**《机器学习》课程设计**

1. **基本要求**

从下述三个案例中选择自己感兴趣的某个案例进行课程设计，完成且不局限于案例中的具体要求。

每个学生必须在充分分析问题和数据的基础上，采用回顾课堂授课、查阅文献和咨询老师等方法，通过建模编程的方式解决课程设计中遇到的各种问题，独立完成该项设计和论文撰写。

报告需要介绍完成设计所用模型的历史、思想、建模步骤，同时重点描述建模过程、所用的算法、实验结果以及结果的分析。整个报告以科技论文的基本格式呈现，附录提供整个设计所用代码。

1. **内容**
2. **人脸识别**
3. **问题**：仔细研究ORL数据库（Olivetti Research Laboratory in Combridge）的人脸图像ORL\_32x32（数据上传至QQ群），对图像的特征和复杂性进行认真分析，使用机器学习中人工神经网络模型或支持向量机，设计人脸识别系统，并对系统做出评估。

。

1. **主要要求**

* 二分类问题

1. 随机选择其中两类数据或者以是否戴眼镜作为类标签，对数据进行预处理，同时确定训练数据集、测试数据集和验证数据集；
2. 基于训练数据集和验证数据集确定网络层数和神经元个数，完成神经网络模型建立；或基于训练数据集和验证数据集确定支持向量机的核函数以及其余超参数。（附模型图）
3. 使用测试数据集对所建模型进行评估，（附混淆矩阵，测试精度表，查准率，查全率和F1表）；
4. 改变神经网络训练过程中学习率，隐层个数，神经元个数（或改变支持向量机中的核函数和其它超参数）等比较不同网络得到的测试精度，查准率，查全率和F1值。探索学习率、隐层个数和神经元个数对于网络性能的影响。（附不同模型得到的混淆矩阵和精度比较表、查准率比较表、查全率比较表以及F1值比较表）。

* 多分类问题

1. 对数据进行预处理，确定训练数据集、测试数据集和验证数据集；
2. 基于训练数据集和验证数据集确定网络层数和神经元个数，完成神经网络模型建立；或基于训练数据集和验证数据集确定支持向量机的核函数以及其余超参数。（附模型图）
3. 使用测试数据集对所建模型进行评估，（附测试精度表）；
4. 改变神经网络训练过程中学习率，隐层个数，神经元个数（或改变支持向量机中的核函数和其它超参数）等比较不同网络得到的测试精度。探索学习率、隐层个数和神经元个数对于网络性能的影响。（附不同模型得到的精度比较）。

我需要做的事情是先将数据可视化，然后观察图像中的特征，然后肯定是看一下各个类别分别表示什么。

完成

每一类数据太少，数据集的划分有点困难。每一类数据才10个，怎么去解决这个问题。

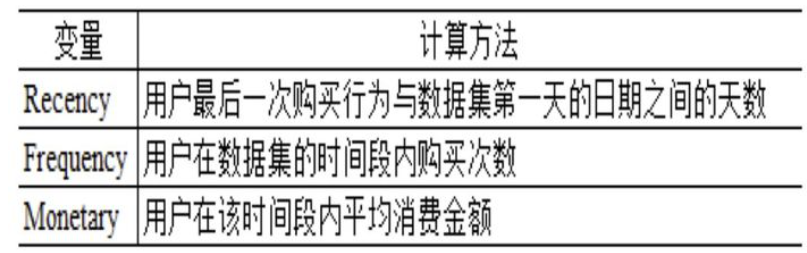
我的想法是使用BP网络。

1. **评估客户的RFM价值，帮助公司合理配置资源**
2. **问题：**不同的公司对客户关系有不同的管理方法。本设计我们根据UCI网站上的一个英国小型在线商城的用户交易数据 Online retail data set（数据上传至QQ群）进行学习，计算客户的RFM价值（R=Recency 代表该用户最近一次交易；F=Frequency 代表用户交易频率；M=Monetary 代表用户平均单次交易金额。），基于RFM值分析客户过去对公司的贡献（可将客户进行划分或分组），以便公司因地制宜，因人而异，合理配置资源。
3. **主要要求**
4. 整个数据集包含了541,909行数据。首先删除无用数据，如财务调整、存货损毁等或出于计算力的考虑删除部分数据，得到用于分析的数据集。

数据的特征说明：



1. 计算每个客户的RMF数据，具体计算方法如下



1. 对上述得到的RFM数据分析其量纲是否一致，若不一致，则进行数据标准化；
2. 基于聚类分析方法同时根据RFM数据对客户进行划分，并可视化聚类分析过程中划分结果以及最终划分结果（可视化图中可将不同的簇用不同的颜色标记，同时需要标记每个客户的ID）；（附可视化图）
3. 根据上述聚类结果，可视化簇内客户数量在总客户中的比例以及簇内贡献销售额对总销售额的占比，为公司提供信息。（附可视化图）
4. **预测无线信号的覆盖强度**
5. **问题：**5G在全球范围内的应用在不断扩大，运营商在部署5G网络的过程中，需要合理选择覆盖区域内的基站站址，进而通过部署基站来满足用户的通信需求。本课程设计要求学生站在设备供应商以及无线运营者的角度，合理运用机器学习模型准确预测无线信号覆盖的强度，从而大大减少网络建设成本，提高网络建设效率。
6. **主要要求：**
7. 整个数据集5Gxinxidataset包含了999，999个样本，每个样本有49个属性，数据同时包含了每个站点的信号接收强度。首先分析是否需要对数据进行预处理。；
8. 将数据集划分为训练集、验证集和测试集；
9. 使用训练数据训练线性回归模型，可视化训练过程中训练集和验证集数据上的均方误差的变化；同时使用训练好的模型对测试集进行测试，进而计算测试集上的误差；（附训练过程中训练集和验证集上的均方误差的变化图）；
10. 使用训练数据训练支持向量回归或者神经网络模型，可视化训练过程中训练集和验证集数据上的均方误差的变化；同时使用训练好的模型对测试集进行测试，进而计算测试集上的误差；（附训练过程中训练集和验证集上的均方误差的变化图）；
11. 将e）中的测试结果与f）中的测试结果进行比较分析。