线性回归

1. 简单介绍一下线性回归。

- 线性：两个变量之间的关系**是**一次函数关系的——图象**是直线**，叫做线性。
- 非线性：两个变量之间的关系**不是**一次函数关系的——图象**不是直线**，叫做非线性。
- 回归：人们在测量事物的时候因为客观条件所限，求得的都是测量值，而不是事物真实的值，为了能够得到真实值，无限次的进行测量，最后通过这些测量数据计算**回归到真实值**，这就是回归的由来。
- 线性回归就是利用的样本D=(xi,yi)，i=1,2,3...N,xi是特征数据，可能是一个，也可能是多个，通过有监督的学习，学习到由x到y的映射h，利用该映射关系对未知的数据进行预估，因为y为连续值，所以是回归问题。

2. 线性回归的假设函数是什么形式？

线性回归的假设函数（θ0表示截距项，x0=1，方便矩阵表达）：

f(x)=θ0x0+θ1x1+θ2x2...+θnxn=θTX

其中θ,x都是列向量

3. 线性回归的代价(损失)函数是什么形式？

MSE:J(θ0,θ1)=12m∑i=1m(yi−hθ(xi))2

4. 简述岭回归与Lasso回归以及使用场景。

- 目的：
    - 解决线性回归出现的过拟合的请况。
    - 解决在通过正规方程方法求解θ的过程中出现的XTX不可逆的请况。
- 本质：
    - 约束(限制)要优化的参数

这两种回归均通过在损失函数中引入**正则化项**来达到目的：

线性回归的损失函数：

J(θ)=12m∑i=1m(hθ(x(i))−y(i))2

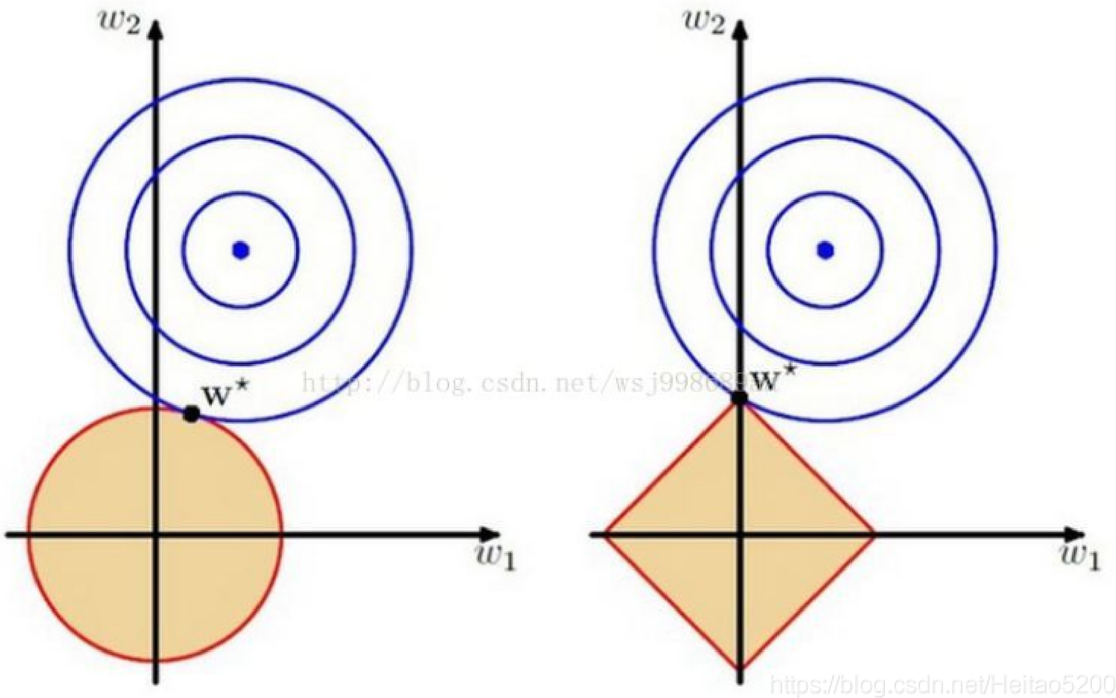
- 岭回归
    - 损失函数：

J(θ)=12m∑i=1m(hθ(x(i))−y(i))2+λ∑j=1nθj2

- Lasso回归
    - 损失函数

J(θ)=12m∑i=1m(hθ(x(i))−y(i))2+λ∑j=1n|θj|

本来Lasso回归与岭回归的解空间是全部区域，但通过正则化添加了一些约束，使得解空间变小了，甚至在个别正则化方式下，解变得稀疏了。



如图所示，这里的w1，w2都是模型的参数，要优化的目标参数，那个红色边框包含的区域，其实就是解空间，正如上面所说，这个时候，解空间“缩小了”，你只能在这个缩小了的空间中，寻找使得目标函数最小的w1，w2。左边图的解空间是圆的，是由于采用了L2范数正则化项的缘故，右边的是个四边形，是由于采用了L1范数作为正则化项的缘故，大家可以在纸上画画，L2构成的区域一定是个圆，L1构成的区域一定是个四边形。

再看看那蓝色的圆圈，再次提醒大家，这个**坐标轴和特征（数据）没关系**，它完全是参数的坐标系，每一个圆圈上，可以取无数个w1，w2，这些w1，w2有个共同的特点，用它们计算的目标函数值是相等的！那个蓝色的圆心，就是实际最优参数，但是由于我们对解空间做了限制，所以最优解只能在“缩小的”解空间中产生。

蓝色的圈圈一圈又一圈，代表着参数w1，w2在不停的变化，并且是在解空间中进行变化（这点注意，图上面没有画出来，估计画出来就不好看了），直到脱离了解空间，也就得到了图上面的那个w*，这便是目标函数的最优参数。

对比一下左右两幅图的w*，我们明显可以发现，右图的w*的w1分量是0，有没有感受到一丝丝凉意？稀疏解诞生了！是的，这就是我们想要的稀疏解，我们想要的简单模型。L1比L2正则化更容易产生稀疏矩阵。

- 补充
  - **ElasticNet 回归**： 线性回归 + L1正则化 + L2 正则化。
    - ElasticNet在我们发现用Lasso回归太过(太多特征被稀疏为0),而岭回归也正则化的不够(回归系数衰减太慢)的时候，可以考虑使用ElasticNet回归来综合，得到比较好的结果。
    - 损失函数

      J(θ)=12∑im(y(i)−θTx(i))2+λ(ρ∑jn|θj|+(1−ρ)∑jnθj2)
  - **LWR(局部加权)回归**：
    - 局部加权线性回归是在线性回归的基础上对每一个测试样本（训练的时候就是每一个训练样本）在其已有的样本进行一个加权拟合，**权重的确定**可以通过一个核来计算，常用的有**高斯核**（离测试样本越近，权重越大，反之越小），这样对每一个测试样本就得到了不一样的权重向量，所以最后得出的拟合曲线不再是线性的了，这样就增加的模型的复杂度来更好的拟合非线性数据。

- 损失函数

  $$J(\theta)=\frac{1}{2}\sum_{i=1}^{m}w^{(i)}(h_\theta(x^{(i)})-y^{(i)})^2$$

5. 线性回归要求因变量服从正态分布吗?

线性回归的假设前提是噪声服从正态分布，即因变量服从正态分布。但实际上难以达到，因变量服从正态分布时模型拟合效果更好。

参考资料： http://www.julyedu.com/question/big/kp_id/23/ques_id/2914

逻辑回归

1. 简单介绍一下逻辑回归

逻辑回归用来解决**分类**问题，线性回归的结果Y带入一个非线性变换的**Sigmoid函数**中，得到[0,1]之间取值范围的数S，S可以把它看成是一个概率值，如果我们设置概率阈值为0.5，那么S大于0.5可以看成是正样本，小于0.5看成是负样本，就可以进行分类了。

- 逻辑回归的本质： 极大似然估计
- 逻辑回归的激活函数：Sigmoid
- 逻辑回归的代价函数：交叉熵

2. 简单介绍一下Sigmoid函数

函数公式如下：

$$S(t)=\frac{1}{1+e^{-t}}$$

函数中t无论取什么值，其结果都在[0,1]的区间内，回想一下，一个分类问题就有两种答案，一种是“是”，一种是“否”，那0对应着“否”，1对应着“是”，那又有人问了，你这不是[0,1]的区间吗，怎么会只有0和1呢？这个问题问得好，我们假设分类的**阈值**是0.5，那么超过0.5的归为1分类，低于0.5的归为0分类，阈值是可以自己设定的。

好了，接下来我们把θ^TX+b带入t中就得到了我们的逻辑回归的一般模型方程：

逻辑回归的**假设函数**：

$$H(\theta,b)=\frac{1}{1+e^{(\theta^TX+b)}}$$

结果P也可以理解为概率，换句话说概率大于0.5的属于1分类，概率小于0.5的属于0分类，这就达到了分类的目的。

3. 逻辑回归的损失函数是什么

逻辑回归的损失函数是**对数似然函数**，函数公式如下：

$$cost(h_\theta(x),y)=\begin{cases}-\log(h_\theta(x)) & y=1\\ -\log(1-h_\theta(x)) & y=0\end{cases}$$

两式合并得到**概率分布表达式**：

$$(P(y|x,\theta)=h_\theta(x)^y(1-h_\theta(x))^{1-y})$$

对数似然函数最大化得到**似然函数的代数表达式**为：

$$L(\theta)=\prod_{i=1}^{m}(h_\theta(x^{(i)}))^{y^{(i)}}(1-h_\theta(x^{(i)}))^{1-y^{(i)}}$$

对似然函数对数化取反得到**损失函数表达式**：

$$J(\theta)=-\ln L(\theta)=-\sum_{i=1}^{m}(y^{(i)}\log(h_\theta(x^{(i)}))+(1-y^{(i)})\log(1-h_\theta(x^{(i)})))$$

- 为何不能用mse

解释

4.可以进行多分类吗?

多分类问题一般将二分类推广到多分类的方式有三种，一对一，一对多，多对多。

- 一对一:
    - 将N个类别两两配对，产生N(N−1)/2个二分类任务，测试阶段新样本同时交给所有的分类器，最终结果通过投票产生。
- 一对多:
    - 每一次将一个例作为正例，其他的作为反例，训练N个分类器，测试时如果只有一个分类器预测为正类，则对应类别为最终结果，如果有多个，则一般选择置信度最大的。从分类器角度一对一更多，但是每一次都只用了2个类别，因此当类别数很多的时候一对一开销通常更小(只要训练复杂度高于O(N)即可得到此结果)。
- 多对多:
    - 若干各类作为正类，若干个类作为反类。注意正反类必须特殊的设计。

5.逻辑回归的优缺点

- 优点
    - LR能以概率的形式输出结果，而非只是0,1判定。
    - LR的可解释性强、可控度高、训练速度快
- 缺点
    - 对模型中自变量多重共线性较为敏感

      例如两个高度相关自变量同时放入模型，可能导致较弱的一个自变量回归符号不符合预期，符号被扭转。需要利用因子分析或者变量聚类分析等手段来选择代表性的自变量，以减少候选变量之间的相关性;
    - 预测结果呈S型，因此从log(odds)向概率转化的过程是非线性的，在两端随着log(odds)值的变化，概率变化很小，边际值太小，slope太小，而中间概率的变化很大，很敏感。 导致很多区间的变量变化对目标概率的影响没有区分度，无法确定阀值。

6. 逻辑斯特回归为什么要对特征进行离散化。

- 逻辑回归属于广义线性模型，表达能力受限；单变量离散化为N个后，每个变量有单独的权重，相当于为模型引入了非线性，能够提升模型表达能力，加大拟合；离散特征的增加和减少都很容易，易于模型的快速迭代;
- 稀疏向量内积乘法运算速度快，计算结果方便存储，容易扩展;
- 简化模型：特征离散化以后，起到了简化了逻辑回归模型的作用，降低了模型过拟合的风险。
- 方便交叉与特征组合：离散化后可以进行特征交叉，由M+N个变量变为M∗N个变量，进一步引入非线性，提升表达能力;

7. 线性回归与逻辑回归的区别

- 线性回归的样本的输出，都是连续值，y∈(−∞,+∞)，而逻辑回归中y∈(0,1)，只能取0和1。
- 对于拟合函数也有本质上的差别:
    - 线性回归：f(x)=θTx=θ1x1+θ2x2+...+θnxn
    - 逻辑回归：f(x)=P(y=1|x;θ)=g(θTx)，其中，g(z)=11+e−z

线性回归的拟合函数，是对f(x)的输出变量y的拟合，而逻辑回归的拟合函数是对为1类样本的概率的拟合。

- 为什么要以1类样本的概率进行拟合呢，为什么可以这样拟合呢？

  θTx=0就相当于是1类和0类的决策边界：

  当θTx>0，则y>0.5；若θTx→+∞，则y→1，即y为1类;

  当θTx<0，则y<0.5；若θTx→−∞，则y→0，即y为0类;

- 这个时候就能看出区别，在线性回归中θTx为预测值的拟合函数；而在逻辑回归中θTx为决策边界。下表为线性回归和逻辑回归的区别。

  线性回归和逻辑回归的区别

	线性回归	**逻辑回归**
目的	预测	分类
y(i)	未知	（0,1）
函数	拟合函数	预测函数
参数计算方式	最小二乘法	极大似然估计

下面具体解释一下：

1. 拟合函数和预测函数什么关系呢？简单来说就是将拟合函数做了一个逻辑函数的转换，转换后使得y(i)∈(0,1);
2. 最小二乘和最大似然估计可以相互替代吗？回答当然是不行了。我们来看看两者依仗的原理：最大似然估计是计算使得数据出现的可能性最大的参数，依仗的自然是Probability。而最小二乘是计算误差损失。

8. 逻辑回归有哪些应用

- CTR预估/推荐系统的learning to rank/各种分类场景。
- 某搜索引擎厂的广告CTR预估基线版是LR。
- 某电商搜索排序/广告CTR预估基线版是LR。
- 某电商的购物搭配推荐用了大量LR。
- 某现在一天广告赚1000w+的新闻app排序基线是LR。