11、下列属于无监督学习的是（）

A、k-means

B、SVM

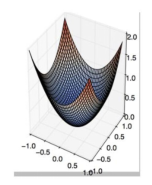
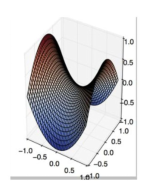
C、最大熵

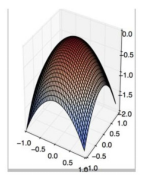
D、CRF

12、“过拟合”只在监督学习中出现，在非监督学习中，没有“过拟合”，这是（）

A、对的 B、错的

13、在下面哪种情况下，一阶梯度下降不一定正确工作（可能会卡住）？ （）

A、 B、

C、 D、以上都不正确

14、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

15、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

16、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：（）

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

E、回归问

F、结构分析问题

17、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（）

A、各类别的先验概率P(C)是相等的

B、以0为均值，sqr(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

18、对于k折交叉验证，以下对k的说法正确的是（）

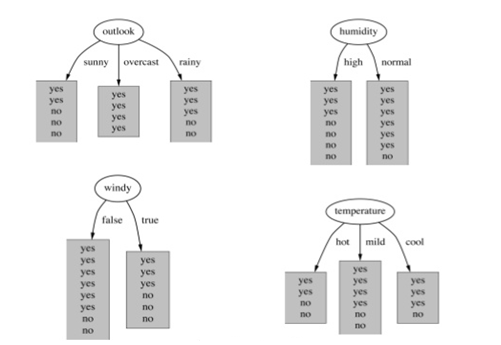
A、k越大, 不一定越好, 选择大的k会加大评估时间

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

19、训练决策树模型，属性节点的分裂，具有最大信息增益的图是哪一个（）



1. Outlook
2. Humidity
3. Windy
4. Temperature

20、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（）：

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