# 机器学习第五章实验内容

## 一、实验内容

### 1.1使用 sklearn 的多层感知机

使用 sklearn 自带的手写数字数据集：

1. 学习标准化处理的方法
2. 使用 sklearn.neural\_network.MLPClassifier 完成手写数字分类任务
3. 绘制学习率为 3，1，0.1，0.01 训练集损失函数的变化曲线

### 1.2神经网络：线性回归

1. 学会梯度下降的基本思想
2. 学会使用梯度下降求解线性回归
3. 了解标准化处理的效果

### 1.3神经网络：对数几率回归

1. 完成对数几率回归
2. 使用梯度下降求解模型参数
3. 绘制模型损失值的变化曲线
4. 调整学习率和迭代轮数，观察损失值曲线的变化
5. 按照给定的学习率和迭代轮数，初始化新的参数，绘制新模型在训练集和测试集上损失值的变化曲线，完成表格内精度的填写

### 1.4神经网络：三层感知机

1. 实现一个三层感知机
2. 对手写数字数据集进行分类
3. 绘制损失值变化曲线
4. 完成 kaggle MNIST 手写数字分类任务，根据给定的超参数训练模型，完成表格的填写

## 二、数据介绍

### 2.1 kaggle房价预测数据集

文件名：

1. 原始数据：kaggle\_hourse\_price\_train.csv
2. 字段说明：kaggle房价预测字段说明.txt
3. 预处理脚本：preprocess.py
4. 处理后数据：kaggle房价预测处理后数据.csv

数据来源：<https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data>

数据预处理步骤：

该数据集是真实数据集，含有大量的缺失值和噪声数据，提供的预处理脚本只对含缺失值的特征进行丢弃处理，学生可根据自己的需要进行其他方式的处理。

### 白葡萄酒质量数据集

文件名：

1. 原始数据：winequality-white.csv
2. 字段说明：葡萄酒字段说明.txt

数据来源：<http://archive.ics.uci.edu/ml/datasets/Wine+Quality>

数据无需预处理，读取即可使用

import pandas as pd

data = pd.read\_csv('data/wine\_quality/winequality-white.csv', delimiter=";")

### spambase垃圾邮件数据集

文件名：

1. 原始数据：spambase.data
2. 字段说明：spambase数据说明.txt

数据来源：<http://archive.ics.uci.edu/ml/datasets/spambase>

数据无需预处理，读取即可使用

import numpy as np

data = np.loadtxt('data/spambase/spambase.data', delimiter = ",")

### Dota2 Games Results Data Set Dota2游戏结果数据集

文件名：

1. 原始数据：dota2Train.csv
2. 字段说明：dota2比赛结果字段说明.txt

数据来源：<http://archive.ics.uci.edu/ml/datasets/Dota2+Games+Results>

数据无需预处理，读取即可使用

import numpy as np

data = np.loadtxt('data/dota2Dataset/dota2Train.csv', delimiter=',')