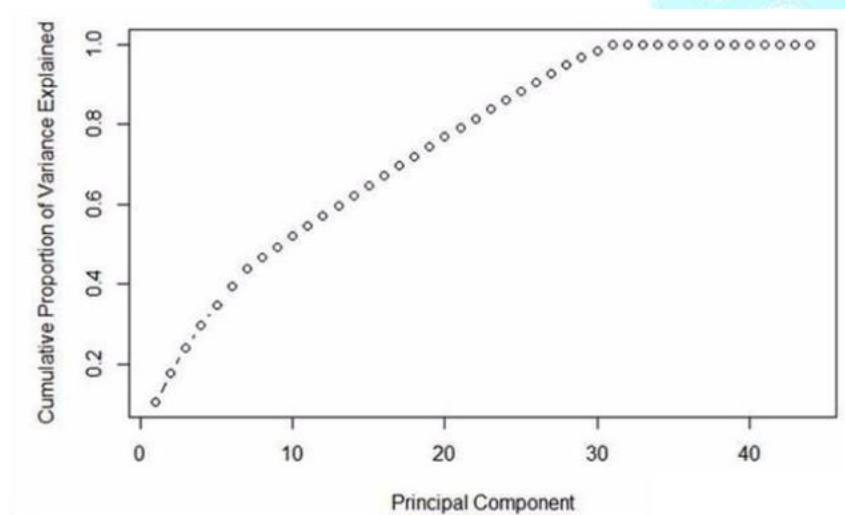
A、 真

B、 假

38. 以下哪种技术对于减少数据集的维度会更好？ A

- A、删除缺少值太多的列
- B、删除数据差异较大的列
- C、删除不同数据趋势的列
- D、都不是

39. 下图中主成分的最佳数量是多少？ B



- A、 7
- B、 30
- C、 40
- D、 不知道

40. 假设正在处理 10 类分类问题，并且想知道 LDA 最多可以产生几个判别向量。以下哪个是正确答案？ B

- A、 20
- B、 9

C、 21

D、 11

41. Logistic Regression 主要用于回归吗？ B

A、 是

B、 否

42. 是否可以对三分问题应用逻辑回归算法？ A

A、 是

B、 否

43. Logistic regression (逻辑回归) 是一种监督式机器学习算法吗？ A

A、 是

B、 否

44. 在训练逻辑回归之前需要对特征进行标准化。 B

A、 是

B、 否

45. 分析逻辑回归表现的一个良好的方法是 AIC, 它与线性回归中的 R 平方相似。有关 AIC,

以下哪项是正确的？ A

A、 具有最小 AIC 值的模型更好

B、具有最大 AIC 值的模型更好

C、视情况而定

D、以上都不是

46. 在逻辑回归输出与目标对比的情况下，以下评估指标中哪一项不适用？ D

A、AUC-ROC

B、准确度

C、Logloss

D、均方误差

47. 使用以下哪种算法进行变量选择？ A

A、LASSO

B、Ridge

C、两者

D、都不是

48. 以下哪些选项为真？ A

A、线性回归误差值必须正态分布，但是在 Logistic 回归的情况下，情况并非如此

B、逻辑回归误差值必须正态分布，但是在线性回归的情况下，情况并非如此

C、线性回归和逻辑回归误差值都必须正态分布

D、线性回归和逻辑回归误差值都不能正态分布

49. 所谓几率，是指发生概率和不发生概率的比值。所以，抛掷一枚正常硬币，正面朝上的

几率 (odds) 为多少? B

A、0.5

B、1

C、都不是

50. 如果以特征向量的相关系数作为模式相似性测度，则影响聚类算法结果的主要因素有

B

A 已知类别样本质量

B 分类准则

C 量纲

52. 在统计模式分类问题中，当先验概率未知时，可以使用 B

A 最小损失准则

B 最小最大损失准则

C 最小误判概率准则

53. 影响基本 K-均值算法的主要因素有 B

A 样本输入顺序

B 模式相似性测度

C 聚类准则

54. 以下不属于影响聚类算法结果的主要因素有 A

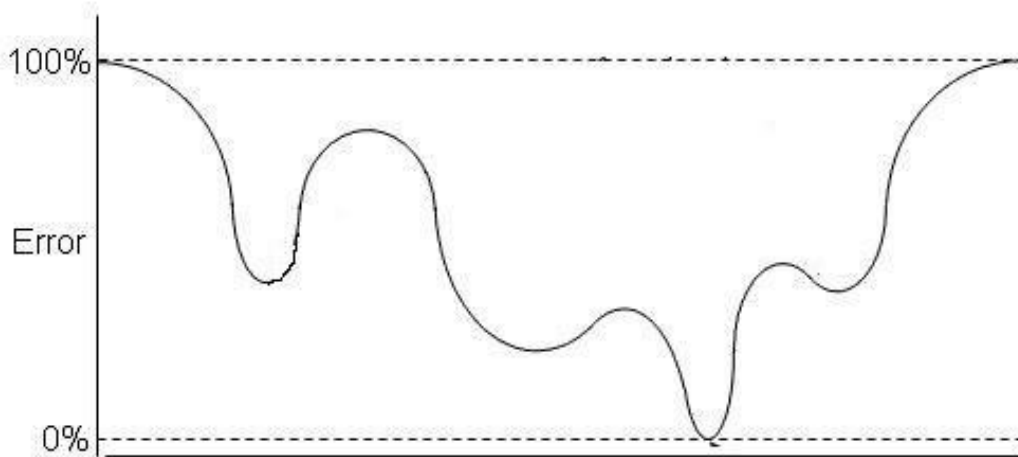
A 已知类别的样本质量

B 分类准则

C 特征选取

D 模式相似性测度

假设，下图是逻辑回归的代价函数



现在，图中有多少个局部最小值？ D

A、 1

B、 2

C、 3

D、 4

55. 对于任意值 “x”，考虑到 B

Logistic (x): 是任意值 “x” 的逻辑 (Logistic) 函数

Logit (x): 是任意值 "x" 的 logit 函数

Logit_inv (x): 是任意值 "x" 的逆逻辑函数

以下哪一项是正确的?

A、 $\text{Logistic}(x) = \text{Logit}(x)$

B、 $\text{Logistic}(x) = \text{Logit_inv}(x)$

C、 $\text{Logit_inv}(x) = \text{Logit}(x)$

D、都不是

56. 下面哪个选项中哪一项属于确定性算法? A

A、PCA

B、K-Means

C、以上都不是

57. 选择 Logistic 回归中的 One-Vs-All 方法中的哪个选项是真实的。 A

A、我们需要在 n 类分类问题中适合 n 个模型

B、我们需要适合 n-1 个模型来分类为 n 个类

C、我们需要只适合 1 个模型来分类为 n 个类

D、这些都没有

58. 下面哪个/些超参数的增加可能会造成随机森林数据过拟合? B

1 树的数量

2 树的深度

3 学习速率

- A、只有 1
- B、只有 2
- C、只有 3
- D、都正确

59. 下面哪一项对梯度下降 (GD) 和随机梯度下降 (SGD) 的描述是正确的? A

- 1 在 GD 和 SGD 中, 每一次迭代中都是更新一组参数以最小化损失函数。
- 2 在 SGD 中, 每一次迭代都需要遍历训练集中的所有样本以更新一次参数。
- 3 在 GD 中, 每一次迭代需要使用整个训练集或子训练集的数据更新一个参数。

- A、只有 1
- B、只有 2
- C、只有 3
- D、都正确

60. 下面哪个/些对「类型 1 (Type-1)」和「类型 2 (Type-2)」错误的描述是正确的? D

类型 1 通常称之为假正类, 类型 2 通常称之为假负类。

类型 2 通常称之为假正类, 类型 1 通常称之为假负类。

类型 1 错误通常在其是正确的情况下拒绝假设而出现。

- A、只有 1
- B、只有 2
- C、只有 3

D、1 和 3

61. 对数损失度量函数可以取负值。 B

A、对

B、错

62. 假定你在神经网络中的隐藏层中使用激活函数假定你在神经网络中的隐藏层中使用激活函数 X。在特定神经元给定任意输入，你会得到输出「-0.0001」。X 可能是以下哪一个激活函数？ B

A、ReLU

B、tanh

C、SIGMOID

D、以上都不是

63. 假定你正在处理类属特征，并且没有查看分类变量在测试集中的分布。现在你想将 one hot encoding (OHE) 应用到类属特征中。

那么在训练集中将 OHE 应用到分类变量可能要面临的困难是什么？ D

A、分类变量所有的类别没有全部出现在测试集中

B、类别的频率分布在训练集和测试集是不同的

C、训练集和测试集通常会有一样的分布

D、A 和 B 都正确

64. 在线性回归问题中，我们用“R 方”来衡量拟合的好坏。在线性回归模型中增加特征值并再训练同一模型。下列哪一项是正确的？ C

- A、如果 R 方上升，则该变量是显著的
- B、如果 R 方下降，则该变量不显著
- C、单单 R 方不能反映变量重要性，不能就此得出正确结论
- D、都不正确

65. 假设线性回归模型完美拟合训练数据（即训练误差为零），则下列哪项是正确的？ C

- A、测试误差一定为零
- B、测试误差一定不为零
- C、以上都不对

66. 导出线性回归的参数时，我们做出下列哪种假定？（ D ）

- 1.因变量 y 和自变量 x 的真实关系是线性的
- 2.模型误差是统计独立的
- 3.误差通常服从一个平均值为零，标准差恒定的分布
- 4.自变量 x 是非随机的，无错的

- A、1,2 和 3
- B、1,3 和 4
- C、1 和 3
- D、以上都对

67. 下列哪一项说明了 X, Y 之间的较强关系 (A)

- A、相关系数为 0.9
- B、Beta 系数为 0 的空假设的 p-value 是 0.0001
- C、Beta 系数为 0 的空假设的 t 统计量是 30
- D、都不对

68. 下列关于异方差性哪项是正确的? (A)

- A、线性回归有变化的误差项
- B、线性回归有恒定的误差项
- C、线性回归有零误差项
- D、以上都不对

69. 我们可以通过一种叫“正规方程”的分析方法来计算线性回归的相关系数, 下列关于“正规方程” 哪一项是正确的? (D)

- 1.我们不必选择学习比率
- 2.当特征值数量很大时会很慢
- 3.不需要迭代

- A、 1 和 2
- B、 1 和 3
- C、 2 和 3
- D、 1,2 和 3

70. 下列哪项可以评价回归模型? (D)

1.R 方

2.调整 R 方

3.F 统计量

4.RMSE/MSE/MAE

A、 2 和 4

B、 1 和 2

C、 2,3 和 4

D、 以上所有

71. 关于 Ridge 和 Lasso 回归在特征值选择上的方法, 一下哪项正确? (B)

A、 Ridge 回归使用特征值的子集选择

B、 Lasso 回归使用特征值的子集选择

C、 二者都使用特征值的子集选择

D、 以上都不正确

72. 相关变量的相关系数可以为零, 对吗? (A)

A、 是

B、 否

73. 若两个变量相关, 它们之间一定有线性关系吗? (B)

A、 是

B、否

74. X 轴是自变量, Y 轴是因变量

下列对 A, B 各自残差和的陈述哪项正确? (C)

A、A 比 B 高

B、A 比 B 低

C、两者相同

D、以上都不对

75. 为了评价一个简单线性回归模型 (单自变量), 需要多少个参数? (B)

A、1

B、2

C、不确定

76. 关于回归和相关, 下列哪项是正确的? (D)

提示: y 是因变量, x 是自变量

A、在两者中, x 、 y 关系都是对称的

B、在两者中, x 、 y 关系都是不对称的

C、 x 、 y 在相关情况下不对称, 在回归中对称

D、 x 、 y 在相关情况下对称, 在回归中不对称

77. 假设用一个复杂回归模型拟合一个数据集, 使用带固定参数 λ 的 Ridge 回归来

减小它的复杂度，下列哪项描述了偏差和方差与 λ 的关系？（ C ）

- A、对于非常大的 λ ，偏差很小，方差很小
- B、对于非常大的 λ ，偏差很小，方差很大
- C、对于非常大的 λ ，偏差很大，方差很小
- D、对于非常大的 λ ，偏差很大，方差很大

78. 观测值是如何影响过拟合的？

提示：余下所有参数都一致

- 1. 观测更少更易过拟合
- 2. 观测更少更不易过拟合
- 3. 观测更多更易过拟合
- 4. 观测更多更不易过拟合

- A、1 和 4
- B、2 和 4
- C、1 和 3
- D、都不正确

79. 可以根据平均值和中位数计算斜率吗？（ B ）

- A、可以
- B、不可以

80. 列哪一种回归方法的相关系数没有闭式解？（ B ）

- A、Ridge 回归
- B、Lasso 回归
- C、Ridge 回归 and Lasso 回归
- D、两者都不是

81. 下列哪一个项对欠拟合和过拟合的权衡影响最大? (A)

- A、多项式次数
- B、是否通过矩阵倒置或梯度下降来学习权重
- C、使用常数项

82. 假设使用逻辑回归模型处理 n 元分类问题, 可以用到 One-vs-rest 方法, 则下列哪一项是正确的? (A)

- A、在 n 元分类问题中, 需要拟合 n 个模型
- B、为了分类为 n 类, 需要拟合 $n-1$ 个模型
- C、为了分类为 n 类, 只需要拟合 1 个模型
- D、都不正确

83. 考虑线性回归和逻辑回归中的重量/相关系数, 关于 cost 函数的偏导, 下列哪一项是正确的? (B)

- A、都不一样
- B、都一样
- C、无法确定

D、以上都不对

84. 逻辑回归是输出结果落在 $[0,1]$ 区间内，下列哪个函数用于转换概率，使其落入 $[0,1]$ ？

(A)

A、Sigmoid

B、Mode

C、Square

D、Probit

85. 下列关于回归分析中的残差表述正确的是：(C)

A、残差的平均值总为零

B、残差的平均值总小于零

C、残差的平均值总大于零

D、残差没有此类规律

86. 在线性回归问题中，我们用 R 方 “R-squared” 来衡量拟合的好坏。在线性回归模型中增加特征值并再训练同一模型。下列哪一项是正确的？(C)

A、如果 R 方上升，则该变量是显著的

B、如果 R 方下降，则该变量不显著

C、单单 R 方不能反映变量重要性，不能就此得出正确结论

D、都不正确

87. 假设线性回归模型完美拟合训练数据（即训练误差为零），则下列哪项是正确的？(C)

A、测试误差一定为零

B、测试误差一定不为零

C、以上都不对

88. 一个人年龄和健康之间的相关系数是-1.09，据此可以得出：(C)

A、年龄是健康预测的好的参考量

B、年龄是健康预测的不好的参考量

C、都不对

89. 下列哪种方法被用于预测因变量？(B)

1.线性回归

2.逻辑回归

A、1 和

B、1

C、2

D、都不是

90. 为了检验连续变量 x , y 之间的线性关系，下列哪种图最合适？(A)

A、散点图

B、条形图

C、直方图

D、都不对

91. 在导出线性回归的参数时，我们做出下列哪种假定？(D)

- 1.因变量 y 和自变量 x 的真实关系是线性的
- 2.模型误差是统计独立的
- 3.误差通常服从一个平均值为零，标准差恒定的分布
- 4.自变量 x 是非随机的，无错的

A、1,2 and 3

B、1,3 and 4

C、1 and 3

D、以上都对

92. 下列哪一项说明了 X , Y 之间的较强关系？(A)

- A、相关系数为 0.9
- B、Beta 系数为 0 的空假设的 p -value 是 0.0001
- C、Beta 系数为 0 的空假设的 t 统计量是 30
- D、都不对

93. 下列哪项可以评价回归模型？(D)

1.R 方 R Squared

2.调整 R 方

3.F 统计量

4. RMSE/MSE/MAE

- A、2 和 4
- B、1 和 2
- C、2,3 和 4
- D、以上所有

94. 假设你在训练一个线性回归模型，以下哪项是正确的？(C)

- 1. 数据越少越易过拟合
- 2. 假设区间小则易过拟合

- A、都是错的
- B、1 错 2 对
- C、1 对 2 错
- D、都是对的

95. 若两个变量相关，它们之间一定有线性关系吗？(B)

- A、是
- B、否

96. 为了评价一个简单线性回归模型（单自变量），需要多少个参数？(B)

- A、1
- B、2
- C、不确定

97. 将原始数据进行集成、变换、维度规约、数值规约是在以下哪个步骤的任务? (C)

A. 频繁模式挖掘 B. 分类和预测 C. 数据预处理 D. 数据流挖掘

98. 下面哪种不属于数据预处理的方法? (D)

A 变量代换 B 离散化 C 聚集 D 估计遗漏值

99. 什么是 KDD? (A)

A. 数据挖掘与知识发现 B. 领域知识发现 C. 文档知识发现 D. 动态知识发现

90. 建立一个模型, 通过这个模型根据已知的变量值来预测其他某个变量值属于数据挖掘的哪一类任务? (C)

A. 根据内容检索 B. 建模描述

C. 预测建模 D. 寻找模式和规则

91. 以下哪种方法不属于特征选择的标准方法? (D)

A 嵌入 B 过滤 C 包装 D 抽样

92. 下列属于无监督学习的是: (A)

k-means

B. SVM

C. 最大熵

D. CRF

93. 下列不是 SVM 核函数的是: (B)

多项式核函数

B. Logistic 核函数

C. 径向基核函数

D. Sigmoid 核函数

94. bootstrap 数据是什么意思? (提示: 考 “bootstrap” 和 “boosting” 区别) (C)

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

95. 解决隐马模型中预测问题的算法是? (D)

前向算法

B. 后向算法

C. Baum-Welch 算法

D. 维特比算法

96. 训练完 SVM 模型后, 不是支持向量的那些样本我们可以丢掉, 也可以继续分类: (A)

A. 正确

B. 错误

97. 以下哪些算法, 可以用神经网络去构造: (B)

1.KNN

2.线性回归

3.对数几率回归

A. 1 和 2

B. 2 和 3

C. 1, 2 和 3

D. 以上都不是

98. 请选择下面可以应用隐马尔科夫(HMM)模型的选项: (D)

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

99. 对于 PCA(主成分分析)转化过的特征, 朴素贝叶斯的“不依赖假设”总是成立, 因为所有主要成分都是正交的, 这个说法是: (B)

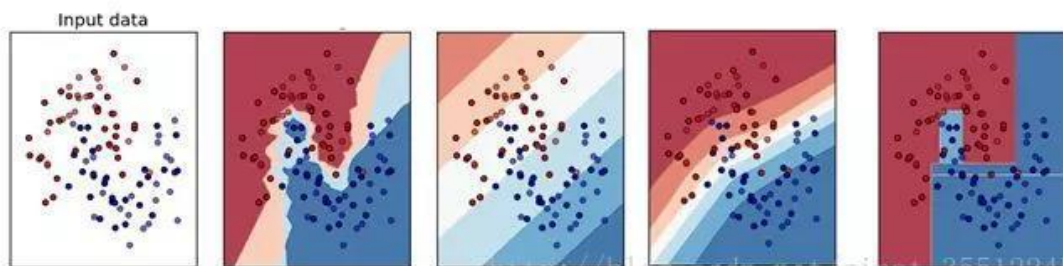
A. 正确的

B. 错误的

100. 数据科学家可能会同时使用多个算法（模型）进行预测，并且最后把这些算法的结果集成起来进行最后的预测（集成学习），以下对集成学习说法正确的是：（B）

- A. 单个模型之间有高相关性
- B. 单个模型之间有低相关性
- C. 在集成学习中使用“平均权重”而不是“投票”会比较好
- D. 单个模型都是用的一个算法

101. 以下哪个图是 KNN 算法的训练边界？（B）



- A) B
- B) A
- C) D
- D) C

102. 对于线性回归，我们应该有以下哪些假设？（D）

- 1.找到利群点很重要，因为线性回归对利群点很敏感
- 2.线性回归要求所有变量必须符合正态分布
- 3.线性回归假设数据没有多重线性相关性

- A. 1 和 2

- B. 2 和 3
- C. 1,2 和 3
- D. 以上都不是

103. 我们注意变量间的相关性。在相关矩阵中搜索相关系数时, 如果我们发现 3 对变量的相关系数是(Var1 和 Var2, Var2 和 Var3, Var3 和 Var1)是-0.98, 0.45, 1.23 . 我们可以得出什么结论? (C)

- 1.Var1 和 Var2 是非常相关的
- 2.因为 Var 和 Var2 是非常相关的, 我们可以去除其中一个
- 3.Var3 和 Var1 的 1.23 相关系数是不可能的

- A. 1 and 3
- B. 1 and 2
- C. 1,2 and 3
- D. 1

104. 如果在一个高度非线性并且复杂的一些变量中 “一个树模型可比一般的回归模型效果更好” 是 (A)

- A. 对的
- B. 错的

105. 下面对集成学习模型中的弱学习者描述错误的是? (C)

- A. 他们经常不会过拟合

- B. 他们通常带有高偏差，所以其并不能解决复杂学习问题
- C. 他们通常会过拟合

106. 最出名的降维算法是 PAC 和 t-SNE。将这两个算法分别应用到数据「X」上，并得到数据集「X_projected_PCA」, 「X_projected_tSNE」。下面哪一项对「X_projected_PCA」和「X_projected_tSNE」的描述是正确的？ (B)

- A. X_projected_PCA 在最近邻空间能得到解释
- B. X_projected_tSNE 在最近邻空间能得到解释
- C. 两个都在最近邻空间能得到解释
- D. 两个都不能在最近邻空间得到解释

107. 在 k-均值算法中，以下哪个选项可用于获得全局最小？ (D)

- A. 尝试为不同的质心 (centroid) 初始化运行算法
- B. 调整迭代的次数
- C. 找到集群的最佳数量
- D. 以上所有

108. 你正在使用带有 L1 正则化的 logistic 回归做二分类，其中 C 是正则化参数，w1 和 w2 是 x1 和 x2 的系数。当你把 C 值从 0 增加至非常大的值时，下面哪个选项是正确的？ (B)

- A. 第一个 w2 成了 0，接着 w1 也成了 0
- B. 第一个 w1 成了 0，接着 w2 也成了 0

C. w_1 和 w_2 同时成了 0

D. 即使在 C 成为大值之后, w_1 和 w_2 都不能成 0

109. 下面哪个选项中哪一项属于确定性算法? (A)

A.PCA

B.K-Means

C. 以上都不是

110. 下面哪个/些超参数的增加可能会造成随机森林数据过拟合? (B)

A. 树的数量

B. 树的深度

C. 学习速率

111. 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

112. 机器学习中做特征选择时, 可能用到的方法有? (E)

A.卡方

B. 信息增益

- C. 平均互信息
- D. 期望交叉熵
- E. 以上都有

113. 一般, K-NN 最近邻方法在 (A) 的情况下效果较好。

- A. 样本较多但典型性不好
- B. 样本呈团状分布
- C. 样本较少但典型性好
- D. 样本呈链状分布

114. 以下描述错误的是 (C)

- A. SVM 是这样一个分类器, 它寻找具有最小边缘的超平面, 因此它也经常被称为最小边缘分类器
- B. 在聚类分析当中, 簇内的相似性越大, 簇间的差别越大, 聚类的效果就越差
- C. 在决策树中, 随着树中结点数变得太大, 即使模型的训练误差还在继续降低, 但是检验误差开始增大, 这是出现了模型拟合不足的原因
- D. 聚类分析可以看作是一种非监督的分类

115. 关于正态分布, 下列说法错误的是 (C)

- A. 正态分布具有集中性和对称性
- B. 正态分布的均值和方差能够决定正态分布的位置和形态
- C. 正态分布的偏度为 0, 峰度为 1

D. 标准正态分布的均值为 0, 方差为 1

116. 在以下不同的场景中,使用的分析方法不正确的有 (B)

A. 根据商家最近一年的经营及服务数据,用聚类算法判断出天猫商家在各自主营类目下所属的商家层级

B. 根据商家近几年的成交数据,用聚类算法拟合出用户未来一个月可能的消费金额公式

C. 用关联规则算法分析出购买了汽车坐垫的买家,是否适合推荐汽车脚垫

D. 根据用户最近购买的商品信息,用决策树算法识别出淘宝买家可能是男还是女

117. 下列表述中,在 k-fold 交叉验证中关于选择 K 说法正确的是: (D)

A、较大的 K 并不总是好的,选择较大的 K 可能需要较长的时间来评估你的结果

B、相对于期望误差来说,选择较大的 K 会导致低偏差 (因为训练 folds 会变得与整个数据集相似)

C、在交叉验证中通过最小化方差法来选择 K 值

D、以上都正确

118. 在决策树中,用作分裂节点的 information gain 说法不正确的是(A)

A、较小不纯度的节点需要更多的信息来区分总体

B、信息增益可以使用熵得到

C、信息增益更加倾向于选择有较多取值的属性

119. 以下哪一种方法最适合在 n ($n > 1$) 维空间中做异常点检测(C)

A、正态分布图

B、盒图

C、马氏距离

D、散点图

120. logistic 回归与多元回归分析有哪些不同? (D)

A、logistic 回归预测某事件发生的概率

B、logistic 回归有较高的拟合效果

C、logistic 回归回归系数的评估

D、以上全选

121. bootstrap 数据的含义是: (C)

A、有放回的从整体 M 中抽样 m 个特征

B、无放回的从整体 M 中抽样 m 个特征

C、有放回的从整体 N 中抽样 n 个样本

D、无放回的从整体 N 中抽样 n 个样本

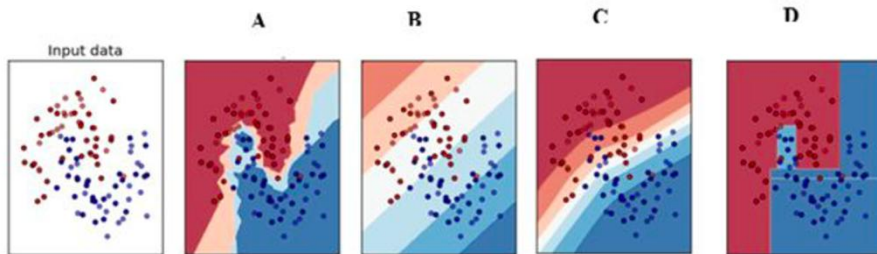
122. 决策树可以用来执行聚类吗? (A)

A、能

B、不能

中等

1. 下面那个决策边界是神经网络生成的？（E）



- A. A
- B. D
- C. C
- D. B
- E. 以上都有

2. 关于 Logit 回归和 SVM 不正确的是（A）

- A. Logit 回归本质上是一种根据样本对权值进行极大似然估计的方法，而后验概率正比于先验概率和似然函数的乘积。logit 仅仅是最大化似然函数，并没有最大化后验概率，更谈不上最小化后验概率
- B. Logit 回归的输出就是样本属于正类别的几率，可以计算出概率，正确
- C. SVM 的目标是找到使得训练数据尽可能分开且分类间隔最大的超平面，应该属于结构风险最小化。
- D. SVM 可以通过正则化系数控制模型的复杂度，避免过拟合

3. Fisher 线性判别函数的求解过程是将 M 维特征矢量投影在 (B) 中进行求解。

A、 $M-1$ 维空间

B、一维空间

C、三维空间

D、二维空间

4. 在其它条件不变的前提下，以下哪种做法容易引起机器学习中的过拟合问题 (D)

A、增加训练集数量

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/RBF 核代替

5. 如果线性回归模型中的随机误差存在异方差性，那么参数的 OLS 估计量是 (B)

A、无偏的，有效的

B、无偏的，非有效的

C、有偏的，有效的

D、有偏的，非有效的

6. 下列哪个不属于 CRF 模型对于 HMM 和 MEMM 模型的优势 B

A、特征灵活

B、速度快

C、可容纳较多上下文信息

D、全局最优

7. 已知一组数据的协方差矩阵 P , 下面关于主分量说法错误的是 (C)

A、主分量分析的最佳准则是对一组数据进行按一组正交基分解, 在只取相同数量分量的条件下, 以均方误差计算截尾误差最小

B、在经主分量分解后, 协方差矩阵成为对角矩阵

C、主分量分析就是 K-L 变换

D、主分量是通过求协方差矩阵的特征值得到

8. 以下哪些方法不可以直接来对文本分类? A

A、Kmeans

B、决策树

C、支持向量机

D、KNN

9. 以下 (B) 不属于线性分类器最佳准则?

A、感知准则函数

B、贝叶斯分类

C、支持向量机

D、Fisher 准则

10. 在统计模式分类问题中, 当先验概率未知时, 可以使用 (B)

- A、最小损失准则
- B、最小最大损失准则
- C、最小误判概率准则

11. 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、 $2x-y=0$

12. 在 Logistic Regression 中,如果同时加入 L1 和 L2 范数,不会产生什么效果 (D)

- A、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

13. 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/RBF 核代替线性核

14. 隐马尔可夫模型三个基本问题以及相应的算法说法错误的是 (D)

A、评估—前向后向算法

B、解码—维特比算法

C、学习—Baum-Welch 算法

D、学习—前向后向算法

从第 15 题开始

15. 假设你使用 $\log\text{-loss}$ 函数作为评估标准。下面这些选项，哪些是对作为评估标准的 $\log\text{-loss}$ 的正确解释。 D

A、如果一个分类器对不正确的分类很自信， $\log\text{-loss}$ 会严重的批评它

B、对一个特别的观察而言，分类器为正确的类别分配非常小的概率，然后对 $\log\text{-loss}$ 的相应分布会非常大

C、 $\log\text{-loss}$ 越低，模型越好

D、以上都是

16. 对于信息增益，决策树分裂节点，下面说法正确的是 (C)

1 纯度高的节点需要更多的信息去区分

2 信息增益可以用“1 比特-熵”获得

3 如果选择一个属性具有许多归类值，那么这个信息增益是有偏差的

- A、 1
- B、 2
- C、 2 和 3
- D、 所有以上

17. 模型的高 bias 是什么意思, 我们如何降低它 ? B

- A、 在特征空间中减少特征
- B、 在特征空间中增加特征
- C、 增加数据点
- D、 B 和 C
- E、 以上所有

18. 回归模型中存在多重共线性, 你如何解决这个问题? D

- 1 去除这两个共线性变量
- 2 我们可以先去除一个共线性变量
- 3 计算 VIF(方差膨胀因子), 采取相应措施
- 4 为了避免损失信息, 我们可以使用一些正则化方法, 比如, 岭回归和 lasso 回归

- A、 1
- B、 2
- C、 2 和 3
- D、 2, 3 和 4

19. 对于 k 折交叉验证, 以下对 k 的说法正确的是 (D)

- 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

20. 下列选项中, 识别模式与其他不一样的是 E

- A、用户年龄分布判断: 少年、青年、中年、老年
- B、医生给病人诊断发病类型
- C、投递员分拣信件
- D、消费者类型判断: 高消费、一般消息、低消费
- E、出行方式判断: 步行、骑车、坐车
- F、商家对商品分级

21. 关于 ARMA 、 AR 、 MA 模型的功率谱, 下列说法正确的是 (D)

- A、MA 模型是同一个全通滤波器产生的
- B、MA 模型在极点接近单位圆时, MA 谱是一个深谷
- C、AR 模型在零点接近单位圆时, AR 谱是一个尖峰
- D、RMA 谱既有尖峰又有深谷

22. 给线性回归模型添加一个不重要的特征可能会造成? A

- A、增加 R-square

B、减少 R-square

23. 线性回归的基本假设不包括哪个？ C

A、随机误差项是一个期望值为 0 的随机变量

B、对于解释变量的所有观测值，随机误差项有相同的方差

C、随机误差项彼此相关

D、解释变量是确定性变量不是随机变量，与随机误差项之间相互独立

E、随机误差项服从正态分布

24. 决策树的父节点和子节点的熵的大小关系是什么？ D

A、A. 决策树的父节点更大

B、B. 子节点的熵更大

C、C. 两者相等

D、D. 根据具体情况而定

25. 统计模式分类问题中，当先验概率未知时，可以使用 (A)

A、最小最大损失准则

B、最小误判概率准则

C、最小损失准则

D、N-P 判决

26. 假定你使用了一个很大 γ 值的 RBF 核，这意味着： B

- A、模型将考虑使用远离超平面的点建模
- B、模型仅使用接近超平面的点来建模
- C、模型不会被点到超平面的距离所影响
- D、以上都不正确

27. SVM 在下列那种情况下表现糟糕 C

- A、线性可分数据
- B、清洗过的数据
- C、含噪声数据与重叠数据点

28. 假设你训练了一个基于线性核的 SVM，多项式阶数为 2，在训练集和测试集上准确率都为 100%。如果增加模型复杂度或核函数的多项式阶数，将会发生什么？ A

- A、导致过拟合
- B、导致欠拟合
- C、无影响，因为模型已达 100%准确率
- D、以上均不正确

29. 最常用的降维算法是 PCA，以下哪项是关于 PCA 的？ D

- 1.PCA 是一种无监督的方法
- 2.它搜索数据具有最大差异的方向
- 3.主成分的最大数量 \leq 特征能数量
- 4.所有主成分彼此正交

A、2、3 和 4

B、1、2 和 3

C、1、2 和 4

D、以上所有

30. 判断: t-SNE 学习非参数映射。 A

A、真

B、假

31. 对于 t-SNE 代价函数, 以下陈述中的哪一个正确? B

A、本质上是不对称的

B、本质上是对称的

C、与 SNE 的代价函数相同

32. 在相同的机器上运行并设置最小的计算能力, 以下哪种情况下 t-SNE 比 PCA 降维效果更好? C

A、具有 1 百万项 300 个特征的数据集

B、具有 100000 项 310 个特征的数据集

C、具有 10,000 项 8 个特征的数据集

D、具有 10,000 项 200 个特征的数据集

33. 假设使用维数降低作为预处理技术, 使用 PCA 将数据减少到 k 维度。然后使用这些 PCA

预测作为特征，以下哪个声明是正确的？ B

- A、更高的“k”意味着更正则化
- B、更高的“k”意味着较少的正则化
- C、都不对

34. 与 PCA 相比，t-SNE 的以下说明哪个正确？ A

- A、数据巨大（大小）时，t-SNE 可能无法产生更好的结果。
- B、无论数据的大小如何，T-NSE 总是产生更好的结果。
- C、对于较小尺寸的数据，PCA 总是比 t-SNE 更好。
- D、都不是

35. 以下对于 t-SNE 和 PCA 的陈述中哪个是正确的？ D

- A、t-SNE 是线性的，而 PCA 是非线性的
- B、t-SNE 和 PCA 都是线性的
- C、t-SNE 和 PCA 都是非线性的
- D、t-SNE 是非线性的，而 PCA 是线性的

36. 以下哪种情况 LDA 会失败？ A

- A、如果有辨识性的信息不是平均值，而是数据的方差
- B、如果有辨识性的信息是平均值，而不是数据方差
- C、如果有辨识性的信息是数据的均值和方差
- D、都不是

37. PCA 和 LDA 的以下比较哪些是正确的？ D

1. LDA 和 PCA 都是线性变换技术
2. LDA 是有监督的，而 PCA 是无监督的
3. PCA 最大化数据的方差，而 LDA 最大化不同类之间的分离

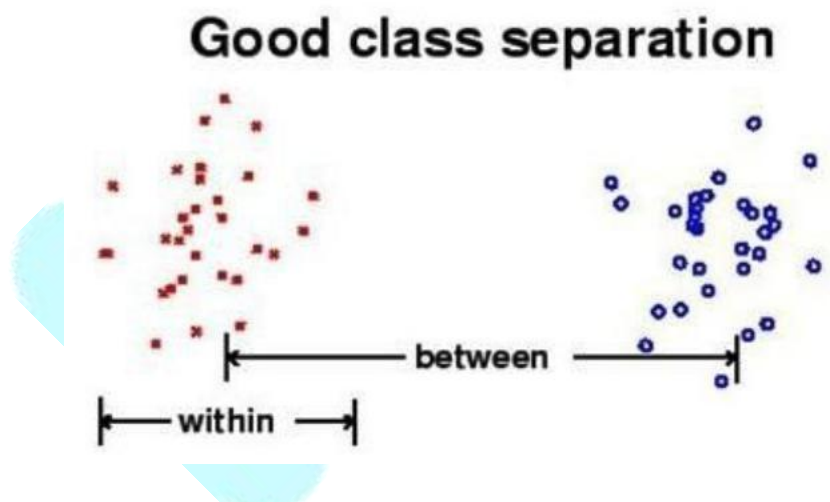
A、1 和 2

B、1 和 3

C、只有 3

D、1、2 和 3

38. LDA 的以下哪项是正确的？ A



A、LDA 旨在最大化之间类别的距离，并最小化类内之间的距离

B、LDA 旨在最小化类别和类内之间的距离

C、LDA 旨在最大化类内之间的距离，并最小化类别之间的距离

D、LDA 旨在最大化类别和类内之间的距离

39. 如果对相同的数据进行逻辑回归，将花费更少的时间，并给出比较相似的精度（也可能不一样），怎么办？ D

假设在庞大的数据集上使用 Logistic 回归模型。可能遇到一个问题，Logistic 回归需要很长时间才能训练。

- A、降低学习率，减少迭代次数
- B、降低学习率，增加迭代次数
- C、提高学习率，增加迭代次数
- D、增加学习率，减少迭代次数

40. 以下哪种方法能最佳地适应逻辑回归中的数据？ B

- A、Least Square Error
- B、Maximum Likelihood
- C、Jaccard distance
- D、Both A and B

41. 逻辑回归的以下模型： $P(y = 1 | x, w) = g(w_0 + w_1 x)$ 其中 $g(z)$ 是逻辑函数。

在上述等式中，通过改变参数 w 可以得到的 $P(y = 1 | x; w)$ 被视为 x 的函数。 C

- A、 $(0, \text{inf})$
- B、 $(-\text{inf}, 0)$
- C、 $(0, 1)$
- D、 $(-\text{inf}, \text{inf})$

42. 逻辑回归的以下模型: $P(y = 1 | x, w) = g(w_0 + w_1 x)$ 其中 $g(z)$ 是逻辑函数。

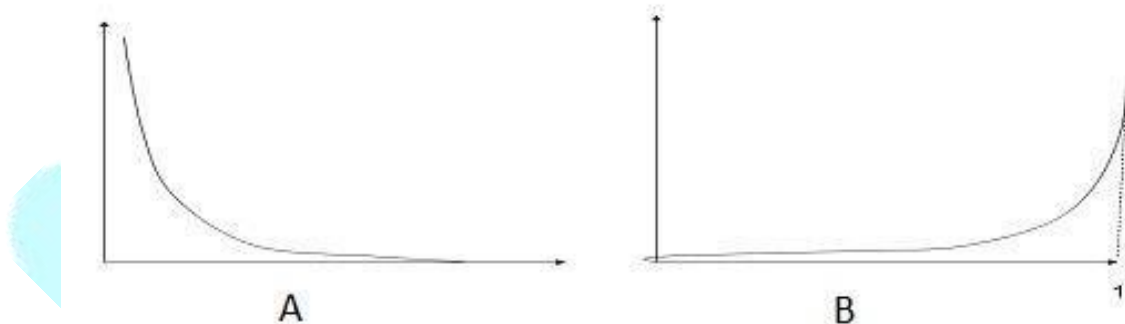
在上述等式中, 通过改变参数 w 可以得到的 $P(y = 1 | x; w)$ 被视为 x 的函数。

在上面的问题中, 你认为哪个函数会产生 $(0,1)$ 之间的 p ? A

- A、逻辑函数
- B、对数似然函数
- C、两者的复合函数
- D、都不会

43. 以下哪个图像显示 $y = 1$ 的代价函数? A

以下是两类分类问题的逻辑回归 (Y 轴损失函数和 x 轴对数概率) 的损失函数。



注: Y 是目标类

- A、A
- B、B
- C、两者
- D、这些都没有

44. 如果对相同的数据进行逻辑回归，将花费更少的时间，并给出比较相似的精度（也可能不一样），怎么办？ D

（假设在庞大的数据集上使用 Logistic 回归模型。可能遇到一个问题，Logistic 回归需要很长时间才能训练。）

- A、降低学习率，减少迭代次数
- B、降低学习率，增加迭代次数
- C、提高学习率，增加迭代次数
- D、增加学习率，减少迭代次数

45. Logit 函数 (给定为 $l(x)$) 是几率函数的对数。域 $x = [0,1]$ 中 logit 函数的范围是多少？

- A
- A、 $(-\infty, \infty)$
- B、 $(0,1)$
- C、 $(0, \infty)$
- D、 $(-\infty, 0)$

46. 模式识别中，不属于马式距离较之于欧式距离的优点的是 A

- A 平移不变性
- B 尺度不变性
- C 考虑了模式的分布

47. 假定特征 F1 可以取特定值：A、B、C、D、E 和 F，其代表着学生在大学所获得的评

分。

在下面说法中哪一项是正确的？ B

A、特征 F1 是名义变量 (nominal variable) 的一个实例。

B、特征 F1 是有序变量 (ordinal variable) 的一个实例。

C、该特征并不属于以上的分类。

D、以上说法都正确。

48. 以下是目标变量在训练集上的 8 个实际值 [0,0,0,1,1,1,1,1]，目标变量的熵是所少？

A

A、 $-(5/8 \log(5/8) + 3/8 \log(3/8))$

B、 $5/8 \log(5/8) + 3/8 \log(3/8)$

C、 $3/8 \log(5/8) + 5/8 \log(3/8)$

D、 $5/8 \log(3/8) - 3/8 \log(5/8)$

49. 假如你在「Analytics Vidhya」工作，并且想开发一个能预测文章评论次数的机器学习算法。你的分析的特征是基于如作者姓名、作者在 Analytics Vidhya 写过的总文章数量等

等。那么在这样一个算法中，你会选择哪一个评价度量标准？ A

1 均方误差

2 精确度

3 F1 分数

A、只有 1

B、只有 2

C、只有 3

50. 假设给定三个变量 X , Y , Z 。 (X, Y) 、 (Y, Z) 和 (X, Z) 的 Pearson 相关性系数分别为 C_1 、 C_2 和 C_3 。现在 X 的所有值加 2 (即 $X+2$)， Y 的全部值减 2 (即 $Y-2$)， Z 保持不变。

那么运算之后的 (X, Y) 、 (Y, Z) 和 (X, Z) 相关性系数分别为 D_1 、 D_2 和 D_3 。

现在试问 D_1 、 D_2 、 D_3 和 C_1 、 C_2 、 C_3 之间的关系是什么？ E

A、 $D_1 = C_1$, $D_2 < C_2$, $D_3 > C_3$

B、 $D_1 = C_1$, $D_2 > C_2$, $D_3 > C_3$

C、 $D_1 = C_1$, $D_2 > C_2$, $D_3 < C_3$

D、 $D_1 = C_1$, $D_2 < C_2$, $D_3 < C_3$

E、 $D_1 = C_1$, $D_2 = C_2$, $D_3 = C_3$

51. 给线性回归模型添加一个不重要的特征可能会造成： A

1 增加 R-square

2 减少 R-square

A、只有 1 是对的

B、只有 2 是对的

C、1 或 2 是对的

D、都不对

52. 在先前问题中，假定你已经鉴别了多元共线特征。那么下一步你可能的操作是什么？ D

1 移除两个共线变量

2 不移除两个变量，而是移除一个

3 移除相关变量可能会导致信息损失。为了保留这些变量，我们可以使用带罚项的回归模型 (如 ridge 或 lasso regression)。

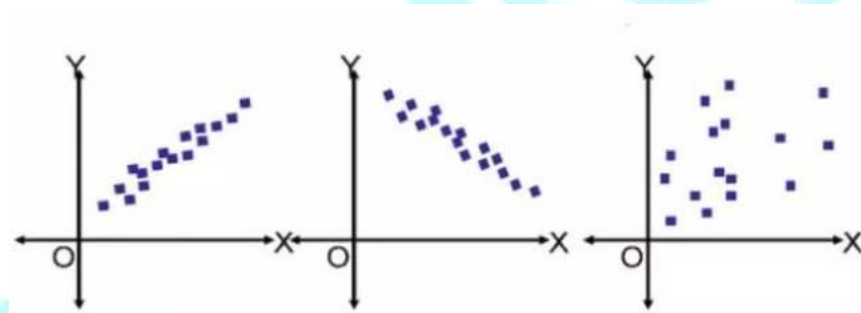
A、只有 1

B、只有 2

C、只有 3

D、2 或 3

53. 给定下面两个特征的三个散点图 (从左到右依次为图 1、2、3)



在上面的图像中，哪一个是多元共线 (multi-collinear) 特征? D

A、图 1 中的特征

B、图 2 中的特征

C、图 3 中的特征

D、图 1、2 中的特征

54. 假设存在一个黑箱算法，其输入为有多个观察 ($t_1, t_2, t_3, \dots, t_n$) 的训练数据和一个新的观察 (q_1)。该黑箱算法输出 q_1 的最近邻 t_i 及其对应的类别标签 c_i 。你可以将这个

黑箱算法看作是一个 1-NN (1-最近邻) 能够仅基于该黑箱算法而构建一个 k-NN 分类算

法? 注: 相对于 k 而言, n (训练观察的数量) 非常大。 A

A、可以

B、不可以

55. 下面哪个/些选项对 K 折交叉验证的描述是正确的 D

1 增大 K 将导致交叉验证结果时需要更多的时间

2 更大的 K 值相比于小 K 值将对交叉验证结构有更高的信心

3 如果 $K=N$, 那么其称为留一交叉验证, 其中 N 为验证集中的样本数量

A、1 和 2

B、2 和 3

C、1 和 3

D、1、2 和 3

56. 在集成学习中, 模型集成了弱学习者的预测, 所以这些模型的集成将比使用单个模型预测效果更好。下面哪个/些选项对集成学习模型中的弱学习者描述正确? A

1 他们经常不会过拟合

2 他们通常带有高偏差, 所以其并不能解决复杂学习问题

3 他们通常会过拟合

A、1 和 2

B、1 和 3

C、2 和 3

D、只有 1

57. 对于下面的超参数来说，更高的值对于决策树算法更好吗？ E

1 用于拆分的样本量

2 树深

3 树叶样本

A、1 和 2

B、2 和 3

C、1 和 3

D、1、2 和 3

E、无法分辨

58. 在 k-均值算法中，以下哪个选项可用于获得全局最小？ D

1 尝试为不同的质心（centroid）初始化运行算法

2 调整迭代的次数

3 找到集群的最佳数量

A、2 和 3

B、1 和 3

C、1 和 2

D、以上所有

59. 列哪一项关于极大似然估计（MLE）的说法是正确的？ C

- 1.MLE 并不总是存在
- 2.MLE 一直存在
- 3.如果 MLE 存在，它可能不特异
- 4.如果 MLE 存在，它一定是特异的

- A、1 和 4
- B、2 和 3
- C、1 和 3
- D、2 和 4

60. 导出线性回归的参数时，我们做出下列哪种假定？（ D ）

- 1.因变量 y 和自变量 x 的真实关系是线性的
- 2.模型误差是统计独立的
- 3.误差通常服从一个平均值为零，标准差恒定的分布
- 4.自变量 x 是非随机的，无错的

- A、1,2 和 3
- B、1,3 和 4
- C、1 和 3
- D、以上都对

61. 假设我们用 Lasso 回归拟合一个有 100 个特征值($X_1, X_2 \dots X_{100}$)的数据集，现在，我们重新调节其中一个值，将它乘 10（将它视作 X_1 ），并再次拟合同一规则化参数。下列哪一项正确？（ B ）

- A、X1 很可能被模型排除
- B、X1 很可能被包含在模型内
- C、很难说
- D、都不对

62. 在线性回归模型中增加一个变量，下列哪一项是正确的？ (A)

- 1.R 方和调整 R 方都上升
- 2.R 方上升，调整 R 方下降
- 3.R 方和调整 R 方都下降
- 4.R 方下降，调整 R 方上升

- A、1 和 2
- B、1 和 3
- C、2 和 4
- D、以上都不对

63. Y 的预期值是关于变量 X(X_1, X_2, \dots, X_n)的线性函数, 回归线定义为 $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$, 下列陈述哪项正确? (D)

- 1.如果 X_i 的变化量为 ΔX_i , 其它为常量, 则 Y 的变化量为 $\beta_i \Delta X_i$, 常量 β_i 可以为正数或负数
- 2. β_i 的值都是一样的, 除非是其它 X 的 β_i
- 3.X 对 Y 预期值的总影响为每个分影响之和

提示: 特征值间相互独立, 互不干扰

- A、1 和 2
- B、1 和 3
- C、2 和 3
- D、1,2 和 3

64. 假设对数据提供一个逻辑回归模型，得到训练精度 X 和测试精度 Y 。在数据中加入新的特征值，则下列哪一项是正确的？(B)

提示：其余参数是一样的

- A、训练精度总是下降
- B、训练精度总是上升或不变
- C、测试精度总是下降
- D、测试精度总是上升或不变

65. 简单线性回归模型中（单自变量），如果改变输入变量 1 单元，输出变量会变化多少？

(D)

- A、1 单元
- B、无变化
- C、截距值
- D、斜率值

66. 关于 Ridge 回归，下列哪项正确？(A)

1. λ 为 0 时，模型作用类似于线性回归模型

2. λ 为 0 时，模型作用与线性回归模型不相像

3. 当 λ 趋向无穷，会得到非常小，趋近 0 的相关系数

4. 当 λ 趋向无穷，会得到非常大，趋近无穷的相关系数

A、1 和 3

B、1 和 4

C、2 和 3

D、2 和 4

67. 下列哪一项关于极大似然估计 (MLE) 的说法是正确的？©

1. MLE 并不总是存在

2. MLE 一直存在

3. 如果 MLE 存在，它可能不特异

4. 如果 MLE 存在，它一定是特异的

A、1 and 4

B、2 and 3

C、1 and 3

D、2 and 4

68. 假设我们用 Lasso 回归拟合一个有 100 个特征值(X_1, X_2, \dots, X_{100})的数据集，现在，我们重新调节其中一个值，将它乘 10 (将它视作 X_1)，并再次拟合同一正则化参数。下列哪一项正确？(B)

A、 X_1 很可能被模型排除

B、X1 很可能被包含在模型内

C、很难说

D、都不对

69. 我们可以通过一种叫“正规方程”的分析方法来计算线性回归的相关系数, 下列关于“正规方程”哪一项是正确的? (D)

1. 我们不必选择学习比率

2. 当特征值数量很大时会很慢

3. 不需要迭代

A、1 和 2

B、1 和 3

C、2 和 3

D、1, 2 和 3

70. 假设对数据提供一个逻辑回归模型, 得到训练精度 X 和测试精度 Y。在数据中加入新的特征值, 则下列哪一项是正确的? (A)

提示: 其余参数是一样的

1. 训练精度总是下降

2. 训练精度总是上升或不变

3. 测试精度总是下降

4. 测试精度总是上升或不变

A、只有 2

B、只有 1

C、只有 3

D、只有 4

71. 在进行聚类分析之前，给出少于所需数据的数据点，下面哪种方法最适合用于数据清理？ (A)

1 限制和增加变量

2 去除异常值

A、 1

B、 2

C、 1 和 2

D、都不能

72. 评估模型之后，得出模型存在偏差，下列哪种方法可能解决这一问题： (B)

A、减少模型中特征的数量

B、向模型中增加更多的特征

C、增加更多的数据

D、B 和 C

E、以上全是

73. 一个回归模型存在多重共线问题。在不损失过多信息的情况下，下列哪个操作不可取(A)

A、移除共线的两个变量

B、移除共线的两个变量其中一个

C、我们可以计算方差膨胀因子 (variance inflation factor)来检查存在的多重共线性并采取相应的措施

D、移除相关变量可能会导致信息的丢失, 为了保留这些变量, 我们可以使用岭回归(ridge)或 lasso 等回归方法对模型进行惩罚

74. 某超市研究销售纪录数据后发现, 买啤酒的人很大概率也会购买尿布, 这种属于数据挖掘的哪类问题? (A)

A. 关联规则发现 B. 聚类 C. 分类 D. 自然语言处理

75. 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? (B)

A. 分类 B. 聚类 C. 关联分析 D. 隐马尔可夫链

76. 已知一组数据的协方差矩阵 P , 下面关于主分量说法错误的是 (C)

主分量分析的最佳准则是对一组数据进行按一组正交基分解, 在只取相同数量分量的条件下, 以均方误差计算截尾误差最小

B、在经主分量分解后, 协方差矩阵成为对角矩阵

C、主分量分析就是 K-L 变换

D、主分量是通过求协方差矩阵的特征值得到

77. 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)

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/RBF 核代替线性核

78. 下面有关分类算法的准确率，召回率，F1 值的描述，错误的是？ (C)

准确率是检索出相关文档数与检索出的文档总数的比率，衡量的是检索系统的查准率

- B. 召回率是指检索出的相关文档数和文档库中所有的相关文档数的比率，衡量的是检索系统的查全率
- C. 正确率、召回率和 F 值取值都在 0 和 1 之间，数值越接近 0，查准率或查全率就越高
- D. 为了解决准确率和召回率冲突问题，引入了 F1 分数

79. 下列哪个不属于 CRF 模型对于 HMM 和 MEMM 模型的优势 (B)

- A. 特征灵活
- B. 速度快
- C. 可容纳较多上下文信息
- D. 全局最优

80. 以下哪个是常见的时间序列算法模型 (C)

- RSI
- B. MACD
- C. ARMA

D. KDJ

81. 以下几种模型方法属于判别式模型(Discriminative Model)的有 (A)

1)混合高斯模型 2)条件随机场模型

3)区分度训练 4)隐马尔科夫模型

2,3

B. 3,4

C. 1,4

D. 1,2

82. 对数几率回归 (logistics regression) 和一般回归分析有什么区别? (D)

对数几率回归是设计用来预测事件可能性的

B. 对数几率回归可以用来度量模型拟合程度

C. 对数几率回归可以用来估计回归系数

D. 以上所有

83. 对于 k 折交叉验证, 以下对 k 的说法正确的是 : (D)

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

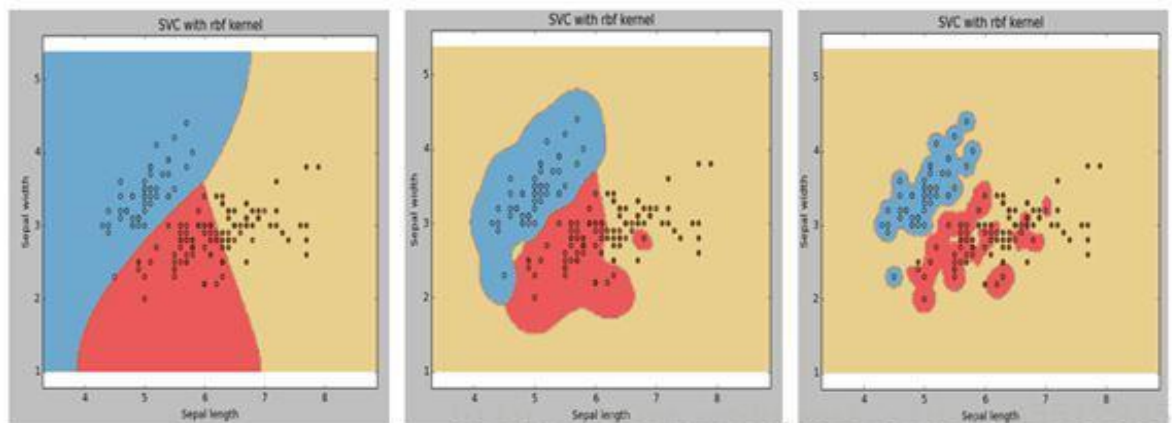
84. 如果 SVM 模型欠拟合, 以下方法哪些可以改进模型 : (A)

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)

85. 下图是同一个 SVM 模型, 但是使用了不同的径向基核函数的 gamma 参数, 依次是 g_1 , g_2 , g_3 , 下面大小比较正确的是 : (C)



$g_1 > g_2 > g_3$

B. $g_1 = g_2 = g_3$

C. $g_1 < g_2 < g_3$

D. $g_1 \geq g_2 \geq g_3$

E. $g_1 \leq g_2 \leq g_3$

86. 我们建立一个 5000 个特征, 100 万数据的机器学习模型. 我们怎么有效地应对这样的大数据训练 : (F)

A. 我们随机抽取一些样本, 在这些少量样本之上训练

- B. 我们可以试用在线机器学习算法
- C. 我们应用 PCA 算法降维, 减少特征数
- D. B 和 C
- E. A 和 B
- F. 以上所有

87. 对于随机森林和 GradientBoosting Trees, 下面说法正确的是: (A)

- 1.在随机森林的单个树中, 树和树之间是有依赖的, 而 GradientBoosting Trees 中的单个树之间是没有依赖的
- 2.这两个模型都使用随机特征子集, 来生成许多单个的树
- 3.我们可以并行地生成 GradientBoosting Trees 单个树, 因为它们之间是没有依赖的, GradientBoosting Trees 训练模型的表现总是比随机森林好

- A. 2
- B. 1 and 2
- C. 1, 3 and 4
- D. 2 and 4

88. 对于 PCA 说法正确的是: (A)

- 1.我们必须在使用 PCA 前规范化数据
- 2.我们应该选择使得模型有最大 variance 的主成分
- 3.我们应该选择使得模型有最小 variance 的主成分
- 4.我们可以使用 PCA 在低维度上做数据可视化

A. 1, 2 and 4

B. 2 and 4

C. 3 and 4

D. 1 and 3

E. 1, 3 and 4

89. 在有监督学习中，我们如何使用聚类方法？（B）

1.我们可以先创建聚类类别，然后在每个类别上用监督学习分别进行学习

2.我们可以使用聚类“类别 id”作为一个新的特征项，然后再用监督学习分别进行学习

3.在进行监督学习之前，我们不能新建聚类类别

4.我们不可以使用聚类“类别 id”作为一个新的特征项，然后再用监督学习分别进行学习

A. 2 和 4

B. 1 和 2

C. 3 和 4

D. 1 和 3

90. 以下说法正确的是：（C）

1.一个机器学习模型，如果有较高准确率，总是说明这个分类器是好的

2.如果增加模型复杂度，那么模型的测试错误率总是会降低

3.如果增加模型复杂度，那么模型的训练错误率总是会降低

A. 1

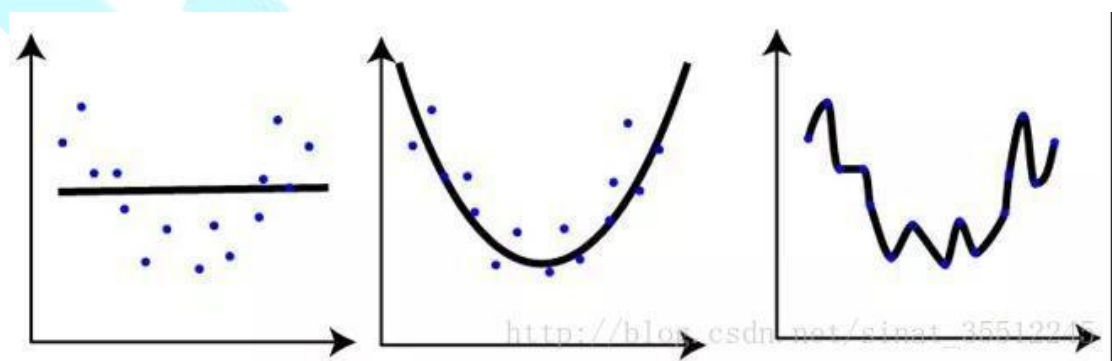
- B. 2
- C. 3
- D. 1 and 3

91. 对于线性回归模型，包括附加变量在内，以下的可能正确的是：(D)

- 1.R-Squared 和 Adjusted R-squared 都是递增的
- 2.R-Squared 是常量的， Adjusted R-squared 是递增的
- 3.R-Squared 是递减的， Adjusted R-squared 也是递减的
- 4.R-Squared 是递减的， Adjusted R-squared 是递增的

- A. 1 和 2
- B. 1 和 3
- C. 2 和 4
- D. 以上都不是

92. 对于下面三个模型的训练情况， 下面说法正确的是：(C)



- 1.第一张图的训练错误与其余两张图相比，是最大的
- 2.最后一张图的训练效果最好，因为训练错误最小
- 3.第二张图比第一和第三张图鲁棒性更强，是三个里面表现最好的模型

4.第三张图相对前两张图过拟合了

5.三个图表现一样，因为我们还没有测试数据集

A. 1 和 3

B. 1 和 3

C. 1, 3 和 4

D. 5

93. 下面哪个/些选项对 K 折交叉验证的描述是正确的？ (D)

1.增大 K 将导致交叉验证结果时需要更多的时间

2.更大的 K 值相比于小 K 值将对交叉验证结构有更高的信心

3.如果 $K=N$ ，那么其称为留一交叉验证，其中 N 为验证集中的样本数量

A. 1 和 2

B. 2 和 3

C. 1 和 3

D. 1、2 和 3

94. 给定三个变量 X, Y, Z。 (X, Y)、(Y, Z) 和 (X, Z) 的 Pearson 相关性系数分别为 C_1 、

C_2 和 C_3 。现在 X 的所有值加 2 (即 $X+2$)，Y 的全部值减 2 (即 $Y-2$)，Z 保持不变。

那么运算之后的 (X, Y)、(Y, Z) 和 (X, Z) 相关性系数分别为 D_1 、 D_2 和 D_3 。现在试问 D_1 、

D_2 、 D_3 和 C_1 、 C_2 、 C_3 之间的关系是什么？ (E)

A. $D_1 = C_1$, $D_2 < C_2$, $D_3 > C_3$

B. $D_1 = C_1$, $D_2 > C_2$, $D_3 > C_3$

C. $D1 = C1, D2 > C2, D3 < C3$

D. $D1 = C1, D2 < C2, D3 < C3$

E. $D1 = C1, D2 = C2, D3 = C3$

95. 两个变量的 Pearson 相关性系数为零，但这两个变量的值同样可以相关。(A)

A. 正确

B. 错误

96. 下面有关序列模式挖掘算法的描述，错误的是？(C)

A. AprioriAll 算法和 GSP 算法都属于 Apriori 类算法，都要产生大量的候选序列

B. FreeSpan 算法和 PrefixSpan 算法不生成大量的候选序列以及不需要反复扫描原数据库

C. 在时空的执行效率上，FreeSpan 比 PrefixSpan 更优

D. 和 AprioriAll 相比，GSP 的执行效率比较高

97. 类域界面方程法中，不能求线性不可分情况下分类问题近似或精确解的方法是？(D)

A. 伪逆法-径向基 (RBF) 神经网络的训练算法，就是解决线性不可分的情况

B. 基于二次准则的 H-K 算法：最小均方差准则下求得权矢量，二次准则解决非线性问题

C. 势函数法 - 非线性

D. 感知器算法 - 线性分类算法

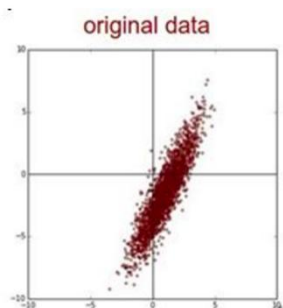
98. 下列方法中，不可以用于特征降维的方法包括 (E)

A. 主成分分析 PCA

- B. 线性判别分析 LDA
- C. 深度学习 SparseAutoEncoder
- D. 矩阵奇异值分解 SVD
- E. 最小二乘法 LeastSquares

困难

1. 假设在训练中我们突然遇到了一个问题，在几次循环之后，误差瞬间降低。你认为数据有问题，于是你画出了数据并且发现也许是数据的偏度过大造成了这个问题。

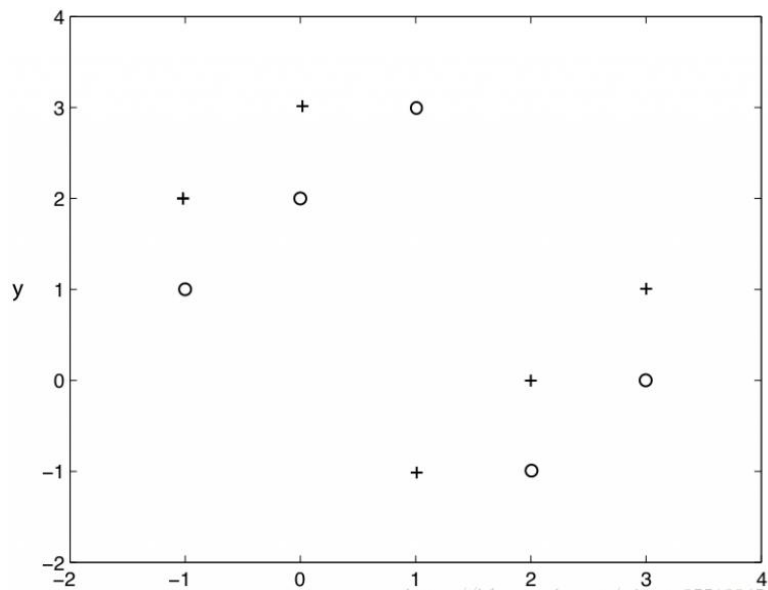


你打算怎么做来处理这个问题？ (D)

- A. 对数据作归一化
- B. 对数据取对数变化
- C. 都不对
- D. 对数据作主成分分析(PCA)和归一化

2. 使用 $k=1$ 的 KNN 算法, 下图二类分类问题, “+” 和 “o” 分别代表两个类, 那么, 用

仅拿出一个测试样本的交叉验证方法, 交叉验证的错误率是多少 : (B)



- A. 0%
- B. 100%
- C. 0% 到 100%
- D. 以上都不是

3. 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. $2x-y=0$

4. 以 $P(w)$ 表示词条 w 的概率, 假设已知 $P(\text{南京}) = 0.8$, $P(\text{市长}) = 0.6$, $P(\text{江大桥}) = 0.4$: $P(\text{南京市}) = 0.3$, $P(\text{长江大桥}) = 0.5$: 如果假设前后两个词的出现是独立的, 那

么分词结果就是 (B)

- A、南京市*长江*大桥
- B、南京*市长*江大桥
- C、南京市长*江大桥
- D、南京市*长江大桥

5. 一个二进制源 X 发出符号集为 $\{-1,1\}$, 经过离散无记忆信道传输, 由于信道中噪音的存在, 接收端 Y 收到符号集为 $\{-1,1,0\}$ 。已知 $P(x=-1)=1/4$, $P(x=1)=3/4$, $P(y=-1|x=-1)=4/5$, $P(y=0|x=-1)=1/5$, $P(y=1|x=1)=3/4$, $P(y=0|x=1)=1/4$, 求条件熵 $H(Y|X)$ (A)

- A、0.2375
- B、0.3275
- C、0.5273
- D、0.5372

6. Naive Bayes 是一种特殊的 Bayes 分类器, 特征变量是 X , 类别标签是 C , 它的一个假定是 (C)

- A、各类别的先验概率 $P(C)$ 是相等的
- B、以 0 为均值, $\sqrt{2}/2$ 为标准差的正态分布
- C、特征变量 X 的各个维度是类别条件独立随机变量
- D、 $P(X|C)$ 是高斯分布

7. 假定你使用 SVM 学习数据 X , 数据 X 里面有些点存在错误。现在如果你使用一个二次核函数, 多项式阶数为 2, 使用松弛变量 C 作为超参之一。

如果使用较小的 C (C 趋于 0), 则: A

- A、误分类
- B、正确分类
- C、不确定
- D、以上均不正确

8. 对于投影数据为 $((\sqrt{2}), (0), (\sqrt{2}))$ 。现在如果在二维空间中重建, 并将它们视为原始数据点的重建, 那么重建误差是多少? A

- A、0%
- B、10%
- C、30%
- D、40%

9. X_i 和 X_j 是较高维度表示中的两个不同点, 其中 Y_i 和 Y_j 是较低维度中的 X_i 和 X_j 的表示。

C

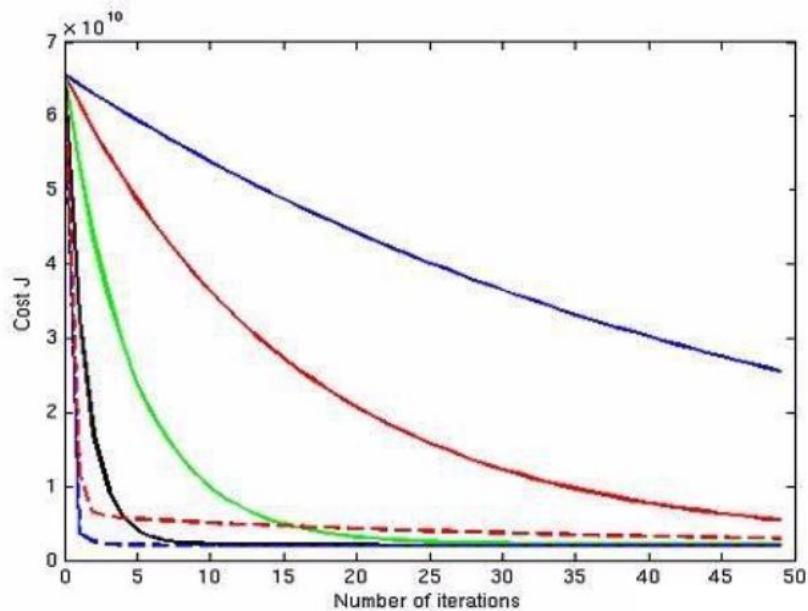
1. 数据点 X_i 与数据点 X_j 的相似度是条件概率 $p(j|i)$ 。

2. 数据点 Y_i 与数据点 Y_j 的相似度是条件概率 $q(j|i)$ 。

对于在较低维度空间中的 X_i 和 X_j 的完美表示, 以下哪一项必须是正确的?

- A、 $p(j|i) = 0, q(j|i) = 1$
- B、 $p(j|i)$
- C、 $p(j|i) = q(j|i)$
- D、 $P(j|i) > q(j|i)$

10. 如下逻辑回归图显示了 3 种不同学习速率值的代价函数和迭代次数之间的关系 (不同的颜色在不同的学习速率下显示不同的曲线)。



为了参考而保存图表后,忘记其中不同学习速率的值。现在需要曲线的倾斜率值之间的关系。

以下哪一个是正确的? C

注:

1.蓝色的学习率是 L1

2.红色的学习率是 L2

3.绿色学习率为 L3

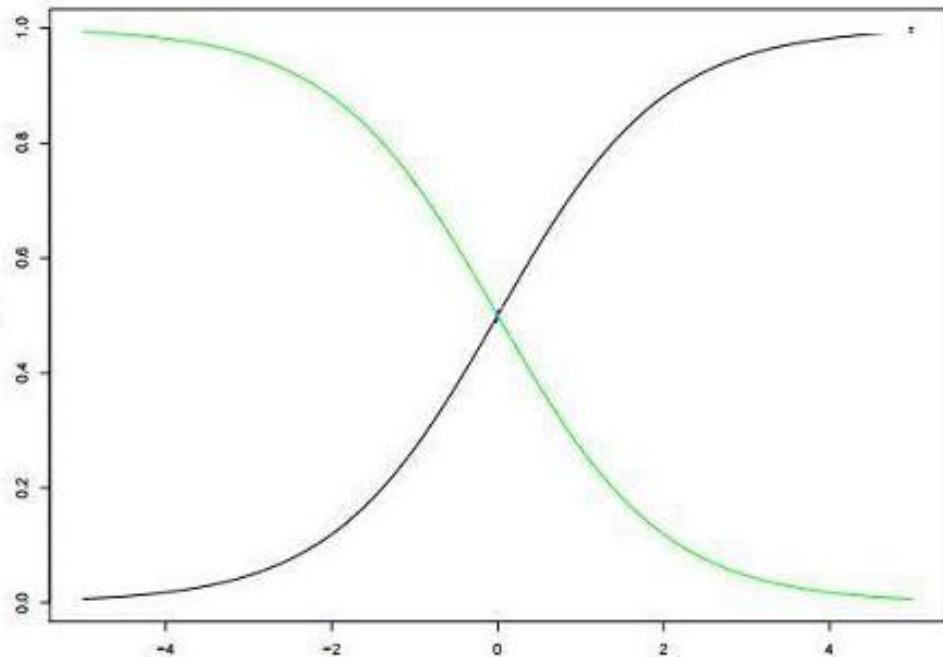
A、 $L1 > L2 > L3$

B、 $L1 = L2 = L3$

C、 L1

D、 都不是

12. 以下是两种不同的对数模型，分别为 β_0 和 β_1 。

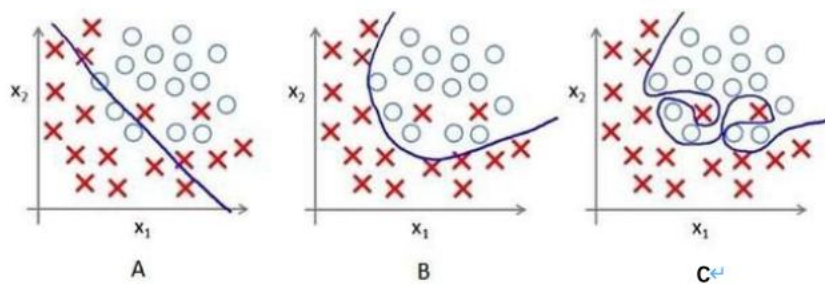


对于两种对数模型（绿色，黑色）的 β_0 和 β_1 值，下列哪一项是正确的？ B

注： $Y = \beta_0 + \beta_1 * X$ 。其中 β_0 是截距， β_1 是系数。

- A、绿色的 β_1 大于黑色
- B、绿色的 β_1 小于黑色
- C、两种颜色的 β_1 相同

13. 下面是三个散点图（A，B，C，从左到右）和和手绘的逻辑回归决策边界。

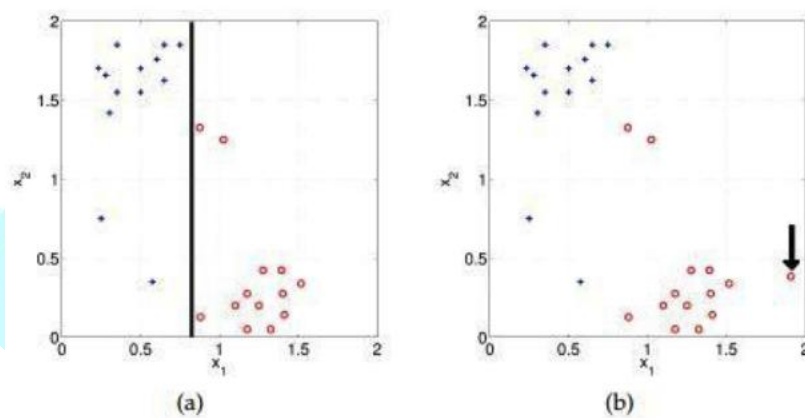


根据可视化后的结果，能得出什么结论？ A

- 1.与第二和第三图相比，第一幅图中的训练误差最大。
- 2.该回归问题的最佳模型是最后（第三个）图，因为它具有最小的训练误差（零）。
- 3.第二个模型比第一个和第三个更强，它在不可见数据中表现最好。
- 4.与第一种和第二种相比，第三种模型过度拟合了。
- 5.所有的模型执行起来都一样，因为没有看到测试数据。

- A、1 和 3
- B、2 和 3
- C、1,3 和 4
- D、5

14. 使用 high(infinite) regularisation 时偏差会如何变化?

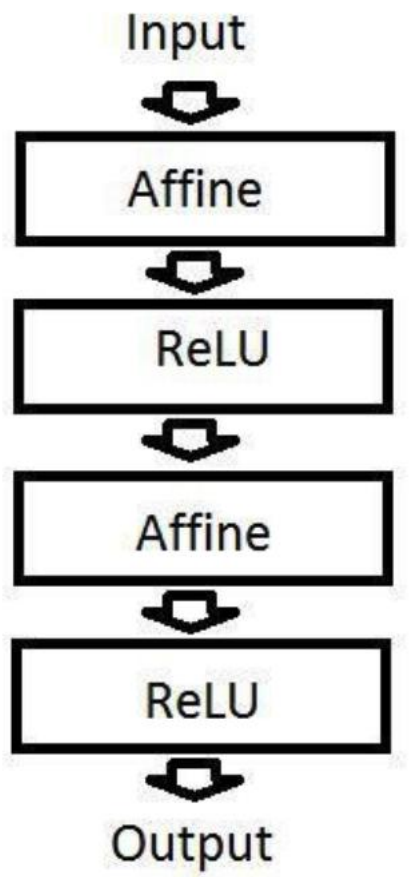


有散点图 “a” 和 “b” 两类（蓝色为正，红色为负）。在散点图 “a” 中，使用了逻辑回归

（黑线是决策边界）对所有数据点进行了正确分类。 A

- A、偏差很大
- B、偏差很小
- C、不确定
- D、都不是

15. 假设有一个如下定义的神经网络：



如果我們去掉 ReLU 层，这个神经网络仍能够处理非线性函数，这种说法是 B

A、正确的

B、错误的

16. 假定你想将高维数据映射到低维数据中，那么最出名的降维算法是 PCA 和 t-SNE。

现在你将这两个算法分别应用到数据「X」上，并得到数据集「X_projected_PCA」，

「X_projected_tSNE」。

下面哪一项对「X_projected_PCA」和「X_projected_tSNE」的描述是正确的？ B

A、X_projected_PCA 在最近邻空间能得到解释

B、X_projected_tSNE 在最近邻空间能得到解释

C、两个都在最近邻空间能得到解释

D、两个都不能在最近邻空间得到解释

17. 为了得到和 SVD 一样的投射 (projection), 你需要在 PCA 中怎样做? A

A、将数据转换成零均值

B、将数据转换成零中位数

C、无法做到

D、以上方法不行

18. 假设我们有一个数据集, 在一个深度为 6 的决策树的帮助下, 它可以使用 100% 的精确度被训练。现在考虑一下两点, 并基于这两点选择正确的选项。 A

注意: 所有其他超参数是相同的, 所有其他因子不受影响。

1 深度为 4 时将有高偏差和低方差

2 深度为 4 时将有低偏差和低方差

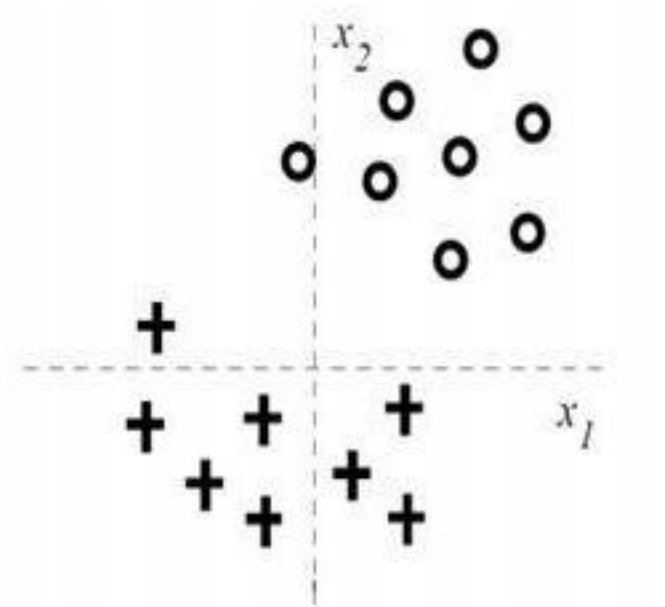
A、只有 1

B、只有 2

C、1 和 2

D、没有一个

19. 假设你被给到以下数据, 你想要在给定的两个类别中使用 logistic 回归模型对它进行分类。



你正在使用带有 L1 正则化的 logistic 回归，其中 C 是正则化参数， w_1 和 w_2 是 x_1 和 x_2 的系数。

$$\sum_{i=1}^n \log P(y_i | x_i, w_0, w_1, w_2) - C(|w_1| + |w_2|).$$

当你把 C 值从 0 增加至非常大的值时，下面哪个选项是正确的？ B

- A、第一个 w_2 成了 0，接着 w_1 也成了 0
- B、第一个 w_1 成了 0，接着 w_2 也成了 0
- C、 w_1 和 w_2 同时成了 0
- D、即使在 C 成为大值之后， w_1 和 w_2 都不能成 0

20. 假设存在一个黑箱算法，其输入为有多个观察 ($t_1, t_2, t_3, \dots, t_n$) 的训练数据和一个新的观察 (q_1)。该黑箱算法输出 q_1 的最近邻 t_i 及其对应的类别标签 c_i 。你可以将这个黑箱算法看作是一个 1-NN (1-最近邻)

我们不使用 1-NN 黑箱，而是使用 j-NN($j > 1$) 算法作为黑箱。为了使用 j-NN 寻找

k-NN，下面哪个选项是正确的？ A

A、j 必须是 k 的一个合适的因子

B、 $j > k$

C、不能办到

21. 假设你正在做一个项目，它是一个二元分类问题。你在数据集上训练一个模型，并在验证数据集上得到混淆矩阵。基于上述混淆矩阵，下面哪个选项会给你正确的预测。 C

1 精确度是~0.91

2 错误分类率是~0.91

3 假正率 (False correct classification) 是~0.95

4 真正率 (True positive rate) 是~0.95

A、1 和 3

B、2 和 4

C、1 和 4

D、2 和 3

22. Y 的预期值是关于变量 X(X_1, X_2, \dots, X_n) 的线性函数, 回归线定义为 $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$, 下列陈述哪项正确? (D)

1. 如果 X_i 的变化量为 ΔX_i , 其它为常量, 则 Y 的变化量为 $\beta_i \Delta X_i$, 常量 β_i 可以为正数或负数

2. β_i 的值都是一样的, 除非是其它 X 的 β_i

3. X 对 Y 预期值的总影响为每个分影响之和

提示：特征值间相互独立，互不干扰

- A、1 和 2
- B、1 和 3
- C、2 和 3
- D、1,2 和 3

23. Nave Bayes 是一种特殊的 Bayes 分类器,特征变量是 X ,类别标签是 C ,它的一个假定是:(C)

- A 各类别的先验概率 $P(C)$ 是相等的
- B.以 0 为均值, $\sqrt{2}/2$ 为标准差的正态分布
- C.特征变量 X 的各个维度是类别条件独立随机变量
- D. $P(X|C)$ 是高斯分布

24. 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 : (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

25. 在一个 n 维的空间中, 最好的检测 outlier(离群点)的方法是: (C)

- A 作正态分布概率图
- B. 作盒形图

C. 马氏距离

D. 作散点图

26. 已知一组数据的协方差矩阵 P , 下面关于主分量说法错误的是(C)

A 主分量分析的最佳准则是对一组数据进行按一组正交基分解, 在只取相同数量分量的条件下, 以均方误差计算截尾误差最小

B. 在经主分量分解后, 协方差矩阵成为对角矩阵

C. 主分量分析就是 K-L 变换

D. 主分量是通过求协方差矩阵的特征值得到

27. 回归模型中存在多重共线性, 你如何解决这个问题?

1. 去除这两个共线性变量

2. 我们可以先去除一个共线性变量

3. 计算 VIF(方差膨胀因子), 采取相应措施

4. 为了避免损失信息, 我们可以使用一些正则化方法, 比如, 岭回归和 lasso 回归.

以下哪些是对的: (D)

A 1

B. 2

C. 2 和 3

D. 2, 3 和 4

28. 模型的高 bias 是什么意思, 我们如何降低它? (B)

- A 在特征空间中减少特征
- B. 在特征空间中增加特征
- C. 增加数据点
- D. B 和 C

29. 对于信息增益, 决策树分裂节点, 下面说法正确的是: (C)

- 1. 纯度高的节点需要更多的信息去区分
- 2. 信息增益可以用“1 比特-熵”获得
- 3. 如果选择一个属性具有许多归类值, 那么这个信息增益是有偏差的

- A 1
- B. 2
- C. 2 和 3
- D. 所有以上

30. 假设我们要解决一个二类分类问题, 我们已经建立好了模型, 输出是 0 或 1, 初始时设阈值为 0.5, 超过 0.5 概率估计, 就判别为 1, 否则就判别为 0; 如果我们现在用另一个大于 0.5 的阈值, 那么现在关于模型说法, 正确的是: (C)

- 1 模型分类的召回率会降低或不变
- 2. 模型分类的召回率会升高
- 3. 模型分类准确率会升高或不变
- 4. 模型分类准确率会降低

B. 2

C. 1 和 3

D. 2 和 4

31. “点击率问题”是这样一个预测问题, 99%的人是不会点击的, 而 1%的人是会点击进去的, 所以这是一个非常不平衡的数据集. 假设, 现在我们已经建了一个模型来分类, 而且有了 99%的预测准确率, 我们可以下的结论是 : (B)

A. 模型预测准确率已经很高了, 我们不需要做什么了

B. 模型预测准确率不高, 我们需要做点什么改进模型

C. 无法下结论

D. 以上都不对

32. 假如我们使用非线性可分的 SVM 目标函数作为最优化对象, 我们怎么保证模型线性可分? (C)

A. 设 $C=1$

B. 设 $C=0$

C. 设 $C=$ 无穷大

D. 以上都不对

33. 我们想要减少数据集中的特征数, 即降维. 选择以下适合的方案 : (D)

1.使用前向特征选择方法

2.使用后向特征排除方法

3.我们先把所有特征都使用, 去训练一个模型, 得到测试集上的表现. 然后我们去掉一个特征, 再去训练, 用交叉验证看看测试集上的表现. 如果表现比原来还要好, 我们可以去除这个特征

4.查看相关性表, 去除相关性最高的一些特征

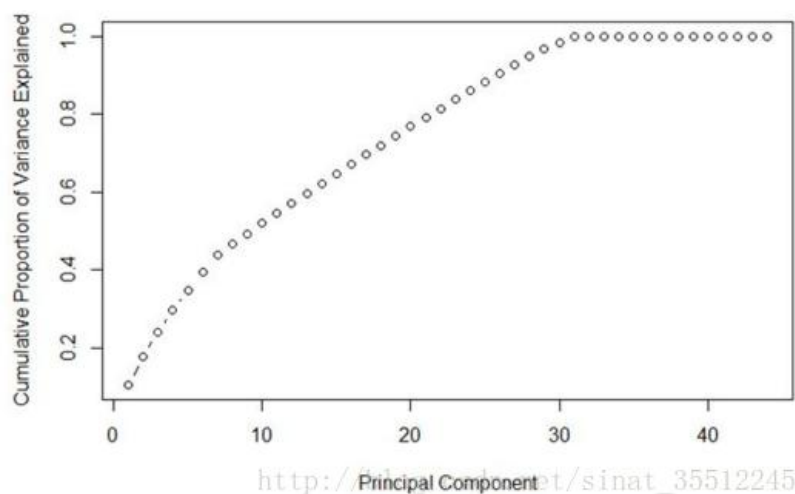
A. 1 和 2

B. 2, 3 和 4

C. 1, 2 和 4

D. All

34. 对于下图, 最好的主成分选择是多少 ? (B)



A. 7

B. 30

C. 35

D. 不确定

35. 对应 GradientBoosting tree 算法, 以下说法正确的是 : (C)

- 1.当增加最小样本分裂个数，我们可以抵制过拟合
- 2.当增加最小样本分裂个数，会导致过拟合
- 3.当我们减少训练单个学习器的样本个数，我们可以降低 variance
- 4.当我们减少训练单个学习器的样本个数，我们可以降低 bias

- A. 2 和 4
- B. 2 和 3
- C. 1 和 3
- D. 1 和 4

36. 如果一个训练好的模型在测试集上有 100%的准确率， 这是不是意味着在一个新的数据集上，也会有同样好的表现？ (B)

- A. 是的，这说明这个模型的泛化能力已经足以支持新的数据集合了
- B. 不对，依然后其他因素模型没有考虑到，比如噪音数据

37. 下面的交叉验证方法：

- i. 有放回的 Bootstrap 方法
- ii. 留一个测试样本的交叉验证
- iii. 5 折交叉验证
- iv. 重复两次的 5 折交叉验证

当样本是 1000 时，下面执行时间的顺序，正确的是：(B)

- A. $i > ii > iii > iv$
- B. $ii > iv > iii > i$

C. $iv > i > ii > iii$

D. $ii > iii > iv > i$

38. 变量选择是用来选择最好的判别器子集，如果要考虑模型效率，我们应该做哪些变量选择的考虑？：(C)

1.多个变量其实有相同的用处

2.变量对于模型的解释有多大作用

3.特征携带的信息

4.交叉验证

A. 1 和 4

B. 1, 2 和 3

C. 1,3 和 4

D. 以上所有

39. 为了得到和 SVD 一样的投射 (projection)，你需要在 PCA 中怎样做？(A)

A. 将数据转换成零均值

B. 将数据转换成零中位数

C. 无法做到

41. 假设你使用 log-loss 函数作为评估标准。下面这些选项，哪些是对作为评估标准的 log-loss 的正确解释。(D)

A.如果一个分类器对不正确的分类很自信，log-loss 会严重的批评它。

B. 对一个特别的观察而言，分类器为正确的类别分配非常小的概率，然后对 log-loss 的相应分布会非常大。

C. log-loss 越低，模型越好

D. 以上都是

42. 影推荐系统是以下哪些的应用实例：(E)

1 分类

2 聚类

3 强化学习

4 回归

A、只有 2

B、1 和 2

C、1 和 3

D、2 和 3

E、1 2 3

F、1 2 3 4

43. 对于一个图像识别问题(在一张照片里找出一只猫)，下面哪种神经网络可以更好地解决这个问题？(D)

A. 循环神经网络

B. 感知机

C. 多层感知机

D. 卷积神经网络

44. 当在卷积神经网络中加入池化层(pooling layer)时,变换的不变性会被保留,是吗? (C)

A. 不知道

B. 看情况

C. 是

D. 否

45. 如果我们用了一个过大的学习速率会发生什么? (D)

A. 神经网络会收敛

B. 不好说

C. 都不对

D. 神经网络不会收敛

46. 在一个神经网络中,下面哪种方法可以用来处理过拟合? (D)

A. Dropout

B. 分批归一化(Batch Normalization)

C. 正则化(regularization)

D. 都可以

47. 批规范化(Batch Normalization)的好处都有啥? (A)

A. 在将所有的输入传递到下一层之前对其进行归一化 (更改)

- B. 它将权重的归一化平均值和标准差
- C. 它是一种非常有效的反向传播(BP)方法
- D. 这些均不是

48. 下列哪个神经网络结构会发生权重共享? (D)

- A. 卷积神经网络
- B. 循环神经网络
- C. 全连接神经网络
- D. 选项 A 和 B

49. 构建一个神经网络, 将前一层的输出和它自身作为输入。下列哪一种架构有反馈连接?

(A)

- A. 循环神经网络
- B. 卷积神经网络
- C. 限制玻尔兹曼机
- D. 都不是

50. 下列哪项关于模型能力 (model capacity) 的描述是正确的? (指神经网络模型能拟合复杂函数的能力) (A)

- A. 隐藏层层数增加, 模型能力增加
- B. Dropout 的比例增加, 模型能力增加
- C. 学习率增加, 模型能力增加

D. 都不正确

51. 在训练神经网络时, 损失函数(loss)在最初的几个 epochs 时没有下降, 可能的原因是?

(A)

A. 学习率(learning rate)太低

B. 正则参数太高

C. 陷入局部最小值

D. 以上都有可能

52. 下列哪一项在神经网络中引入了非线性? (B)

A. 随机梯度下降

B. 修正线性单元 (ReLU)

C. 卷积函数

D. 以上都不正确

53. 下面哪项操作能实现跟神经网络中 Dropout 的类似效果? (B)

A. Boosting

B. Bagging

C. Stacking

D. Mapping

54. 已知:

- 大脑是有很多个叫做神经元的东西构成，神经网络是对大脑的简单的数学表达。
- 每一个神经元都有输入、处理函数和输出。
- 神经元组合起来形成了网络，可以拟合任何函数。
- 为了得到最佳的神经网络，我们用梯度下降方法不断更新模型

给定上述关于神经网络的描述，什么情况下神经网络模型被称为深度学习模型？（A）

- A. 加入更多层，使神经网络的深度增加
- B. 有维度更高的数据
- C. 当这是一个图形识别的问题时
- D. 以上都不正确

55. 一监狱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 4 种不同的人员：狱警，小偷，送餐员，其他。下面哪种学习方法最适合此种应用需求：（B）。

- A. 二分类问题
- B. 多分类问题
- C. 层次聚类问题
- D. k-中心点聚类问题
- E. 回归问题
- F. 结构分析问题

56. 梯度下降算法的正确步骤是什么？（D）

1. 计算预测值和真实值之间的误差
2. 重复迭代，直至得到网络权重的最佳值

3.把输入传入网络，得到输出值

4.用随机值初始化权重和偏差

5.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. 1, 2, 3, 4, 5

B. 5, 4, 3, 2, 1

C. 3, 2, 1, 5, 4

D. 4, 3, 1, 5, 2

57. 下列方法中，不可以用于特征降维的方法包括 C

A、主成分分析 PCA

B、线性判别分析 LDA

C、深度学习 SparseAutoEncoder

D、矩阵奇异值分解 SVD

58. 神经网络模型（Neural Network）因受人类大脑的启发而得名。神经网络由许多神经元（Neuron）组成，每个神经元接受一个输入，对输入进行处理后给出一个输出。请问下列关于神经元的描述中，哪一项是正确的？（E）

A.每个神经元只有一个输入和一个输出

B.每个神经元有多个输入和一个输出

C.每个神经元有一个输入和多个输出

D.每个神经元有多个输入和多个输出

E.上述都正确

59. 深度学习是当前很热门的机器学习算法, 在深度学习中, 涉及到大量的矩阵相乘, 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 $m \times n$, $n \times p$, $p \times q$, 且 $m < n < p < q$, 以下计算顺序效率最高的是 (A)

- A. (AB)C
- B. AC(B)
- C. A(BC)
- D. 所以效率都相同

60. 基于统计的分词方法为 (D)

- A、正向最大匹配法
- B、逆向最大匹配法
- C、最少切分
- D、条件随机场

61. 以下几种模型方法属于判别式模型(Discriminative Model)的有(A)

- 1)混合高斯模型
- 2)条件随机场模型
- 3)区分度训练
- 4)隐马尔科夫模型

A、 2,3

B、 3,4

C、1,4

D、1,2

62. 下面有关分类算法的准确率，召回率，F1 值的描述，错误的是？ C

A、准确率是检索出相关文档数与检索出的文档总数的比率，衡量的是检索系统的查准率

B、召回率是指检索出的相关文档数和文档库中所有的相关文档数的比率，衡量的是检索系统的查全率

C、正确率、召回率和 F 值取值都在 0 和 1 之间，数值越接近 0，查准率或查全率就越高

D、为了解决准确率和召回率冲突问题，引入了 F1 分数

63. 基于统计的分词方法为 (D)

A、正向最大匹配法

B、逆向最大匹配法

C、最少切分

D、条件随机场

64. 在大规模的语料中，挖掘词的相关性是一个重要的问题。以下哪一个信息不能用于确定两个词的相关性。 B

A、互信息

B、最大熵

C、卡方检验

D、最大似然比

65. 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

66. 中文同义词替换时, 常用到 Word2Vec, 以下说法错误的是 C

A、Word2Vec 基于概率统计

B、Word2Vec 结果符合当前预料环境

C、Word2Vec 得到的都是语义上的同义词

D、Word2Vec 受限于训练语料的数量和质量

67. 想象一下, 你有一个 28×28 的图片, 并使用输入深度为 3 和输出深度为 8 在上面运行一个 3×3 的卷积神经网络。注意, 步幅 padding 是 1, 你正在使用相同的填充(padding)。

当使用给定的参数时, 输出特征图的尺寸是多少? A

A、28 宽、28 高、8 深

B、13 宽、13 高、8 深

C、28 宽、13 高、8 深

D、13 宽、28 高、8 深

68. 输入图片大小为 200×200 , 依次经过一层卷积 (kernel size 5×5 , padding 1, stride

2), pooling (kernel size 3×3 , padding 0, stride 1), 又一层卷积 (kernel size 3×3 , padding 1, stride 1) 之后, 输出特征图大小为 (C):

- A. 95
- B. 96
- C. 97
- D. 98

69. 在选择神经网络的深度时, 下面那些参数需要考虑? (C)

- 1 神经网络的类型(如 MLP,CNN)
- 2 输入数据
- 3 计算能力(硬件和软件能力决定)
- 4 学习速率
- 5 映射的输出函数

- A. 1,2,4,5
- B. 2,3,4,5
- C. 都需要考虑
- D. 1,3,4,5

70. 当数据过大以至于无法在 RAM 中同时处理时, 哪种梯度下降方法更加有效? (A)

- A. 随机梯度下降法(Stochastic Gradient Descent)
- B. 不知道
- C. 整批梯度下降法(Full Batch Gradient Descent)

D. 都不是

71. 假设我们已经在 ImageNet 数据集(物体识别)上训练好了一个卷积神经网络。然后给这张卷积神经网络输入一张全白的图片。对于这个输入的输出结果为任何种类的物体的可能性都是一样的，对吗？ (D)

A. 对的

B. 不知道

C. 看情况

D. 不对

72. 假设你需要调整参数来最小化代价函数(cost function), 可以使用下列哪项技术? (D)

A. 穷举搜索

B. 随机搜索

C. Bayesian 优化

D. 以上任意一种

73. 对于神经网络的说法, 下面正确的是 : (A)

1.增加神经网络层数, 可能会增加测试数据集的分类错误率

2.减少神经网络层数, 总是能减小测试数据集的分类错误率

3.增加神经网络层数, 总是能减小训练数据集的分类错误率

A. 1

B. 1 和 3

C. 1 和 2

D. 2

74. 想像正在处理文本数据，使用单词嵌入 (Word2vec) 表示使用的单词。在单词嵌入中，最终会有 1000 维。现在想减小这个高维数据的维度，这样相似的词应该在最邻近的空间中具有相似的含义。在这种情况下，您最有可能选择以下哪种算法？ A

A. t-SNE

B. PCA

C. LDA

D.都不是

75. 隐马尔可夫模型 (HMM)，设其观察值

空间为

$$O = \{o_1, o_2, \dots, o_N\}$$

状态空间为

$$S = \{s_1, s_2, \dots, s_K\}$$

如果用维特比算法(Viterbi algorithm)进行解码，时间复杂度为 (D)

A、 $O(NK)$

B、 $O(NK^2)$

C、 $O(N^2K)$

D、以上都不是

76. (假设 $\text{precision} = TP / (TP + FP)$, $\text{recall} = TP / (TP + FN)$ 。) 在二分类问题中, 当测试集的正例和负例数量不均衡时, 以下评价方案哪个是相对不合理的 (A)

A、Accuracy: $(TP + TN) / \text{all}$

B、F-value: $2 * \text{recall} * \text{precision} / (\text{recall} + \text{precision})$

C、G-mean: $\sqrt{\text{precision} * \text{recall}}$

D、AUC: 曲线下面积

77. 语言模型的参数估计经常使用 MLE (最大似然估计)。面临的一个问题是没有出现的项概率为 0, 这样会导致语言模型的效果不好。为了解决这个问题, 需要使用 (A)

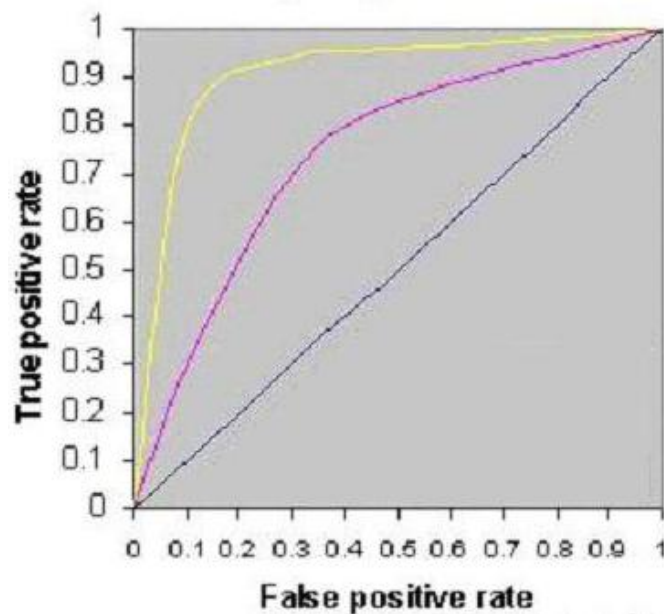
A、平滑

B、去噪

C、随机插值

D、增加白噪音

78. 下图显示了三个逻辑回归模型的 AUC-ROC 曲线。不同的颜色表示不同超参数值的曲线。以下哪个 AUC-ROC 会给出最佳结果? A



- A、黄色
- B、粉红色
- C、紫色
- D、都相同

79. 假定你现在解决一个有着非常不平衡类别的分类问题，即主要类别占据了训练数据的 99%。现在你的模型在测试集上表现为 99% 的准确度。那么下面哪一项表述是正确的？ A

- 1 准确度并不适合于衡量不平衡类别问题
- 2 准确度适合于衡量不平衡类别问题
- 3 精确率和召回率适合于衡量不平衡类别问题
- 3 精确率和召回率不适合于衡量不平衡类别问题

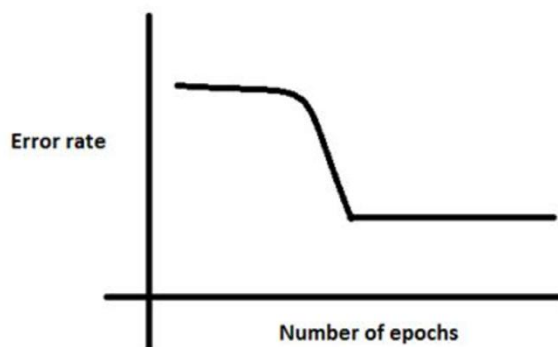
- A、1 and 3
- B、1 and 4
- C、2 and 3
- D、2 and 4

80. 考虑某个具体问题，你可能只有少量数据来解决这个问题。不过幸运的是你有一个类似问题已经预先训练好的神经网络。可以用下面哪种方法来利用这个预先训练好的网络？

(C)

- A. 把除了最后一层外所有的层都冻住，重新训练最后一层
- B. 对新数据重新训练整个模型
- C. 只对最后几层进行调参(fine tune)
- D. 对每一层模型进行评估，选择其中的少数来用

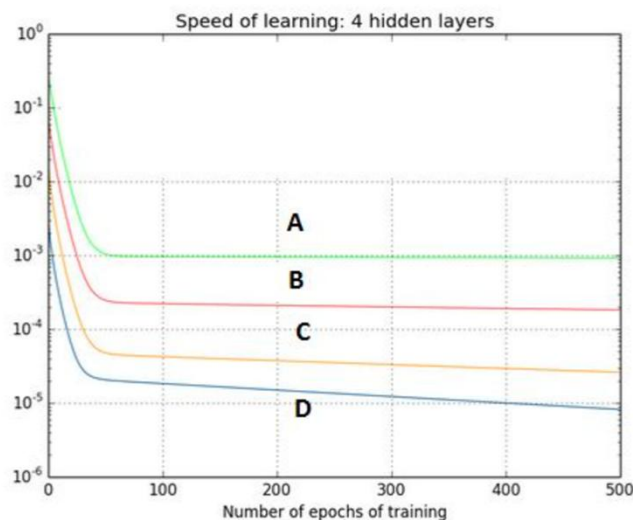
81. 下图显示，当开始训练时，误差一直很高，这是因为神经网络在往全局最小值前进之前一直被卡在局部最小值里。为了避免这种情况，我们可以采取下面哪种策略？ (A)



- A. 改变学习速率，比如一开始的几个训练周期不断更改学习速率
- B. 一开始将学习速率减小 10 倍，然后用动量项(momentum)
- C. 增加参数数目，这样神经网络就不会卡在局部最优处
- D. 其他都不对

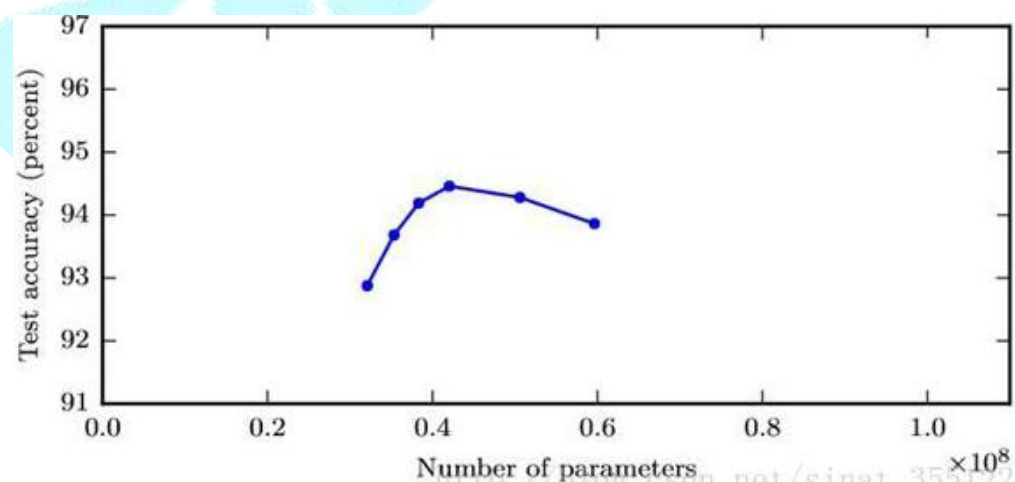
82. 下图是一个利用 sigmoid 函数作为激活函数的含四个隐藏层的神经网络训练的梯度下

降图。这个神经网络遇到了梯度消失的问题。下面哪个叙述是正确的？ (A)



- A. 第一隐藏层对应 D, 第二隐藏层对应 C, 第三隐藏层对应 B, 第四隐藏层对应 A
- B. 第一隐藏层对应 D, 第二隐藏层对应 C, 第三隐藏层对应 B, 第四隐藏层对应 A
- C. 第一隐藏层对应 A, 第二隐藏层对应 B, 第三隐藏层对应 C, 第四隐藏层对应 D
- D. 第一隐藏层对应 B, 第二隐藏层对应 D, 第三隐藏层对应 C, 第四隐藏层对应 A

83. 下图显示了训练过的 3 层卷积神经网络准确度, 与参数数量(特征核的数量)的关系。



从图中趋势可见, 如果增加神经网络的宽度, 精确度会增加到一个特定阈值后, 便开始降低。

造成这一现象的可能原因是什么？ (C)

- A. 即使增加卷积核的数量, 只有少部分的核会被用作预测

- B. 当卷积核数量增加时，神经网络的预测能力（Power）会降低
- C. 当卷积核数量增加时，它们之间的相关性增加(correlate)，导致过拟合
- D. 以上都不正确

85. 想像正在处理文本数据，使用单词嵌入（Word2vec）表示使用的单词。在单词嵌入中，最终会有 1000 维。现在想减小这个高维数据的维度，这样相似的词应该在最邻近的空间中具有相似的含义。在这种情况下，您最有可能选择以下哪种算法？ A

- A、t-SNE
- B、PCA
- C、LDA
- D、都不是

86. 你可以使用不同的标准评估二元分类问题的表现，例如准确率、log-loss、F-Score。

让我们假设你使用 log-loss 函数作为评估标准。 D

下面这些选项，哪个 / 些是对作为评估标准的 log-loss 的正确解释。

$$\log Loss = \frac{-1}{N} \sum_{i=1}^N (y_i (\log p_i) + (1 - y_i) \log(1 - p_i))$$

- 1 如果一个分类器对不正确的分类很自信，log-loss 会严重的批评它。
- 2 对一个特别的观察而言，分类器为正确的类别分配非常小的概率，然后对 log-loss 的相应分布会非常大。

3 log-loss 越低，模型越好。

- A、1 和 3
- B、2 和 3
- C、1 和 2
- D、1、2、3

二、多选题 (简单共 12 题 中等共 11 题 困难共 2 题 一共 25 题)

简单

1. 影响聚类算法效果的主要原因有：(ABC)
 - A. 特征选取
 - B. 模式相似性测度
 - C. 分类准则
 - D. 已知类别的样本质量
2. 在统计模式分类问题中，当先验概率未知时，可以使用 (BD)
 - A. 最小损失准则；
 - B. 最小最大损失准则；
 - C. 最小误判概率准则；
 - D. N-P 判决

3. 下列哪些方法可以用来对高维数据进行降维 (A B C D E F)

- A. LASSO
- B. 主成分分析法
- C. 聚类分析
- D. 小波分析法
- E. 线性判别法
- F. 拉普拉斯特征映射

4. 处理类别型特征时，事先不知道分类变量在测试集中的分布。要将 one-hot encoding (独热码) 应用到类别型特征中。那么在训练集中将独热码应用到分类变量可能要面临的困难是什么？(A、B)

- A. 分类变量所有的类别没有全部出现在测试集中
- B. 类别的频率分布在训练集和测试集是不同的
- C. 训练集和测试集通常会有一样的分布

5. 如果以特征向量的相关系数作为模式相似性测度，则影响聚类算法结果的主要因素有 (BC)

- A. 已知类别样本质量；
- B. 分类准则；
- C. 特征选取；
- D. 量纲

6. 影响基本 K-均值算法的主要因素有(ABD)

- A. 样本输入顺序;
- B. 模式相似性测度;
- C. 聚类准则;
- D. 初始类中心的选取

7. 以下哪个属于线性分类器最佳准则? (ACD)

- A. 感知准则函数
- B. 贝叶斯分类
- C. 支持向量机
- D. Fisher 准则

8. 以下说法中正确的是 (BD) ?

- A. SVM 对噪声(如来自其他分布的噪声样本)鲁棒
- B. 在 AdaBoost 算法中,所有被分错的样本的权重更新比例相同
- C. 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
- D. 给定 n 个数据点,如果其中一半用于训练,一般用于测试,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

9. 机器学习中 L1 正则化和 L2 正则化的区别是? (AD)

- A. 使用 L1 可以得到稀疏的权值

- B. 使用 L1 可以得到平滑的权值
- C. 使用 L2 可以得到稀疏的权值
- D. 使用 L2 可以得到平滑的权值

10. 位势函数法的积累势函数 $K(x)$ 的作用相当于 Bayes 判决中的 (AD)

- A. 后验概率
- B. 先验概率
- C. 类概率密度
- D. 类概率密度与先验概率的乘积

11. 下列方法中，可以用于特征降维的方法包括 (ABCD)

- A. 主成分分析 PCA
- B. 线性判别分析 LDA
- C. 深度学习 SparseAutoEncoder
- D. 矩阵奇异值分解 SVD
- E. 最小二乘法 LeastSquares

12. 数据清理中，处理缺失值的方法是？ (ABCD)

- A. 估算
- B. 整例删除
- C. 变量删除
- D. 成对删除

中等

1. 下面哪些是基于核的机器学习算法?(BCD)

- A. Expectation Maximization (EM) (最大期望算法)
- B. Radial Basis Function (RBF) (径向基核函数)
- C. Linear Discriminate Analysis (LDA) (主成分分析法)
- D. Support Vector Machine (SVM) (支持向量机)

2. 在统计模式识分类问题中, 当先验概率未知时, 可以使用(BC)?

- A. 最小损失准则
- B. N-P 判决
- C. 最小最大损失准则
- D. 最小误判概率准则

3. 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,打乱顺序参与分类
- 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

4. 关于线性回归的描述,以下正确的有: (ACEF)

- A. 基本假设包括随机干扰项是均值为 0,方差为 1 的标准正态分布
- B. 基本假设包括随机干扰项是均值为 0 的同方差正态分布
- C. 在违背基本假设时,普通最小二乘法估计量不再是最佳线性无偏估计量
- D. 在违背基本假设时,模型不再可以估计
- E. 可以用 DW 检验残差是否存在序列相关性
- F. 多重共线性会使得参数估计值方差减小

5. 基于二次准则函数的 H-K 算法较之于感知器算法的优点是 (BD) ?

- A. 计算量小
- B. 可以判别问题是否线性可分
- C. 其解完全适用于非线性可分的情况
- D. 其解的适应性更好

6. 欧式距离具有 (AB); 马式距离具有 (ABCD)。

- A. 平移不变性;
- B. 旋转不变性;
- C. 尺度缩放不变性;
- D. 不受量纲影响的特性

7. 假定目标变量的类别非常不平衡,即主要类别占据了训练数据的 99%。现在你的模型在

测试集上表现为 99% 的准确度。那么下面哪一项表述是正确的？（A、C）

- A. 准确度并不适合于衡量不平衡类别问题
- B. 准确度适合于衡量不平衡类别问题
- C. 精确率和召回率适合于衡量不平衡类别问题
- D. 精确率和召回率不适合于衡量不平衡类别问题

8. 鉴别了多元共线特征。那么下一步可能的操作是什么？（B、C）

- A. 移除两个共线变量
- B. 不移除两个变量，而是移除一个
- C. 移除相关变量可能会导致信息损失，可以使用带罚项的回归模型（如 ridge 或 lasso regression）。

9. 下面哪些对「类型 1 (Type-1)」和「类型 2 (Type-2)」错误的描述是正确的？（A、C）

- A. 类型 1 通常称之为假正类，类型 2 通常称之为假负类。
- B. 类型 2 通常称之为假正类，类型 1 通常称之为假负类。
- C. 类型 1 错误通常在其是正确的情况下拒绝假设而出现。

10. 模式识别中，马式距离较之于欧式距离的优点是（CD）

- A. 平移不变性；
- B. 旋转不变性；
- C. 尺度不变性；
- D. 考虑了模式的分布

11. 影响聚类算法结果的主要因素有 (BCD)

A. 已知类别的样本质量;

B. 分类准则;

C. 特征选取;

D. 模式相似性测度

困难

1. 假定某同学使用 Naive Bayesian (NB) 分类模型时, 不小心将训练数据的两个维度搞重复了, 那么关于 NB 的说法中正确的是: (BD)

A. 这个被重复的特征在模型中的决定作用会被加强

B. 模型效果相比无重复特征的情况下精确度会降低

C. 如果所有特征都被重复一遍, 得到的模型预测结果相对于不重复的情况下的模型预测结果一样。

D. 当两列特征高度相关时, 无法用两列特征相同时所得到的结论来分析问题

E. NB 可以用来做最小二乘回归

F. 以上说法都不正确

2. 隐马尔可夫模型三个基本问题以及相应的算法说法正确的是 (ABC)

A. 评估—前向后向算法

B. 解码—维特比算法

C. 学习—Baum-Welch 算法

D. 学习—前向后向算法

