

Task2

Review: 未完成上一次任务的尽快完成，考完试放松一下，设置期限元月16日晚，在群文件中将记录和源代码打包上传，希望这个task可以提前2-3天完成。

思考: 大家中学开始求解的几乎所有的问题都是精确解，最小二乘法就是求得线性模型参数的精确解(解析解)，而梯度下降法求解的是数值解。数值解肯定是不够精确的，那我们为什么还要求数值解呢？

第二个机器学习模型——Logistics Regression

分类问题是机器学习的又一大问题，给出身高体重你是否能判断一个人的性别？我们可以将这个任务交给机器实现。

1 前置知识

极大似然估计: 数理统计的知识，对大一的同学有一定的难度，参考资料：[极大似然估计详解](#)。

这个有问题的可以直接在群里问。难度其实也不能算大，只是抽象，今后的机器学习模型的数学难度比这个还要大。这里的关键不是背方法，而是理解极大似然估计方法的motivation。

时间安排: 1.16-1.18

2 模型学习

理论学习: [\[中英字幕\]吴恩达机器学习系列课程P26-P42](#)

时间安排: 1.19-1.21

3 编码实现

同样建议使用面向对象程序设计的思想进行程序设计。示例：

```
import numpy as np

class LogisticModel(object):
    """
    描述模型功能
    """
    def __init__(self):
        """
        构造函数
        """
        pass

    def logisticRegression(self, data):
        """
        模型主体，尽量将这个函数拆成多个原子性的函数以方便修改模型
        """
        pass

    def showOutput(self, data, label):
        """
        展示数据
        """
```

```

        pass
    def otherFunction(self):
        pass

class DataReader(object):
    """
    获取数据
    """
    def __init__(self):
        """
        构造函数
        """
        pass

    def returnData(self, filepath = 'xxxx'):
        """
        返回数据
        """
        pass

if __name__ == "__main__":
    data_reader = DataReader()
    data = data_reader.returnData()
    logistic_model = LogisticModel()
    logistic_model.logisticRegression(data)

```

初级：实现二维平面的分类任务并做可视化（完成这个就已经相当OK了）。

中级：将维度进行拓展，代码模块的入口能识别输入的数据的维度，并根据这个维度进行计算。

高级：实现MNIST数据集中0、1两个手写数字的分类（思路：将二维的照片拉伸成一个向量）。

时间安排：1.22-1.24