# 机器学习编程总结与研究

这里的主要的目的就是对各个部分的知识进行总结，然后明确各个插件的作用，为以后的调用与参考做好准备。因为每一个部分都很实用。（这里的要几个feature，正则化，等等内容必须在后面引入学习曲线后才能够准确判断）

1. 线性回归：

线性回归我们必须要明确线性回归并非是直线，一维时可以是其的N次方，n维也是这样，这样一维的时候是曲线的线，也就是线性表达，（这里只用改变我们可以给X带入任何一个函数去作为新的feature，而结果曲线可能是特别复杂的，而对于参数依然是同样的方法。）而参数的拟合依据于最优化costfunction来达到最优的预期（这里未来要加入正则化和学习曲线提高准确度）。

线性回归的具体的做法和函数位置：

**机器学习总结1**：1.数据的载入，2.画图的简单方法3.一维线性回归的J的求，与theta的求。4画图原来的图和预测的图，在一起5.在一定范围内画图J对应theta0和1的3D图，6.画出轮廓线，也就是2D的上图。

**wd.m**：这个是用来修改二维的变量画出3d的图的，最后可以看到可以画出各种各样的图。

**文件夹一维线性拟合中的：ex1**可以直接修改里面的参数，调入数据，进行一维的线性拟合，得到结果，然后可以看到有关的预期值与原数据的图和（J,theta0,theta1）的3D和轮廓图，

当中的costfunction和gd是可以直接复制去调用的，作为一维的两个计算。

**文件夹多维线性拟合中的ex1：首先用了特征正常化，**再调用两个函数去计算出我们要的theta，最后画出迭代次数和J的关系图，用来调整我们想要调整的alpha和次数，并且最后对于一组的特征值下的结果进行了预测。这里因为是多维，所以没有办法去画出相关的图，超出了三维了。

总结 ：一维和多位的线性回归的目的就是通过已知的特征量，对于这个特征量下的Y进行预测，得到我们想要的结果。这里是通过大量的训练从而得到最后的参数，得到拟合结果。

1. 逻辑回归：

逻辑回归和神经网络的关系是什么样的呢，多元逻辑回顾也可以实现对于样本的分类，不过他的theta本质上依然是对于特征的函数，最终，去判断该特征下，属于哪一个class，这样我们就可以依据一些过去的特征去做预测，也就是可以做到的是根据一些特征去判断他的一些分类，属于哪一些，那么在股市里，我们的想法就是对于未来的分类，利用这种方法去预测未来的分类。

**机器学习总结2：主要包含：**

**一、（1）画出散点图，画出圈与十字，加上图例**

**（2）计算出costfunction，这里的没有正则化**

**（3）fminunc函数用来求出最优解（theta与cost）的使用方法（内含plotboundary的使用）**

**（4）预测一组数据的为1概率，并且预测准确率**

**二、（1）如何创造出多变量，6次28项式**

**（2）正则化**

**（3）求出J，以及用fminunc函数**

**（4）预测概率**

**三、多元逻辑回归：**

**（1）正则化有关函数，完成多元的目标最优求解**

**（2）预测样本的准确率**

1. 神经网络

神经网络的主要是构建类似于人类大脑的复杂处理体系，去训练处一个能够处理复杂问题的网络体系，而一般的逻辑回归和线性回归都是线性的，很难去处理复杂额问题，而神经网络的各个层之间是类似于多元逻辑回归的结构，通过这样的结构的结合，最终可以实现接近完美的拟合，这样处理能力也就极其强大了。因此相比之下的逻辑回归，准确性和各方面都不如神经网络更加优异。

因为是一步步的操作，因此目前没有办法实现后面高偏差方差等问题的调整 ，目前对于手头的问题，我们可以将神经网络总结为，手里有一定的特征，而这些的特征可以十分的巨大，（n可能几百几千几万）而有了这些特征，从而分析出已知的分类，比如说，这个图片里的东西是不是人，未来的一周的股票走势的分类，如按照涨幅分成10类，对数据标记后实现完整的一个推测，（可以以过去一段时间的价格变化和换手率为特征）通过这样的方式，让她自己学习当中可能的规律，最终实现我们想要的目标。也就是由特征A，到结论B。可以由很多的拓展，重要的是思路。

神经网络的具体的操作步骤，也会在下面体现：

**机器学习总结3**：

1. 正向传播和正向传播的结果计算（调用预测函数）
2. 反向传播去训练参数，并预测
3. 计算反向传播的误差函数
4. 调用函数计算误差函数
5. 正则化的误差函数
6. 求出有关偏导
7. 初始化参数值
8. 进行偏导数的检查（巨复杂）
9. 正则化的修改参数
10. 训练结果，训练与调用预测函数预测最后的准确率

下面就是要对有关的内容进行实践。

（1）经过了实践，把握了数据的整理，搜集，处理的步骤，另外对于程序的过程有了更加深刻的理解，创作了一个比较不错的小程序，实现了对于过去100交易日的数据处理后推断出一周后的涨幅，成功了达到了57.88%

（2）理解了神经网络在目前三层的时候用sigmoid函数，而没有办法拓展到4层，后续依然需要学习神经网络有关的深度学习以及卷积等知识，学习更多

（3）对于之后的随机的迭代以及学习曲线的应用也十分的关键。是对程序优化的关键。

（4）进行了大量的测试后，发现了在训练集的培训，理论上是可以达到百分之百的成功率的，但是由于迭代是一步步的，导致了一定程度上的速度太慢，未来看看随机的迭代有没有机会结果速度慢的问题，另外神经网络目前可以培训仅仅可以用于简单的使用，也就是对于复杂的问题，比如说非人能判断的问题，其也未必能够解决，因此需要通过实践，确定其训练的准确率，通过多组的实验，说不定可以找到特别的规律，而这个规律则十分的有价值，因此必然不容易找到。因此多去分析，去把握，说不定可以找到不错的。

(4)

1. 机器学习优化

机器学习的优化重点解决在实践中怎么优化机器学习的代码，怎么去处理，让我们的机器学习表现的更好。

首先在数据有大量的时候，很难会过度拟合，而且会出现实际的拟合程度和CV差不多，而对于高BIAS，拟合度不足：用减小LAMBDA，用更多的特征，增加层数或者层数单元，

相比之下，过度拟合可以提高lambda，减小特征，更多数据来处理。

对于股票的问题，数据没有问题，关键是如何才能拟合好，这是问题，对于相关的线性问题以及相关的简单均衡，后面遇到再看。