

# 支持向量机通俗导论 (理解SVM的三层境界)

July, pluskid 著  
Codejedi 编辑



# Contents

<b>1</b>	<b>第一层——了解SVM</b>	<b>5</b>
1.1	什么是支持向量机SVM	5
1.2	线性分类	5
1.2.1	分类标准	5
1.2.2	1或-1分类标准的起源: logistic回归	6
<b>2</b>	<b>第二层——深入SVM</b>	<b>7</b>
2.1	从线性可分到线性不可分	7
2.1.1	从原始问题到对偶问题的求解	7
<b>3</b>	<b>第三层——证明SVM</b>	<b>9</b>
	<b>Bibliography</b>	<b>11</b>
	<b>Books</b>	<b>11</b>
	<b>Articles</b>	<b>11</b>
	<b>Index</b>	<b>13</b>



## 1 — 第一层——了解SVM

### 1.1 什么是支持向量机SVM

要明白什么是SVM，便得从分类说起。

分类作为数据挖掘领域中一项非常重要的任务，它的目的是学会一个分类函数或分类模型(或者叫做分类器)，而支持向量机本身便是一种监督式学习的方法(至于具体什么是监督学习与非监督学习，请参见此系列*Machine L&Data Mining*第一篇)，它广泛的应用于统计分类以及回归分析中。

支持向量机（SVM）是90年代中期发展起来的基于统计学习理论的一种机器学习方法，通过寻求结构化风险最小来提高学习机泛化能力，实现经验风险和置信范围的最小化，从而达到在统计样本量较少的情况下，亦能获得良好统计规律的目的。

通俗来讲，它是一种二类分类模型，其基本模型定义为特征空间上的间隔最大的线性分类器，即支持向量机的学习策略便是间隔最大化，最终可转化为一个凸二次规划问题的求解。

对于不想深究SVM原理的同学或比如就只想看看SVM是干嘛的，那么，了解到这里便足够了，不需上层。而对于那些喜欢深入研究一个东西的同学，甚至究其本质的，咱们则还有很长的一段路要走，万里长征，咱们开始迈第一步吧，相信你能走完。

### 1.2 线性分类

OK，在讲SVM之前，咱们必须先弄清楚一个概念：线性分类器(也可以叫做感知机，这里的机表示的是一种算法，本文第三部分、证明SVM中会详细阐述)。

#### 1.2.1 分类标准

这里我们考虑的是一个两类的分类问题，数据点用 $x$ 来表示，这是一个 $n$ 维向量， $w^T$ 中的 $T$ 代表转置，而类别用 $y$ 来表示，可以取1或者-1，分别代表两个不同的类。一个线性分类器的学习目标就是要在 $n$ 维的数据空间中找到一个分类超平面，其方程可以表示为：

$$w^T x + b = 0$$

上面给出了线性分类的定义描述，但或许读者没有想过：为何用 $y$ 取1 或者-1来表示两个不同的类别呢？其实，这个1或-1的分类标准起源于logistic回归，为了完整和过渡的自然性，咱们就再来看看这个logistic回归。

### 1.2.2 1或-1分类标准的起源：logistic回归

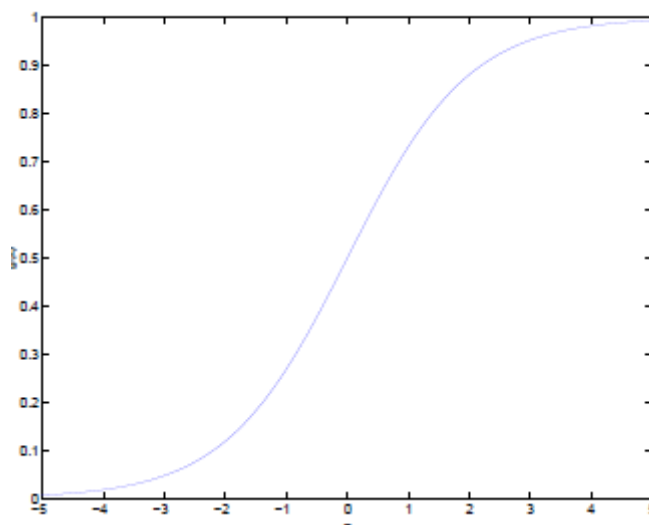
Logistic回归目的是从特征学习出一个0/1分类模型，而这个模型是将特性的线性组合作为自变量，由于自变量的取值范围是负无穷到正无穷。因此，使用logistic函数（或称作sigmoid函数）将自变量映射到(0,1)上，映射后的值被认为是属于 $y=1$ 的概率。形式化表示就是

假设函数

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

其中 $x$ 是 $n$ 维特征向量，函数 $g$ 就是logistic函数。

而 $g(z) = \frac{1}{1+e^{-z}}$ 的图像是



可以看到，将无穷映射到了(0,1)。

而假设函数就是特征属于 $y=1$ 的概率。

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

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

当我们要判别一个新来的特征属于哪个类时，只需求，若大于0.5就是 $y=1$ 的类，反之属于 $y=0$ 类。

再审视一下 $h_{\theta}(x)$ ，发现 $h_{\theta}(x)$ 只和 $\theta^T x$ 有关， $\theta^T x > 0$ ，那么 $h_{\theta}(x) > 0.5$ ， $g(z)$ 只不过是用来映射，真实的类别决定权还在 $\theta^T x$ 。还有当时 $\theta^T x \gg 0$ ， $h_{\theta}(x) = 1$ ，反之 $h_{\theta}(x) = 0$ 。如果我们只从 $\theta^T x$ 出发，希望模型达到的目标无非就是让训练数据中 $y = 1$ 的特征 $\theta^T x \gg 0$ ，而是 $y = 0$ 的特征 $\theta^T x \ll 0$ 。Logistic回归就是要学习得到 $\theta$ ，使得正例的特征远大于0，负例的特征远小于0，强调在全部训练实例上达到这个目标。



从线性可分到线性不可分  
从原始问题到对偶问题的求解


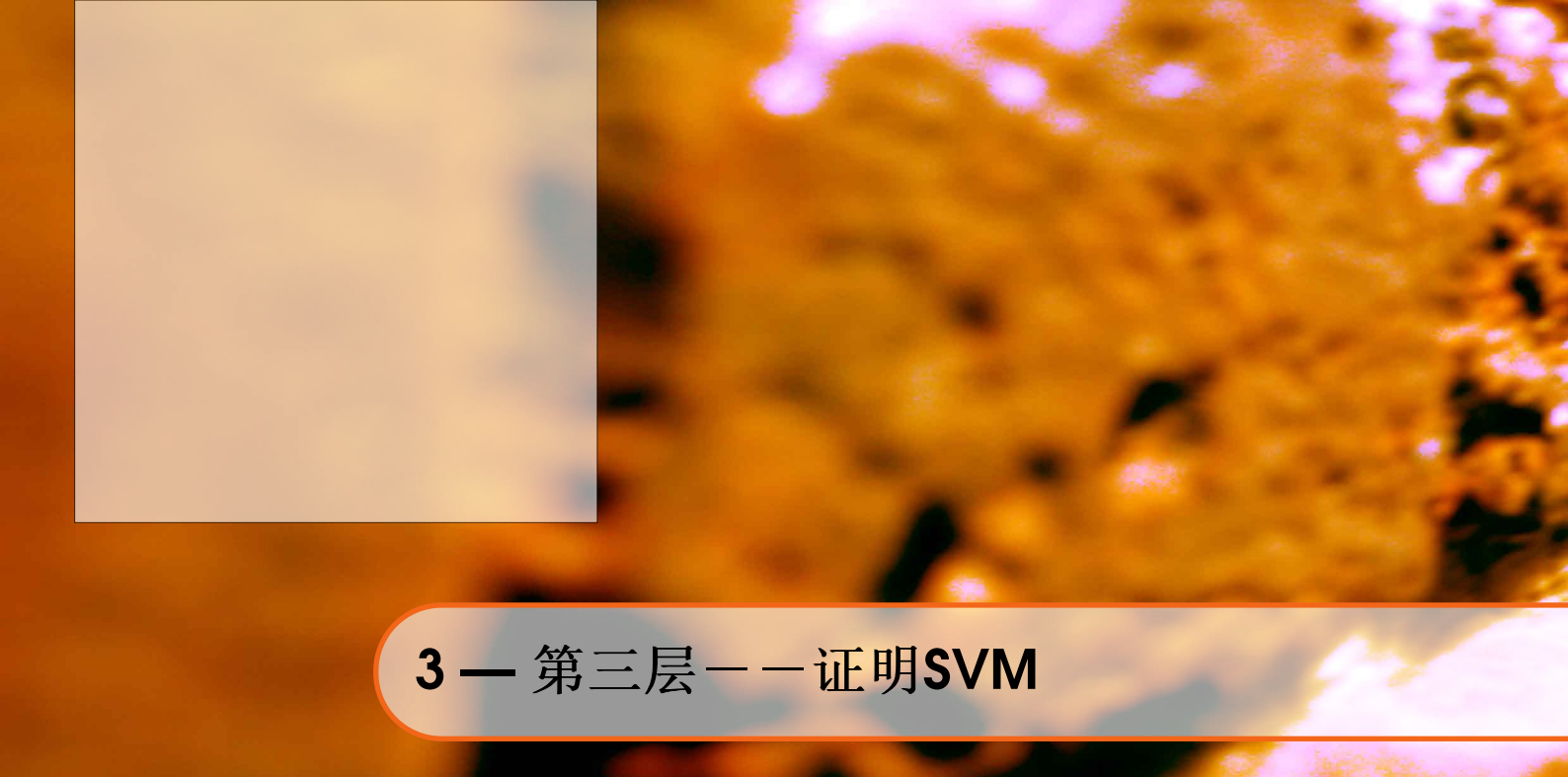
## 2 — 第二层——深入SVM

**2.1** 从线性可分到线性不可分

**2.1.1** 从原始问题到对偶问题的求解







### 3 — 第三层——证明SVM





## Bibliography

**Books**

**Articles**



# Index

1或-1分类标准的起源: logistic回归, 6

Paragraphs of Text, 5

从线性可分到线性不可分, 7

从原始问题到对偶问题的求解, 7

分类标准, 5

线性分类, 5