线性回归与逻辑回归

Notebook:	学习		
Created:	8/11/2019 2:35 PM	**Updated:**	8/11/2019 2:39 PM
Author:	yikun_wang@shannonai.com		
URL:	about:blank		

线性回归与逻辑回归

本文试图在复习线性回归和逻辑回归的过程中进行一个记录与总结。## 线性模型在进入线性回归之前，首先记录一下线性模型的基本形式。用一句话总结，线性模型是一个试图通过样本不同属性的线性组合进行预测的函数。这类模型往往形式简单、易于建模，同时解释性足够强。经典的线性模型包括线性回归、逻辑回归、线性判别分析（LDA）等。

线性回归

- 定义 给定数据集，线性回归试图从中学得一个针对样本各个属性的线性组合形成的线性模型，用于尽可能准确地预测实值输出标记。

- 属性的数学表达
  线性回归需要将样本的各个属性都以数学形式表达出来，如果属性本身就是连续属性，则可以直接将其映射到某个取值范围（如 [0,1]）；如果属性为离散值，则可以分为两种情况讨论。如果离散属性存在有序关系，如 “高” 和 “矮”、“青年”、“中年”、“老年”，都可以转化为连续属性{1.0, 0.0}、{1.0, 0.5, 0.0}等。而无序的属性，如颜色属性，最好用k维向量表示，如one-hot向量，其中k即是该属性的类别数。如果在无序属性中不恰当地引入序关系，可能会给后续的距离计算等造成误导。
- 误差度量
  线性回归的经典误差度量方法就是“最小二乘法”，即通过最小化数据集预测值和真实值之间的均方误差来进行模型求解。均方误差的公式为：

在线性回归中的解释即为试图在样本空间中找到一条直线，使得所有样本到直线上的欧氏距离之和最小。

- 求解过程
  均方误差显然是可导的，因此线性回归的最小二乘“参数估计”就是通过将误差方程对 w 和 b 分别进行求导，使导数为零从而得到最优解的闭式解。
  在多元线性回归的过程中，通常会将 b 吸收入向量 w