

Logistic regression 学习总结

于雪 经济系 15320171151914

2019.4.10

[摘要]: 逻辑回归就是将线性模型的输出值经过 Sigmoid 函数处理, 将输出值定位在 $(0, 1)$ 之间, 用于二分类的任务。它不仅能够预测出类别, 还能得到属于某个类别的概率预测。是目前应用比较广泛的一种优化算法, 利用 logistic regression 进行分类的主要思想是: 根据现有数据对分类边界线建立回归公式, 以此进行分类。“回归”一词源于最佳拟合, 表示要找到最佳拟合参数集。Logistic 回归的因变量可以是二分类的, 也可以是多分类的, 但是二分类的更为常用, 也更加容易解释。所以实际中最常用的就是二分类的 Logistic 回归。

[关键词]: 逻辑回归 二分类 多分类 Sigmoid 函数 代价函数

1. Logistic regression 基本内容

1.1 Logistic Regression 的推导过程

对于多元线性回归, 线性边界:

$$h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

其中:

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} \in R^{n+1}$$

$$x = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} \in R^{n+1} \quad \text{且 } x_0 = 1$$

预测函数:

$$h_{\theta}(x) = g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}}$$

此时根据代表着函数输出为 1 的概率, 把分类结果看成类别 1 和类别 2, 故可得概率函数:

$$P(y = 1|x; \theta) = h_{\theta}(x)$$

$$P(y = 0|x; \theta) = 1 - h_{\theta}(x)$$

将两个公式进行合并：

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

取似然函数：

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

取对数：

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

最后目标变成取 $J(\theta)$ 最小值时的 θ 为最佳参数。到这里可以采用梯度上升进行优化或者添加负系数用梯度下降进行优化，这里添加 $-1/m$ 系数：

$$J(\theta) = -\frac{1}{m} l(\theta)$$

梯度下降法进行参数更新：

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

$$\begin{aligned} \frac{\partial}{\partial \theta_j} J(\theta) &= -\frac{1}{m} \sum_{i=1}^m \left(y^i \frac{1}{h_{\theta}(x_i)} \frac{\partial}{\partial \theta_j} h_{\theta}(x_i) - (1 - y^i) \frac{1}{1 - h_{\theta}(x_i)} \frac{\partial}{\partial \theta_j} h_{\theta}(x_i) \right) \\ &= -\frac{1}{m} \sum_{i=1}^m \left(y^i \frac{1}{h_{\theta}(x_i)} - (1 - y^i) \frac{1}{1 - h_{\theta}(x_i)} \right) \frac{\partial}{\partial \theta_j} h_{\theta}(x_i) \end{aligned}$$

$$\begin{aligned}
&= -\frac{1}{m} \sum_{i=1}^m \left(y^i \frac{1}{g(\theta^T x_i)} - (1 - y^i) \frac{1}{1 - g(\theta^T x_i)} \right) g(\theta^T x_i) (1 - g(\theta^T x_i)) \frac{\delta}{\delta \theta_j} \theta^T x_i \\
&= -\frac{1}{m} \sum_{i=1}^m (y^i - g(\theta^T x_i)) \frac{\delta}{\delta \theta_j} \theta^T x_i \\
&= -\frac{1}{m} \sum_{i=1}^m (y^i - g(\theta^T x_i)) x_i^j \\
&= \frac{1}{m} \sum_{i=1}^m (g(\theta^T x_i) - y^i) x_i^j
\end{aligned}$$

故参数更新过程如下：

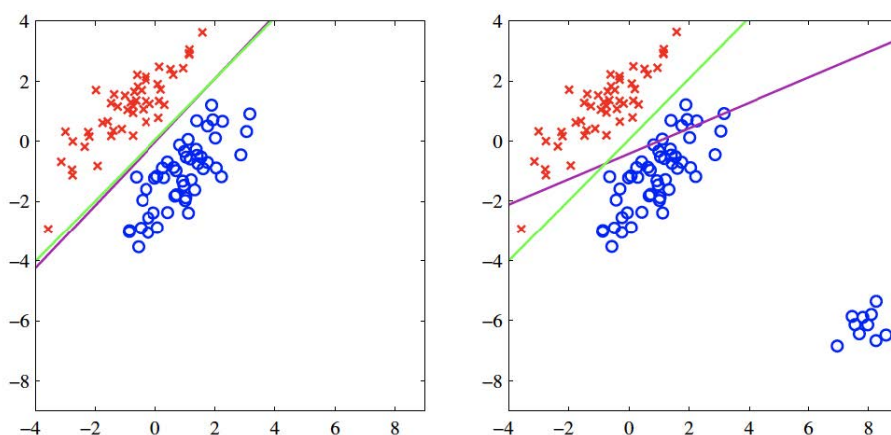
$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m (g(\theta^T x_i) - y^i) x_i^j$$

1.2 优缺点

优点：计算代价不高，易于理解和实现；预测结果是界于 0 和 1 之间的值。

缺点：容易欠拟合，分类精度可能不高；预测结果呈“S”型，概率变化很小，边际值太小，斜率太小，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，不容易确定阈值。

2. Linear vs. Logistic Regression



来自两个类别的数据由红色十字和蓝色圆圈表示，通过最小二乘（红色）和逻辑回归（绿色）找到决策边界。与逻辑回归不同，最小二乘法对异常值高度敏感。

3. 逻辑回归与机器学习相联系

机器学习最通俗的解释就是让机器学会决策。对于我们人来说，比如去菜市场里挑选芒果，从一堆芒果中拿出一个，根据果皮颜色、大小、软硬等属性或叫做特征，我们就会知道它甜还是不甜。类似的，机器学习就是把这些属性信息量化后输入计算机模型，从而让机器自动判断一个芒果是甜是酸，这实际上就是一个分类问题。

分类和回归是机器学习可以解决两大主要问题，从预测值的类型上看，连续变量预测的定量输出称为回归；离散变量预测的定性输出称为分类。例如：预测明天多少度，是一个回归任务；预测明天阴、晴、雨，就是一个分类任务。

逻辑回归由于存在易于实现、解释性好以及容易扩展等优点，被广泛应用于点击率预估（CTR）、计算广告（CA）以及推荐系统（RS）等任务中。逻辑回归虽然名字叫做回归，但实际上却是一种分类学习方法。线性回归完成的是回归拟合任务，而对于分类任务，我们同样需要一条线，但不是去拟合每个数据点，而是把不同类别的样本区分开来。

4. 总结

4.1 逻辑回归（二分类）

（1）线性回归模型： $y = W^T X + b$ 其中，

Logistics regression 问题的常规步骤为：寻找 h 函数（即 hypothesis）；构造 J 函数（损失函数）；使 J 函数最小并求得回归参数（ θ ）。

Sigmoid function:

$$\varphi(y) = \frac{1}{1 + e^{-y}}$$

映射到区间 $[0, 1]$ ，以 0.5 为阈值分成两类

cost function: (代价函数/目标函数):

首先想到 $\sum (\varphi(y) - y^{true})^2$ 利用预测值与真实值的差来作为目标，但是将 Sigmoid function 代入后发现是非凸函数，即有许多局部最小值不利于求解。所以转换思路，利用

已知 $p(y=1) = \varphi(y)$, $p(y=0) = 1 - \varphi(y)$

其中 $\varphi(y)$ 相当于属于 1 类的概率，用极大似然来估计：

$$\prod p(y) = \prod \varphi(y)^y (1 - \varphi(y))^{1-y}$$

用梯度下降求参数，梯度的负方向为代价函数下降最快的方向。

(2) 逻辑回归要点

假设数据服从伯努利分布（0-1 分布），用极大化似然函数的方法，用梯度下降来求解函数，最后达到二分类的目的。

Logistic regression 最大的优势在于它的输出结果不仅可以用于分类，还可以表征某个样本属于某类别的概率。将原本输出结果从范围 $(-\infty, +\infty)$ 映射到 $(0, 1)$ ，从而完成概率的估测。

逻辑回归判定的阈值能够映射为平面的一条判定边界，随着特征的复杂化，判定边界可能是多种多样的样貌，但是它能够较好地把两类样本点分隔开，解决分类问题。

求解逻辑回归参数的传统方法是梯度下降，构造为凸函数的代价函数后，每次沿着偏导方向（下降速度最快方向）迈进一小部分，直至 N 次迭代后到达最低点。

4.2 逻辑回归多分类

LR 是一个传统的二分类模型，它也可以用于多分类任务，其基本思想是：将多分类任务拆分成若干个二分类任务，然后对每个二分类任务训练一个模型，最后将多个模型的结果进行集成以获得最终的分类结果。一般来说，可以采取的拆分策略有：

one vs one 策略：假设我们有 N 个类别，该策略基本思想就是不同类别两两之间训练一个分类器，这时我们一共会训练出 C_N^2 种不同的分类器。在预测时，我们将样本提交给所有的分类器，一共会获得 $N(N-1)$ 个结果，最终结果通过投票产生。

one vs all 策略：该策略基本思想就是将第 i 种类型的所有样本作为正例，将剩下的所有样本作为负例，进行训练得到一个分类器。这样我们就一共可以得到 N 个分类器。在预测时，我们将样本提交给所有的分类器，一共会获得 N 个结果，我们选择其中概率值最大的那个作为最终分类结果。

参考文献

[1]Peter Harrington(著), 李锐等(译).机器学习实战. 人民邮电出版社,2013

[2]Andrew NG 《机器学习》

[3]<https://blog.csdn.net/jk123vip/article/details/80591619>

[4]应用微观计量 classification 课件,茅家铭