**读书报告**

王贵涛

自然语言处理的第五章主要介绍了逻辑回归。

**分类：sigmoid函数（Classification: the sigmoid）**

对于给定的输入，经过线性运算后得到的结果的取值是整个实数域，而我们需要将结果映射到0-1之间，因此我们使用sigmoid函数。在0时结果是0.5且接近线性，在无穷时分别逼近0和1。

正特征的权值一般为正数，负特征的权值一般为负数。

逻辑回归也可以作用于句点消歧的任务。

朴素贝叶斯分类器在遇到强相关的特征时，往往会高估特征，而逻辑回归通过分配权重的方法避免了这个问题，在一些较大文档上表现更好。逻辑回归易于实现，训练速度快。

**学习逻辑回归（Learning in Logistic Regression）**

学习逻辑回归时需要做两件事。一是决定损失函数，这里我们使用交叉熵损失。二是优化算法，这里我们使用随机梯度下降。

**交叉熵损失函数（The cross-entropy loss function）**

损失函数是预测值是对应值的概率取负对数，范围是0到无穷。

**梯度下降法（Gradient Descent）**

梯度下降法是取梯度的逆方向为下降方向，取一定的步长。

随机梯度下降每次更新时取一个样本，用一个样本来近似所以样本。这种方法收敛速度更快。

mini-batch梯度下降在每次更新时用部分个样本，是梯度下降和随机梯度下降的一种折中的方法。

**正则化（Regularization）**

如果一个特征只在一个类中出现，我们将会给这个特征非常高的权值，这可能会导致过拟合的产生，因此我们使用范数进行正则化，权值越高惩罚越大。

L2范数是距离，L1范数是权值的绝对值之和。L2正则化的导数简单，更易于优化，L1正则化在部分值不可导。L2正则化倾向于许多小权值，L1正则化倾向于大权值和零权值的稀疏解。L1正则化可以看作权重的拉普拉斯先验，L2正则化可以看作均值为0的高斯分布的先验。

**多项式逻辑回归（Multinomial logistic regression）**

有时我们需要分类三类及以上，这时我们使用softmax函数，它计算样本属于每一类的概率，且概率只和为1。

多项式逻辑回归的损失函数与二元逻辑回归的损失函数有所不同，单个例子的损失函数为k个输出类的对数之和。

**解释模型（Interpreting models）**

逻辑回归的权值通常为人为设计的，因此有较强的可解释性。

逻辑回归可以与统计检验相结合，调查一个特定的特征是否重要，可以检验某些特征是否与某些结果相关。