1. 关于支持向量机 SVM,下列说法错误的是（ ）

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

2. 关于 Logit 回归和 SVM 不正确的是（ ）

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。A 错误

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. 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/RBF 核代替线性核

4. 关于 Logit 回归和 SVM 不正确的是（ ）。

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

5. 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（ ）

A. 2x+y=4

B. x+2y=5

C. x+2y=3

D. 2x-y=0

6. 关于支持向量机 SVM,下列说法错误的是（ ）

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 1/||w||，||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

7. 下列不是 SVM 核函数的是：（ ）

A. 多项式核函数

B. Logistic 核函数

C. 径向基核函数

D. Sigmoid 核函数

8. 带核的 SVM 为什么能分类非线性问题？

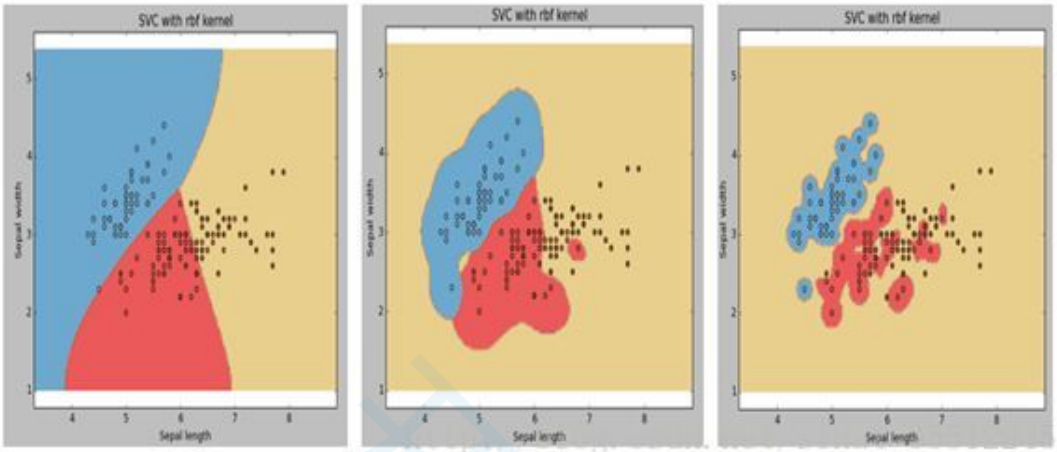
9. 如果 SVM 模型欠拟合，以下方法哪些可以改进模型 ：（ ）

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)

10. 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 g1, g2, g3，下面大小比较正确的是 ：（ ）



A. g1 > g2 > g3

B. g1 = g2 = g3

C. g1 < g2 < g3

D. g1 >= g2 >= g3

E. g1 <= g2 <= g3

11. 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？（ ）

A. 设 C=1

B. 设 C=0

C. 设 C=无穷大

D. 以上都不对

12. 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

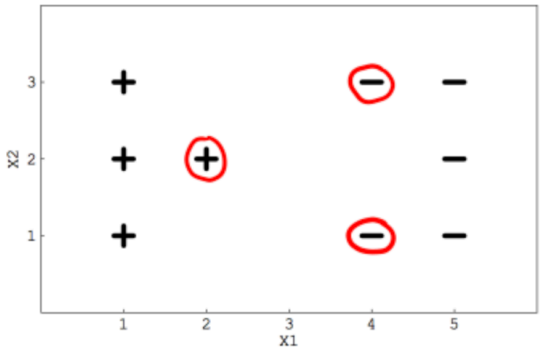
B. 错误

13. 简述 LR 和 SVM 的联系与区别？

14. 简述 L1 和 L2 的区别。

15. 现有一个点能被正确分类且远离决策边界。如果将该点加入到训练集，为什么 SVM 的决策边界不受影响，而已经学好的 logistic 回归会受影响？

16. 假设有一个线性 SVM 分类器用来处理二分类问题，下图显示给定的数据集，其中被红色圈出来的代表支持向量。



1) 若移动其中任意一个红色圈出的点，决策边界是否会变化？（ ）
   A. 会
   B. 不会
2) 若移动其中任意一个没有被圈出的点，决策边界会发生变化？（ ）
   A. 会
   B. 不会

17. SVM 中的泛化误差代表什么？（ ）

A. 分类超平面与支持向量的距离
B. SVM 对新数据的预测准确度
C. SVM 中的误差阈值

18. 若参数 C（cost parameter）被设为无穷，下面哪种说法是正确的？（ ）

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. 二者都不对

19. 怎样理解“硬间隔”？（ ）

A. SVM 只允许极小误差
B. SVM 允许分类时出现一定范围的误差
C. 二者都不对

20. SVM 算法的最小时间复杂度是 O（n^2），基于此，以下哪种规格的数据集并不适该算法？（ ）

A. 大数据集
B. 小数据集
C. 中等数据集
D. 不受数据集大小影响

21. SVM 算法的性能取决于：

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

22. 支持向量是最靠近决策表面的数据点

A. 正确
B. 错误

23. 以下哪种情况会导致 SVM 算法性能下降？

A. 数据线性可分

B. 数据干净、格式整齐
C. 数据有噪声，有重复值

24. 假设你选取了高 Gamma 值的径向基核（RBF），这表示：

A. 建模时，模型会考虑到离超平面更远的点
B. 建模时，模型只考虑离超平面近的点
C. 模型不会被数据点与超平面的距离影响

25. SVM 中的代价参数 C 表示什么？

A. 交叉验证的次数
B. 用到的核函数
C. 在分类准确性和模型复杂度之间的权衡
D. 以上都不对

26. 假定有一个数据集 S，但该数据集有很多误差（这意味着不能太过依赖任何特定的数据点）。若要建立一个 SVM 模型，它的核函数是二次多项式核，同时，该函数使用变量 C（cost parameter）作为一个参数。

1) 若 C 趋于无穷，以下哪种说法正确？
   A. 数据仍可正确分类
   B. 数据无法正确分类
   C. 不确定
   D. 以上都不对
2) 若 C 的值很小，以下哪种说法正确？
   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

27. 若训练时使用了数据集的全部特征，模型在训练集上的准确率为 100%，验证集上准确率为 70%。出现的问题是？

A. 欠拟合
B. 过拟合
C. 模型很完美

28. 下面哪个是 SVM 在实际生活中的应用？

A. 文本分类
B. 图片分类
C. 新闻聚类
D. 以上都对

29. 假定你现在训练了一个线性 SVM 并推断出这个模型出现了欠拟合现象。

1) 在下一次训练时，应该采取下列什么措施？
   A. 增加数据点
   B. 减少数据点
   C. 增加特征
   D. 减少特征
2) 假定你上一道题回答正确，那么根本上发生的是：
   1 偏差（bias）降低
   2 方差（variance）降低

3 偏差增加

4 方差增加

A. 1 和 2

B. 2 和 3

C. 1 和 4

D. 2 和 4

3) 还是上面的问题，如果不在特征上做文章，而是改变一个模型的参数，使得模型效果改善，以下哪种方法是正确的？

A. 增加代价参数 C

B. 减小代价参数 C

C. 改变 C 的值没有作用

D. 以上都不对

30. 在应用高斯核 SVM 之前，通常都会对数据做正态化（normalization），下面对特征正态化的说法哪个是正确的？

1 对特征做正态化处理后，新的特征将主导输出结果

2 正态化不适用于类别特征

3 对于高斯核 SVM，正态化总是有用

A. 1

B. 1 和 2

C. 1 和 3

D. 2 和 3

31. 假定现在有一个四分类问题，你要用 One-vs-all 策略训练一个 SVM 的模型。请看下面的问题：

1) 由题设可知，你需要训练几个 SVM 模型？

A. 1

B. 2

C. 3

D. 4

2) 假定数据集中每一类的分布相同，且训练一次 SVM 模型需要 10 秒，若完成上面的任务，共花费多少秒？

A. 20

B. 40

C. 60

D. 80

3) 现在问题变了，如果目前只需要将数据集分为 2 类，需要训练多少次？

A. 1

B. 2

C. 3

D. 4

32. 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%。

1) 假定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. 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

2) 在增加了模型复杂度之后，你发现训练准确率仍是 100%，原因可能是？、

1. 数据是固定的，但我们在不断拟合更多的多项式或参数，这会导致算法开始记忆数据中的所有内容

2. 由于数据是固定的，SVM 不需要在很大的假设空间中搜索

A. 1

B. 2

C. 1 和 2

D. 二者都不对

33. 下面关于 SVM 中核函数的说法正确的是？

1 核函数将低维空间中的数据映射到高维空间

2 它是一个相似度函数

A. 1

B. 2

C. 1 和 2

D. 以上都不对

34. SVM 的原理是什么？

35. SVM 为什么采用间隔最大化？

36. 为什么要将求解 SVM 的原始问题转换为其对偶问题？

37. 为什么 SVM 对缺失数据敏感？

38. svm RBF 核函数的具体公式？

39. 为什么 SVM 要引入核函数？