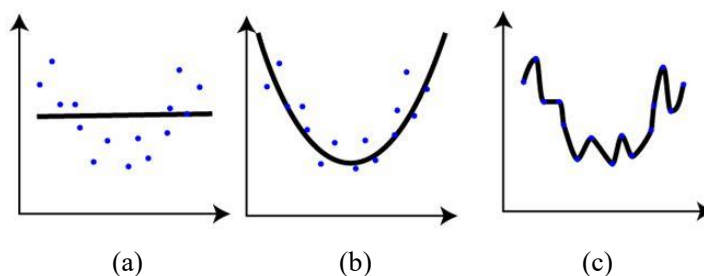


一、选择题（本题 30 分，每小题 2 分）

1. 如果一个计算机程序在某类任务  $T$  上以  $P$  衡量的性能随着经验  $E$  而不断提高，则称该程序在经验  $E$  中进行学习。假定根据前 10 年长沙的温度-耗电量数据作为一个学习算法的输入，要求其能根据该城市的天气预报情况估计当天的耗电量，下列情况属于经验  $E$  的是（ ）  
A、前 10 年的温度-耗电量数据                      B、估计值与实际值的偏离程度  
C、耗电量估计    D、以上全都不是
2. 如果采用基于多项式核的 SVM 分类器出现了过拟合，有如下改进方案：1) 减少惩罚因子  $C$ ；2) 增加惩罚因子  $C$ ；3) 减少多项式核参数  $d$ ；4) 增加多项式核参数  $d$ 。可采用的是（ ）  
A、1) 和 3)    B、1) 和 4)  
C、2) 和 3)    D、2) 和 4)
3. 关于 Batch 梯度下降法、Mini-batch 梯度下降法和 Stochastic 梯度下降法，下列说法中正确的是（ ）  
A、Batch 梯度下降法使用部分训练样本计算梯度来更新参数  
B、Mini-batch 梯度下降法使用单个训练样本计算梯度来更新参数  
C、Stochastic 梯度下降法使用所有训练样本来计算梯度来更新参数  
D、Mini-batch 梯度下降法适用于训练数据集的数量非常大的情况
4. 下面关于损失函数的描述中，错误的是（ ）  
A、平方损失函数公式是  $J(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$   
B、Logistic 回归的损失函数公式是  $J(\theta) = \frac{1}{m} \sum_{i=1}^m (-y^{(i)} \log(\hat{y}^{(i)}) - (1 - y^{(i)}) \log(1 - \hat{y}^{(i)}))$   
C、Softmax 的损失函数公式是  $J(\theta) = \frac{1}{m} \sum_{i=1}^m (-\log(\hat{y}_{y^{(i)}}^{(i)}))$ ，其中  $y^{(i)}$  表示第  $i$  个样本所属类别的序号  
D、使用平方损失函数的梯度下降法有可能得到局部最优值。
5. 下列关于特征归一化的描述中，正确的是（ ）  
A、范围归一化把  $x$  限制在  $[0, 1]$                       B、零均值归一化需要计算方差  
C、零均值单位方差归一化只需要计算  $x$  的最大最小值  
D、零均值单位方差归一化把  $x$  限制在  $[-0.5, 0.5]$

6. 下图分别给出了三个训练好的回归模型，关于这些模型描述下面说法正确的是( )



- A、图(a)中的模型方差高                      B、图(b)中的模型效果最好  
C、图(c)中的模型偏差高                      D、图(b)出现过拟合
7. 下列关于正则化的描述中，正确的是( )
- A、正则化可以缓解模型过拟合问题                      B、正则化参数  $\lambda$  越大越好  
C、正则化参数  $\lambda$  越小越好                      D、决策树算法不能使用正则化
8. SMO 算法中，拉格朗日乘子  $\alpha_1^{\text{old}}$  和  $\alpha_2^{\text{old}}$  是初始可行解，最优解为  $\alpha_1^{\text{new}}$  和  $\alpha_2^{\text{new}}$ ，但是  $\alpha_2^{\text{new}}$  需满足一定的不等式，其取值范围为  $L \leq \alpha_2^{\text{new}} \leq H$ 。关于 L 和 H 的取值下列哪个选项是正确的( )
- A、 $L = \min(C, C + \alpha_2^{\text{old}} - \alpha_1^{\text{old}})$ ,  $H = \max(0, \alpha_2^{\text{old}} - \alpha_1^{\text{old}})$   $y_1 = y_2$   
B、 $L = \max(0, \alpha_2^{\text{old}} + \alpha_1^{\text{old}} - C)$ ,  $H = \min(C, \alpha_2^{\text{old}} + \alpha_1^{\text{old}})$   $y_1 \neq y_2$   
C、 $H = \max(0, \alpha_2^{\text{old}} + \alpha_1^{\text{old}} - C)$ ,  $L = \min(C, \alpha_2^{\text{old}} + \alpha_1^{\text{old}})$   $y_1 = y_2$   
D、 $H = \min(C, C + \alpha_2^{\text{old}} - \alpha_1^{\text{old}})$ ,  $L = \max(0, \alpha_2^{\text{old}} - \alpha_1^{\text{old}})$   $y_1 \neq y_2$
9. 下列关于决策树的说法，正确的是( )
- A、ID3 算法趋向于选择取值个数多的属性作为测试属性  
B、C4.5 算法使用基尼系数计算信息增益率  
C、CART 分类树不需要对连续属性离散化  
D、CART 回归树需要对连续属性离散化
10. 下列不是随机森林特点的是( )
- A、可高度并行化  
B、随机选择一个已训练好的弱分类器进行预测  
C、克服完全生长树的缺点  
D、继承 CART 的优点
11. 下列属于原型聚类的是( )
- (1) K-Means (2) GMM (3) DBSCAN (4) MeanShift
- A、1、3                      B、2、4  
C、1、2                      D、3、4
12. 下列说法中，正确的是( )
- A、Adaboost 算法不会出现过拟合  
B、GMM 算法容易陷入局部极值，但 k-means 算法不会  
C、随机森林算法中增加决策树的数量容易导致过拟合  
D、GBDT 算法中增加决策树的数量容易导致过拟合

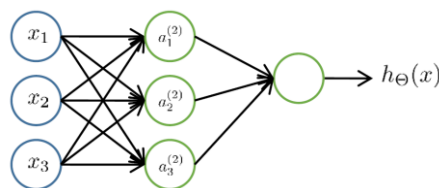
13. 下列哪一项在神经网络中引入了非线性( )
- A、随机梯度下降                      B、修正线性单元 (ReLU)
- C、卷积函数                              D、以上都不正确
14. 降维方法分为线性和非线性，其中非线性降维方法又可以分为基于核函数和基于特征值。下列哪种降维方法不属于基于核函数的降维方法( )
- A、KPCA                                      B、LLE
- C、KDA                                        D、MDS
15. 对某个模型进行评估，经统计得到下表信息，在下列选项中，错误的是( )

		实际结果	
		1	0
预测结果	1	80	10
	0	5	5

- A、precision=0.8                              B、recall=16/17
- C、accuracy=0.85                              D、F1-Score=32/35

## 二、分析与计算题（本题 70 分）

1. 拟用  $k$ -means 聚类算法将  $n$  个  $d$  维度数据划分成  $k$  簇。（本题 10 分）
- (1) 试给出  $k=1$  和  $k=n$  这两种情况下的聚类结果和对应的参数（提示： $k$ -means 算法的目标函数）；
- (2) 有多位同学采用相同的数据集和参数设置  $k$  来训练  $k$ -means，是否会得到不同的解？如有不同，如何选择？
2. 采用 Bagging 的方法来在训练决策树，假设采用自助采样(Bootstrap Sampling)的方法得到  $N$  组训练样本集(bootstrapped training sets):（本题 10 分）
- (1) 若  $N=1$ ，试说明训练得到的决策树与不采用自助采样训练得到的决策树是否一致？
- (2) 增加  $N$  的值，试说明训练得到的决策树的偏差(Bias)和方差(Variance)如何变化？
3. 假设有如下结构的神经网络结构：



另外，有如下数据集：

	$x_1$ (特征 1)	$x_2$ (特征 2)	$x_3$ (特征 3)	回归标签
$x^{(1)}$ (样本 1)	1	2	1	0
$x^{(2)}$ (样本 2)	3	1	2	1

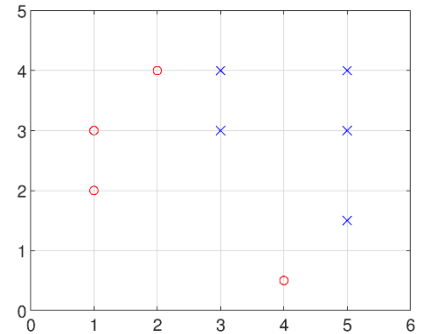
初始化时偏置量  $b$  初始化为全 0，权重  $W^{(1)}$  初始化为： $\begin{bmatrix} 0.1 & 0.2 & 0.3 \\ 0.4 & 0.5 & 0.6 \\ 0.3 & 0.2 & 0.4 \end{bmatrix}$ ，权重  $W^{(2)}$  初始化为：

$[0.2 \quad 0.4 \quad 0.6]$ ，使用线性激活函数，学习率设为 1，损失函数为  $J(\theta) = \frac{1}{2} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$ ，其中  $y^{(i)}$  为第  $i$  个样本的标签值，试回答下列问题：（本题 10 分）

- (1) 分别计算  $x^{(1)}$ 、 $x^{(2)}$  前向传播后的输出值。
- (2) 计算第一次反向传播后参数  $W^{(1)}$  更新后的值。

4. 给定训练集如右图所示，'x'和'o'分别表示正样本和负样本，采用 Adaboost 算法来学习分类器，弱分类器采用 decision stump（本题 15 分）

- (1) 画出 Adaboost 选择出的第一个 decision stump，用实线表示，并在决策边界画出对应类别；
- (2) 在图中圈出第一轮样本权重更新后权重最大的样本，并计算出样本权重更新后分类器的错误率；
- (3) 给出 Adaboost 选择出的第二个 decision stump，用虚线表示，并在决策边界画出对应类别；
- (4) 这两个弱分类器组合形成的分类器是否可以  
对所有数据进行正确分类？为什么？



5. 给定如下训练集（本题 10 分）

- (1) 求训练集的信息熵。
- (2) 求用特征 1 对样本集进行划分获得的信息增益。
- (3) 求用特征 2 对样本集进行划分获得的信息增益率。
- (4) 根据基尼系数，哪个特征可以作为最优划分属性。

提示： $\log_2(3/4)=-0.415$ ,  $\log_2(1/3)=-1.585$ ,  $\log_2(2/3)=-0.585$ ,  $\log_2(2/5)=-1.322$ ,  $\log_2(3/5)=-0.737$ 。

No.	类别	特征 1	特征 2
1	+1	1	0
2	+1	0	1
3	+1	2	0
4	-1	1	1
5	-1	0	0
6	-1	1	0
7	+1	0	0
8	+1	2	1
9	-1	1	1
10	-1	2	1

6. 假设训练集仅包含 4 个 2 维数据，分布如下图所示，其中  $x_1 = [0, 0]^T$ ,  $x_2 = [2, 2]^T$  为正样本， $x_3 = [h, 1]^T$ ,  $x_4 = [0, 3]^T$  为负样本， $h$  为参数 ( $0 \leq h \leq 3$ )（本题 15 分）

- (1) 若数据为线性可分，求出  $h$  的最大值？
- (2) 若数据为线性可分，则学习得到的最大间隔分类边界的方向是否会随着参数  $h$  的改变而改变？为什么？
- (3) 试写出最大间隔(The maximum margin)与参数  $h$  的关系。

