一轮考核答辩论文

袁颖茵

广东工业大学计算机学院计算机类3班

摘要：工作室考核

关键词：一轮考核

1. 解决线性回归
2. 线性回归的原理和含义

在统计学中，线性回归（Linear Regression）是利用称为线性回归方程的最小平方函数对一个或多个自变量和因变量之间关系进行建模的一种回归分析。这种函数是一个或多个称为回归系数的模型参数的线性组合。只有一个自变量的情况称为简单回归，大于一个自变量情况的叫做多元回归。（这反过来又应当由多个相关的因变量预测的多元线性回归区别，而不是一个单一的标量变量。）而在本次考核中，考核内容为多元线性回归。在线性回归中，数据使用线性预测函数来建模，并且未知的模型参数也是通过数据来估计。这些模型被叫做线性模型。最常用的线性回归建模是给定X值的y的条件均值是X的仿射函数。不太一般的情况，线性回归模型可以是一个中位数或一些其他的给定X的条件下y的条件分布的分位数作为X的线性函数表示。像所有形式的回归分析一样，线性回归也把焦点放在给定X值的y的条件概率分布，而不是X和y的联合概率分布（多元分析领域）。线性回归是回归分析中第一种经过严格研究并在实际应用中广泛使用的类型。这是因为线性依赖于其未知参数的模型比非线性依赖于其未知参数的模型更容易拟合，而且产生的估计的统计特性也更容易确定。线性回归模型经常用最小二乘逼近来拟合，但他们也可能用别的方法来拟合，比如用最小化“拟合缺陷”在一些其他规范里（比如最小绝对误差回归），或者在桥回归中最小化最小二乘损失函数的惩罚.相反,最小二乘逼近可以用来拟合那些非线性的模型.因此，尽管“最小二乘法”和“线性模型”是紧密相连的，但他们是不能划等号的。在学习线性回归的含义时，主要查阅了csdn，以及看了相关视频，主要看了李沐老师的视频，这部分较好理解，因此能比较快速解决且理解较深刻，原理也挺容易听懂。（部分取自百度）

1. 梯度下降法

梯度下降法（英语：Gradient descent）是一个一阶最优化，通常也称为最速下降法。 要使用梯度下降法找到一个函数的局部最极值，必须向函数上当前点对应梯度（或者是近似梯度）的*反方向*的规定步长距离点进行迭代搜索。如果相反地向梯度*正方向*迭代进行搜索，则会接近函数的局部极大值点；（取自百度）这个过程则被称为梯度上升法，该方法难度较大，需有较强的数学思维，这部分学习相对吃力，但是最后也解决了。

1. 代码实现

这部分对我来说是最吃力的，首先我的基础不够牢固，刚接触python，函数较多，基本遇到一个函数就要查阅相关资料，题目不同，编译环境不同，代码实现基本要靠自己摸索，我也查阅了各种资料，但大部分都是直接套用线性回归模型，没有找到很好的案例。最后比较遗憾的是还是没有完整实现一次线性回归。

1. Softmax回归

1. Softmax回归简介

Softmax回归其实就相当于多类别情况下的逻辑回归，对比如下：

逻辑回归的假设函数（hypothesis）：

整个逻辑回归模型的参数就是theta，h(\*)是sigmoid函数，输出在0～1之间，一般作为二分类算法。对于具体的问题，找出最合适的theta便是最重要的步骤，这是最优化问题，一般通过定义代价函数，然后最小化代价函数来求解，逻辑回归的代价函数为：图示

描述已自动生成

最小化J(theta)，一般采用梯度下降算法，迭代计算梯度并更新theta。

Softmax的假设函数：

文本

中度可信度描述已自动生成

逻辑回归里将-theta\*x作为sigmoid函数的输入，得到的是0或者1，两个类别。而softmax有有k个类别，并且将-theta\*x作为指数的系数，所以就有e^(-theta\_1\*x)至e^( -theta\_k\*x)共k项，然后除以它们的累加和，这样做就实现了归一化，使得输出的k个数的和为1，而每一个数就代表那个类别出现的概率。因此：softmax的假设函数输出的是一个k维列向量，每一个维度的数就代表那个类别出现的概率。

Softmax的代价函数：文本

描述已自动生成

本质上跟逻辑回归是一样的，采用NLL，如果加上权重衰减项（正则化项），则为：文本

描述已自动生成

最小化代价函数，同样可以采用简单而有效的梯度下降，需要提到的是，在程序实现中，我们一般采用批量随机梯度下降，即MSGD，minibatch Stochastic Gradient Descent，简单来说，就是每遍历完一个batch的样本才计算梯度和更新参数，一个batch一般有几十到几百的单个样本。PS：随机梯度下降则是一个样本更新一次。（取自CSDN博主「wepon\_」的原创文章。原文链接：<https://blog.csdn.net/u012162613/article/details/43157801>）

2.代码实现

Softmax回归实际上是分类问题，相对于线性回归来说难度更大，因此我在写这个部分的代码时感到非常吃力，只能跟着网上的博主写。

以上是一轮考核的成果，我受益匪浅，接触了一个新的方向---机器学习，粗略学习了一门新的语言，有了很多很不错的收获，但是我对自己这段时间来的表现并不是特别满意，首先，基础知识学习地不够扎实，总在很简单的问题方面浪费太多时间，其次，独立思考的能力不够，会经常遇到问题，总是想着有人能帮我解决，一遇到不会的就去问师兄师姐同学们，这样的习惯非常不好，我应该培养自己的自学能力，最后，我觉得自己的心态不够好，遇到问题就很容易想放弃。针对以上问题，我一定会做出妥善的处理，争取在以后的学习中有所收获。