1. 在基于 SGD 随机梯度下降算法的神经网络中，每次打乱数据是非常重要和必不可

少的。(TRUE)

2. 决策树的适用面较广，对于分类应用和回归应用，决策树都可以被用来构建模型。(TRUE)

3. 由于贝努利贝叶斯比适合于贝努利（二项分布）分布，因此，贝努利贝叶斯只能用

于二分类任务。(FALSE)

4. 通常，我们认为对于一个系统来说，误差越小越好，因此无论是泛化误差还是经验 误差，都是越小越好。(FALSE)

5. 常见的性能度量标准有很多，例如：均方误差、准确率、错误率、精度、查全率、

查准率等，其中均方误差只能用于回归模型。(FALSE)

6. Anaconda 是 Python 语言最著名的第三方库之一，它可以实现基于矩阵的数据处理、

科学运算、可视化以及机器学习等功能。 (FALSE)

7. K-Fold交叉验证法比较适合大规模数据集。(FALSE)

9. 线性模型，不仅可以用来模拟线性关系的数据集，同时也可以用来模拟非线性关系

的数据集，甚至是高度非线性关系的数据集。(TRUE)

10. 在线性回归模型中， 参数 w 表示的是特征的权重，它可以用来衡量某个特征的重

要性。 (TRUE)

11. 随机森林算法利用训练数据构建了一系列的决策树，它根据损失函数最大化原则建

立决策树模型。 (FALSE)

12. 支持向量机只能用于二分类，无法应用到多分类中。 (FALSE)

13. 套索回归（Lasso Regression），是目前机器学习中性能最好的模型。(FALSE)

14. batch\_size 用于设置单次加入进行训练的样本数，它适用于 SGD 随机梯度下降算

法，一般设置为内存或显存所能支持的最大容量。(TRUE)

15. 与决策树不同，在构建随机森林的时候，树的深度参数 max\_depth 只能等于 1. (FALSE)

16. SVM 的核心因素也是它的根本原理是最大化类间间隙。 (TRUE)

17. 由于随机森林是由 1 个或多个决策树构成，因此随机森林的性能总是优于决策树。(FALSE)

20. 大规模数据集不存在样本不平衡的问题。 (FALSE)

21. 聚类是有监督的过程，而分类是无监督的过程。 (FALSE)

22. 激活函数对于神经网络来说是必须的。 (TRUE)

23. 一般来说，岭回归和套索回归的性能要优于最基本的线性回归模型。 (FALSE)

24. 在 SVM 数据集中，二维平面的分类线，三维空间的分界面都可以称为超平面。(TRUE)

25. 数据的最大方差给出了数据最重要的信息。(TRUE)

26. 一般用来表示预测结果的正确值。(FALSE)

27. 基于 KNN 的回归算法所生成的模型不需要进行训练。(TRUE)

28. 与决策树不同，在构建随机森林的时候，树的深度参数 max\_depth 只能等于 1. (FALSE)

29. 基于高斯模型的径向基核和非线性的多项式核的支持向量机一定比基于线性核的支

持向量机更好。(FALSE)

30. 由于贝努利贝叶斯比适合于贝努利（二项分布）分布，因此，贝努利贝叶斯只能用

于二分类任务。 (FALSE)

31. 随机森林算法利用训练数据构建了一系列的决策树，它根据损失函数最大化原则建

立决策树模型。 (FALSE)

32. 决策树的适用面较广，对于分类应用和回归应用，决策树都可以被用来构建模型。(TRUE)

33. 在主成分分析中，要求每个主成分之间是不相交的，即正交的。(TRUE)

34. 分层采样算法是指一种非常特殊的数据划分算法，它可以有效解决原始数据集类别

不平衡的问题。(FALSE)

35. PCA 主成分分析不仅能够降低数据的复杂性，实现降维，最重要的是在这个过程中

没有任何信息的损失。(FALSE)

36. 演绎是从特殊到一般的"泛化"过程，即从具体的事实归结出一般性规律。(FALSE)

37. 训练数据较少时，更容易发生欠拟合。(FALSE)

38. 线性回归模型中，对数几率函数ln(y/(1-y))描述了样本作为反例的相对可能性。 (FALSE)

39. 现实分类学习任务中，经常会遇到类别不平衡,会对学习过程造成困扰。(TRUE)

40. 决策树学习的目的是为了产生一棵泛化能力强，即处理未见示例能力强的决策树 。(TRUE)

41. 支持向量机模型中，划分超平面f (x) = ωTx+b中的ω为法向量，决定了超平面的方向。(TRUE)

42. 支持向量机模型中间隔，即为两个最近的同类样本之间的距离。(FALSE)

43. 贝叶斯模型中，聚类的形状由均值向量决定，聚类中心由协方差矩阵决定。 (FALSE)

44. 若个体学习器间存在强依赖关系、则必须并行生成的序列化方法。(FALSE)

45. K均值算法，是常见的聚类算法之一。(TRUE)

46. 机器学习中的过拟合问题，可以通过增加训练样本的方式彻底解决。(FALSE)

47. 现实分类学习任务中，经常会遇到类别不平衡,会对学习过程造成困扰。(TRUE)

48. 决策树模型中，随着划分过程不断进行，我们希望结点的“纯度”越来越小。 (FALSE)

49. 神经网络的输出是通过输出模块的处理得到的？(TRUE)

50. 聚类算法属于有监督学习算法。 (FALSE)

51. 如果两个变量相关，那么它们一定是线性关系. (FALSE)

52. 两个变量相关，它们的相关系数 r 可能为 0。(TRUE)