

## SVM, AdaBoost, Logistics Regression学习策略与算法的比较

学院	年级	班级	姓名	学号
控制科学与工程	2017级	人工智能与机器人	陈逸群	201700181055

### 1 学习策略比较

对于Logistic回归模型，其学习策略为：学习一个对数几率的线性回归模型，利用该模型来学习样本的生成概率分布，属于生成模型。根据logistic分布，以二分类为例，给定输入数据 $x$ ，其输出 $y$ 的概率为

$$P(y = 1|x) = \frac{e^{w \cdot x}}{1 + e^{w \cdot x}}$$

$$P(y = 0|x) = \frac{1}{1 + e^{w \cdot x}}$$

其中 $w$ 是可学习参数。若一个事件的发生概率为 $p$ ，将一个事件几率定义为 $\frac{p}{1-p}$ ，那么logistic回归模型的对数几率为 $\text{logit}(p) = \log(\frac{p}{1-p}) = w \cdot x$ ，为线性模型。

对于SVM模型，其学习策略为：对于一个线性可分的样本空间，学习到一个具有最大几何间隔的超平面将正负样本分开，从而达到分类的目的。对于线性不可分的样本，可以采用核函数将其投影到非线性空间，对于噪声干扰大的数据，可以采用带松弛变量与惩罚系数的支持向量机，从而学到能够使正负例样本的几何间隔最大的超平面。

对于AdaBoost模型，其学习策略为：通过训练一系列比随机预测略好的弱分类器，并根据训练时的分类误差率，为每一个弱分类器分配一个权重，从而组成一个最终分类器，达到训练出一个强分类器的目标。

### 2 算法比较

对于Logistic回归模型，可以采用梯度下降法或者拟牛顿法来求解。

对于SVM，由于当样本容量变得很大的时候，通常的凸优化算法往往变得很低效，以至于无法使用，因此，可以采用SMO（序列最小最优化算法），通过每一次选取两个需要更新的参数，逐步迭代直至满足KKT条件以求得满意解。

对于AdaBoost，其求解算法为：在每一次迭代中，根据前一次的分类误差率对每一个训练样本分配不同的权重，进而对当前的基本的弱分类器进行训练，同时根据每一次迭代的分类误差率，对该基本分类器分配相应的权重。