一、简答题

1、什么是机器学习？简述机器学习的一般过程。

答：机器学习是通过算法使得机器从大量历史数据中学习规律，从而对新样本做分类或预测。一般分为训练阶段、测试阶段和工作阶段。训练阶段的主要工作是根据训练数据建立模型，测试阶段的主要工作是利用验证集对模型评估与选择，工作阶段的主要工作是利用建立好的模型对新的数据进行预测与分类。

2、监督学习和非监督学习是什么

在监督学习中，算法通过训练数据来学习，这些训练数据包含了输入和与之对应的输出（也称为标签）。算法的目标是学习出一个模型，能够对新的未见过的数据做出准确的预测或分类。

无监督学习处理的是未标记的数据。这意味着训练数据不包含任何标签或者预定的输出结果。无监督学习的目标是探索数据本身的结构，发现数据中的模式、关系或者数据的分布特征。

3、什么是监督学习与无监督学习，监督学习算法中回归与分类的区别是什么？

监督学习和无监督学习都是种机器学习的方式，监督学习其训练数据集包括标签数据，这些标签数据告诉了机器学习算法输入数据与输出结果之间的对应关系。无监督学习其训练数据集不包含标签数据，算法必须从数据中自己发现模式和结构。

分类与回归都是监督型机器学习算法，连续变量的预测叫回归，离散变量的预测是分类。

4、什么是训练数据集和测试数据集？

在类似于机器学习的各个信息科学相关领域中，一组数据被用来发现潜在的预测关系，称为“训练数据集”。训练数据集是提供给学习者的案例，而试验数据集是用于测试由学习者提出的假设关系的准确度。

5、什么是线性回归？你能解释它的主要原理吗？

线性回归是一种统计方法，用于建立一个或多个自变量（预测变量）和一个因变量（响应变量）之间的关系模型。它假设变量之间存在线性关系，通过找到最佳拟合直线（在简单线性回归中）或超平面（在多元线性回归中），来预测响应变量的值。

其主要原理基于最小二乘法，目的是最小化实际观测值和模型预测值之间差异的平方和。通过这种方式，线性回归尝试画出一条直线（或超平面），尽可能接近所有观测点，从而能够用来预测新的数据点的响应值。

6、逻辑回归与线性回归有何不同？

逻辑回归与线性回归的主要区别在于它们处理的预测结果类型。线性回归用于预测连续的数值型结果，例如房价、温度或销售额。逻辑回归则用于预测分类结果，特别是用于二分类问题，如是/否决策、成功/失败结果等。

此外，逻辑回归和线性回归在模型构建时使用的函数也不同。线性回归使用的是线性函数，其预测结果可以是任何实数。而逻辑回归使用的是逻辑函数（或称 Sigmoid 函数），这使得它的输出被压缩在 0 和 1 之间，通常表示为发生某事件的概率。

7、解释 k-means 聚类算法的工作原理

k-means 聚类算法是一种广泛使用的无监督学习方法，用于将数据点分组成预定数量的聚类。算法的工作原理可以分为以下几个步骤：

随机选择 k 个数据点作为初始聚类中心。

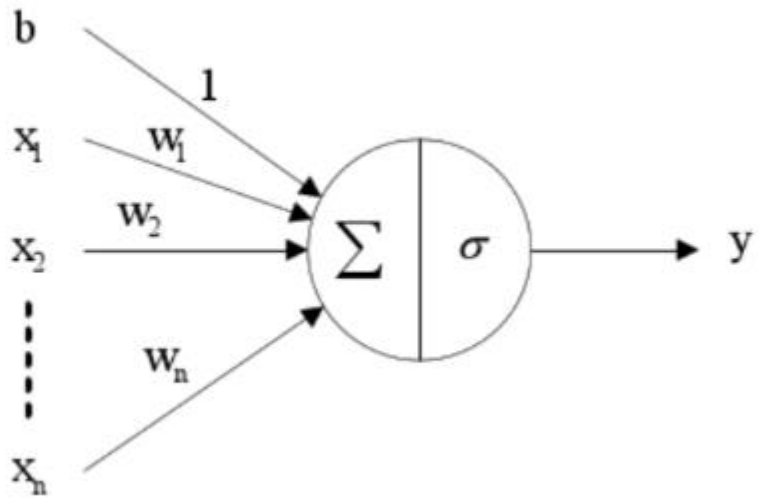
计算每个数据点到各个聚类中心的距离，并将每个点分配到最近的聚类中心所代表的聚类。

重新计算每个聚类的中心，通常是聚类中所有点的均值。

重复步骤 2 和 3，直到聚类中心的变化小于某个预设阈值，或者达到预定的迭代次数，此时算法结束。

8、什么感知机？感知机的基本结构是什么？单层感知机与多层感知机的区别是什么？

感知机是一种最简单的前馈神经网络，是一种二元线性分类器，其基本结构如下图所示。



单层感知机只有输入层和输出层中，多层感知机有输入层、隐藏层和输出层组成。

单层感知机只能解决线性可分问题，多层感知机还可以解决非线性可分问题。

二、分析计算题

1、基于一个学生在大学一年级的表现，预测他在大学二年级表现。令 x 等于学生在大学第一年得到的“A”的个数（包括 A-，A 和 A+成绩），以此表示学生在大学第一年得到的成绩。预测 y 的值：第二年获得的“A”级的数量。下表中每一行是一个训练数据。在线性回归中，我们的假设$h_\theta(x)=\theta_0+\theta_1x$，并且我们使用 m 来表示训练示例的数量。

x	y
3	2
1	2
0	1
4	3

（1） 代价函数的定义是$J(\theta_0,\theta_1)=\frac{1}{2m}\sum_{i=1}^{m}(h_\theta\left(x^i\right)-y^{(i)})^2$，求 J(0,1)。

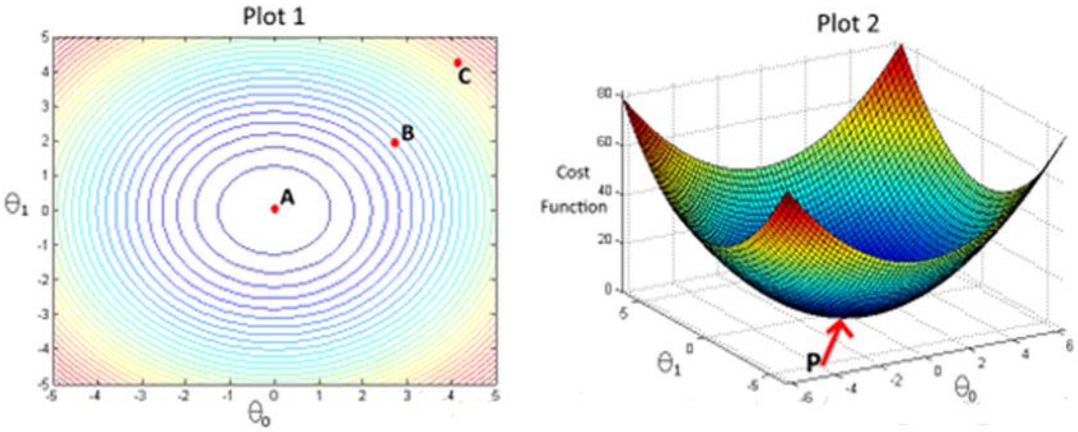
解析：$J(0,1)=\frac{1}{2*4}\left((3-2)^2+(1-2)^2+(0-1)^2+(4-3)^2\right)=0.5$

（2） 线性回归中，假设$\theta_0=-1,\theta_1=2$，求$h_\theta(6)$的值。

解析：$h_\theta(6)=-1+2*6=11$

（3） 代价函数$J(\theta_0,\theta_1)$与θ_0,θ_1的关系如下图 plot2 所示。下图 plot1 中给出了相同代价函数的等高线图。根据图示，选择正确的选项（选出所有正确项）。

Plots for Cost Function $J(\theta_0,\theta_1)$



A. 从 B 点开始，学习率合适的梯度下降算法会最终帮助我们到达或者接近 A 点，即代价函数$J(\theta_0,\theta_1)$在 A 点有最小值

B. 点 P（图 plot2 的全局最小值）对应于图 plot1 的点 C

C. 从 B 点开始，学习率合适的梯度下降算法会最终帮助我们到达或者接近 C 点，即代价函数$J(\theta_0,\theta_1)$在 C 点有最小值

D. 从 B 点开始，学习率合适的梯度下降算法会最终帮助我们到达或者接近 A 点，即代价函数$J(\theta_0,\theta_1)$在 A 点有最大值

E. 点 P（图 plot2 的全局最小值）对应于图 plot1 的点 A

解析：AE，P 是全局最小值，对应的是 A。

2、用两个硬币玩抛硬币的游戏，硬币 1 得到正面的概率为 θ，硬币 2 得到正面的概率为 2θ，你一共抛了五次，得到的结果是这样的（硬币 1，正面）（硬币 2，反面）（硬币 2，反面）（硬币 2，反面）（硬币 2，正面），用极大似然法求参数 θ。

参考答案：

用 x 表示正面向上事件，则根据题意，似然函数为：

$$P(\mathbf{x}|\theta,2\theta)=\theta*(1-2\theta)^3*2\theta=2\theta^2(1-2\theta)^3,$$

对数似然函数为：

$$logP(\mathrm{x}|\theta,2\theta)=2log\theta+3log(1-2\theta)+log2,$$

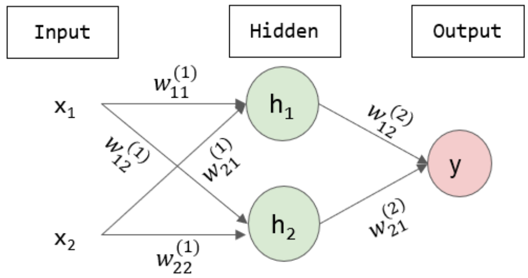
最大化对数似然函数，求导置为零，得到参数 θ 的估计值为：

$$|\theta^*=\frac{1}{5}$$

3、试设计一个前馈神经网络来解决 XOR 问题，要求该前馈神经网络具有两个隐藏神经元和一个输出神经元，并使用 ReLU 作为激活函数，要求写出各层的表达式。

参考答案：

（1）前馈神经网络如下图：



（2）权重矩阵：

$$W_1=\begin{pmatrix} w_{11}^{(1)} & w_{21}^{(1)} \\ w_{12}^{(1)} & w_{22}^{(1)} \end{pmatrix}=\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix},$$

$$W_2=\begin{pmatrix} w_{12}^{(2)} & w_{21}^{(2)} \end{pmatrix}=\begin{pmatrix} 1 & 1 \end{pmatrix}$$

（3）

第一层：

$$Relu\left(\begin{pmatrix} w_{11}^{(1)} & w_{21}^{(1)} \\ w_{12}^{(1)} & w_{22}^{(1)} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right)=Relu\left(\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}\right)$$

$$=Relu\begin{pmatrix} 1 & 0 & -1 & 0 \\ -1 & 0 & 1 & 0 \end{pmatrix}=\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

第二层：

$$\mathrm{Re}\,lu\left(\begin{pmatrix} w_{12}^{(2)} & w_{21}^{(2)} \end{pmatrix}\begin{pmatrix} h_1 \\ h_2 \end{pmatrix}\right)=\mathrm{Re}\,lu\left(\begin{pmatrix} 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}\right)=\mathrm{Re}\,lu\begin{pmatrix} 1 & 0 & 1 & 0 \end{pmatrix}$$

$$=\begin{pmatrix} 1 & 0 & 1 & 0 \end{pmatrix}$$