Logistic回归做多分类和Softmax回归

我们已经知道，普通的logistic回归只能针对二分类(Binary Classification)问题，要想实现多个类别的分类，我们必须要改进logistic回归，让其适应多分类问题。

关于这种改进，有两种方式可以做到。

第一种方式是直接根据每个类别，都建立一个二分类器，带有这个类别的样本标记为1，带有其他类别的样本标记为0。假如我们有k个类别，最后我们就得到了k个针对不同标记的普通的logistic分类器。

第二种方式是修改logistic回归的损失函数，让其适应多分类问题。这个损失函数不再笼统地只考虑二分类非1就0的损失，而是具体考虑每个样本标记的损失。这种方法叫做softmax回归，即logistic回归的多分类版本。

我们首先简单介绍第一种方式。

对于二分类问题，我们只需要一个分类器即可，但是对于多分类问题，我们需要多个分类器才行。假如给定数据集$\mathbf{X} \in \mathbb{R}^{m\times n}$，它们的标记$\mathbf{Y} \in \mathbb{R}^{k}$，即这些样本有$k$个不同的类别。

我们挑选出标记为$c(c \le k)$的样本，将挑选出来的带有标记c的样本的标记置为1，将剩下的不带有标记c的样本的标记置为0。然后就用这些数据训练出一个分类器，我们得到$h_c(x)$（表示针对标记c的logistic分类函数）。

按照上面的步骤，我们可以得到k个不同的分类器。针对一个测试样本，我们需要找到这k个分类函数输出值最大的那一个，即为测试样本的标记：

$$\arg\max_c h_c(x) \quad c = 1,2,\cdots,k$$

下面我们介绍softmax回归。

对于有k个标记的分类问题，分类函数是下面这样：

$$h_\theta(x^{(i)}) = \begin{bmatrix} p(y^{(i)}=1|x^{(i)},\theta) \\ p(y^{(i)}=2|x^{(i)},\theta) \\ \vdots \\ p(y^{(i)}=k|x^{(i)},\theta) \end{bmatrix} = \frac{1}{\sum_{c=1}^{k} e^{\theta_c^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \vdots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix}$$

在这里，我们将上式的所有的$\theta_1,\theta_2,\cdots,\theta_k$组合起来，用矩阵$\theta$来表示，即：

$$\theta = \begin{bmatrix} \theta_1^T \\ \theta_2^T \\ \vdots \\ \theta_k^T \end{bmatrix}$$

这时候，softmax回归算法的代价函数如下所示（其中$\mathrm{sign}(expression\ is\ true) = 1$）：

$$J(\theta) = -\sum_{i=1}^{m}\sum_{c=1}^{k} \mathrm{sign}(y^{(i)}=c)\log p(y^{(i)}=c|x^{(i)},\theta) = -\sum_{i=1}^{m}\sum_{c=1}^{k} \mathrm{sign}(y^{(i)}=c)\log \frac{e^{\theta_c^T x(i)}}{\sum_{l=1}^{k} e^{\theta_l^T x^{(i)}}}$$

很明显，上述公式是logistic回归损失函数的推广。

我们可以把logistic回归的损失函数改为如下形式：

$$J(\theta) = -\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) - (1-y^{(i)}(1-h_\theta x^{(i)})) = -\sum_{i=1}^{m}\sum_{c=0}^{1} \mathrm{sign}(y^{(i)} = c) \log p(y^{(i)} = c|x^{(i)}, \theta)$$

但是，需要特别注意的是，对于$p(y^{(i)} = c|x^{(i)}, \theta)$，softmax回归和logistic回归的计算方式是不同的。

对于选择softmax分类器还是k个logistic分类器，取决于所有类别之间是否互斥。所有类别之间明显互斥用softmax分类器，所有类别之间不互斥有交叉的情况下最好用k个logistic分类器。

参考资料：

http://deeplearning.stanford.edu/wiki/index.php/Softmax_Regression

http://blog.csdn.net/u013239871/article/details/51291277

分类: Machine Learning