**作业1：请简述KNN算法的原理和应用，并找出一个与KNN算法思想相近的例子（诸如：身边发生的案例、谚语等）。**

原理：从所有训练样本中找出和未知样本最近的K个样本，将K个样本中出现最多的类别赋给该未知样本；

应用：文本分类，样本预测

谚语：1.物以类聚，人以群分2.近朱者赤，近墨者黑3.少数服从多数

**作业2：请简述决策树算法的原理和应用，并领悟出该算法所蕴含的哲理，试着讲出该哲理。**

**原理：**决策树是通过一系列规则对数据进行分类的过程。它提供一种在什么条件下会得到什么值的类似规则的方法。决策树分为分类树和回归树两种，分类树对离散变量做决策树，回归树对连续变量做决策树。

**应用：数据分类，数据挖掘**

**哲理：每一步的选择都会有很大的不同。**

**作业3：请简述朴素贝叶斯算法的原理和应用，你觉得该算法有何科学道理？**

**原理：**朴素贝叶斯（naive Bayes）是基于贝叶斯定理与特征（features）条件独立性假设的一种分类算法，属于生成式模型，对联合分布建模。对于给定的训练数据集,首先基于条件独立性假设，学习输入和输出的联合概率分布，然后基于此模型，对于给定的输入 ,利用贝叶斯定理求出后验概率最大的输出。

**应用：垃圾邮件分类，文本分类**

#### 哲理：接受改变个人认知贝叶斯定理通过计算相关事件的条件概率来更新其先验概率。同样, 我们都对人、文化和社会有先入为主的认知和评判。这些认知来源于我们的社会、经济和人口状况。当我们处于与我们旧认知相悖的新境遇, 有意识地提升和改变自己是至关重要的。永远不要畏惧改变。

**作业4：请简述支持向量机算法的原理和应用。只依据少数支持向量，你觉得该方法找到的分隔界面是否科学？**

**原理：**支持向量机（support vector machines, SVM）是一种二分类模型，它的基本模型是定义在特征空间上的**间隔最大的线性分类器**，间隔最大使它有别于感知机；SVM还包括**核技巧**，这使它成为实质上的非线性分类器。SVM的的学习策略就是间隔最大化，可形式化为一个求解凸二次规划的问题，也等价于正则化的合页损失函数的最小化问题。SVM的的学习算法就是求解凸二次规划的最优化算法。

**应用：模式识别，分类，回归分析**

答：科学，因为SVM是建立在统计学习理论中结构风险最小化原理基础上，根据有限的样本信息，在模型的复杂性(即对特定训练样本的学习精度)和学习能力(即无错误地识别样本的能力)之间寻求最佳折衷，以期获得最好的推广能力**。**SVM

**作业5：请简述k均值算法的原理和应用，并试着说出该算法给你带来了哪些思想方面的启迪？**

**原理：**在数据集中根据一定策略选择K个点作为每个簇的初始中心，然后观察剩余的数据，将数据划分到距离这K个点最近的簇中，也就是说将数据划分成K个簇完成一次划分，**但形成的新簇并不一定是最好的划分，**因此生成的新簇中，**重新计算**每个簇的中心点，然后在重新进行划分，直到每次划分的结果保持不变。在实际应用中往往经过很多次的迭代仍然达不到每次划分结果保持不变，甚至因为数据的关系，根本就达不到这个终止条件，实际应用中**往往采用变通的方法设置一个最大迭代次数，当达到最大迭代次数时，终止计算。**

**应用：预测，分类**

#### 启迪：****关系随着时间的推移而变化，****K-means算法通过对数据进行分配并反复迭代，从而达到聚类的目的。人际交往和情绪在建立我们的社交关系中起着至关重要的作用。当关系紧张时，我们会感到沮丧。我们应该意识到再强大的关系也可能面对改变或结束，即使你没有做错任何事，就像在K-means中一样。

**作业6：请简述梯度下降法的原理和应用，并说出该算法所蕴含的人生哲理。**

**原理：梯度下降算法包含多种不同的算法，有批量梯度算法，随机梯度算法，折中梯度算法等等。对于随机梯度下降算法而言，它通过不停的判断和选择当前目标下最优的路径，从而能够在最短路径下达到最优的结果。我们可以在一个人下山坡为例，想要更快的到达山低，最简单的办法就是在当前位置沿着最陡峭的方向下山，到另一个位置后接着上面的方式依旧寻找最陡峭的方向走，这样每走一步就停下来观察最下路线的方法就是随机梯度下降算法的本质。**

**应用：用于搜索优化**

**人生哲理：基于当前情况做出决定，**从回归方法到人工神经网络，梯度下降法是许多机器学习方法的重要组成部分，其始终沿着梯度下降的方向最快地到达预期的目的地。我们经常遇到需要我们在众多替代方案中选择一种方案并作出决定的情况。这些决定都对将来有潜在的影响。梯度下降为我们提供了解决问题的捷径，基于当前的情况提供最佳选择，而不过度考虑将来的过程。**不要等待最理想的方案，你永远无法预测未来。**