

# 问题

机器学习中LR (Logistic Regression) 和SVM (Support Vector Machine) 有什么区别与联系？

## 背景

LR和SVM的概念大家都有了解甚至很熟悉了，不过在面试中可能不止是简单地单独考察你对LR或SVM的理解，可能会让你对这两个算法进行比较分析，因此就有必要将两者放在一起比较一下。

## LR和SVM的联系

### 1. LR和SVM都是分类算法

普通的LR和SVM算法只能处理二分类问题，当然，通过改进后的LR和SVM都可以用来处理多分类问题（后面会详细解释）。

### 2. 在不考虑核函数时，两者都是线性分类算法

注意，不考虑核函数时两者都是**线性**分类器。LR、SVM加了核函数后分别为KLR、KSVM，只不过一般而言采用KSVM较多而KLR用得较少。

### 3. 两者都属于监督学习算法

### 4. 两者都是判别式模型

什么是判别式模型？假设给定观测集合X，需要预测的变量集合为Y，那么判别式模型就是直接**对条件概率分布P(Y|X)进行建模**来预测Y；而生成式模型是指，先对联合概率模型P(X,Y)进行建模，然后在给定观测集合X的情况下，通过计算边缘分布来求解出P(Y|X)。

常见的判别式模型有：LR、SVM、KNN、神经网络、最大熵模型、条件随机场等

常见的生成式模型有：隐马尔科夫模型HMM、朴素贝叶斯、贝叶斯网络、高斯混合模型GMM等

## LR和SVM的区别

### 1. 采用的Loss Function不同

从目标函数来看，LR采用的是Logistic Loss，而SVM采用的是Hinge Loss。

$$LR\ Loss : L(\omega, b) = \sum_{i=1}^m \ln(y_i p_1(x; \beta) + (1 - y_i) p_0(x; \beta)) = \sum_{i=1}^m (-y_i \beta^T x_i + \ln(1 + e^{\beta^T x_i})) \quad (1)$$

$$\text{其中, } \beta = (\omega; b), p_1 = p(y = 1|x; \beta), p_0 = p(y = 0|x; \beta)$$

$$SVM\ Loss : L(\omega, b, \alpha) = \frac{1}{2} \|\omega\|^2 + \sum_{i=1}^m \alpha_i (1 - y_i (\omega^T x_i + b))$$

LR：基于概率理论，通过极大似然估计方法估计参数值

SVM：基于几何间隔最大化原理

补充：	Logistic Loss：	$L_{log}(z) = \log(1 + e^{-z})$	// 用于 LR
	Hinge Loss：	$L_{hinge}(z) = \max(0, 1 - z)$	// SVM

## 2. SVM只考虑边界线上局部的点（即support vector），而LR考虑了所有点。

影响SVM决策分类面的只有少数的点，即边界线上的支持向量，其他样本对分类决策面没有任何影响，即SVM不依赖于数据分布；而LR则考虑了全部的点（即依赖于数据分布），通过非线性映射，减少远离分类平面的点的权重，即对不平衡的数据要先做balance。

## 3. 在解决非线性问题时，SVM采用核函数机制，而LR一般很少采用核函数的方法。

SVM使用的是hinge loss，可以方便地转化成对偶问题进行求解，在解决非线性问题时，引入核函数机制可以大大降低计算复杂度。

## 4. SVM依赖于数据分布的距离测度，所以需对数据先做normalization，而LR不受影响。

normalization的好处：进行梯度下降时，数值更新的速度一致，少走弯路，可以更快地收敛。

## 5. SVM的损失函数中自带正则化项( $\frac{1}{2} ||w||^2$ )，而LR需要另外添加。

# LR和SVM什么时候用？

来自Andrew Ng的建议：

- ①若feature数远大于样本数量，使用LR算法或者Linear SVM
- ②若feature数较小，样本数量适中，使用kernel SVM
- ③若feature数较小，样本数很大，增加更多的feature然后使用LR算法或者Linear SVM

# LR和SVM如何处理多分类问题？

## SVM：

方式一：组合多个二分类器来实现多分类器（两种方法OvO或OvR）

①OvO（one-versus-one）：任意两个类别之间设计一个二分类器，N个类别一共 $\frac{N(N-1)}{2}$ 个二分类器

②OvR（one-versus-rest）：每次将一个类别作为正例，其余的作为反例，共N个分类器。

注：OvO和OvR先训练出多个二分类器，在测试时，新样本将同时提交给所有的分类器进行预测，投票产生最终结果，将被预测的最多的类别作为最终的分类结果

方式二：直接修改目标函数，将多个分类面的参数合并到一个最优化问题中，一次性实现多分类。

## LR：

方式一：OvR：同上，组合多个logistic 二分类器

方式二：修改目标函数，改进成softmax回归

## 参考资料

[LR和SVM的相同和不同](#)

[SVM学习笔记——SVM解决多分类问题的方法](#)

[逻辑回归解决多分类和softmax](#)