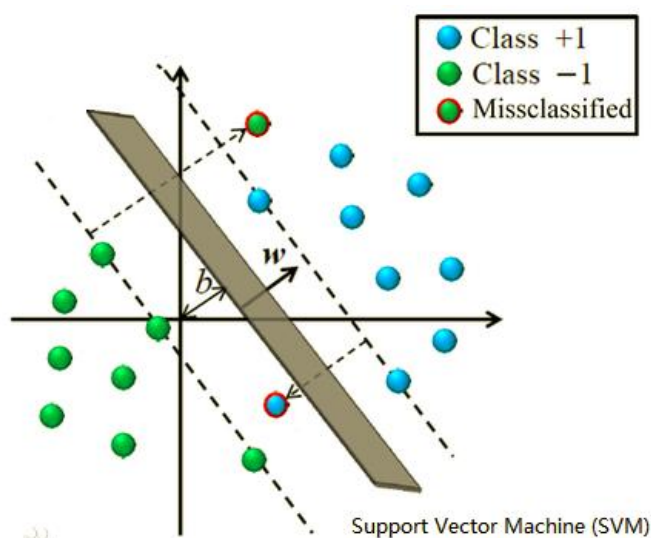


支持向量机 SVM

Support Vector Machine

支持向量机 (SVM) 是一类有监督学习算法, SVM 对数据进行二元分类的广义线性分类器, 其决策边界是对学习样本求解的最大边距超平面。SVM 通常用来进行模式识别、分类以及回归分析。

SVM 使用铰链损失函数 (hinge loss) 计算经验风险 (empirical risk) 并在求解系统中加入了正则化项以优化结构风险 (structural risk), 是一个具有稀疏性和稳健性的分类器。SVM 可以通过核方法 (kernel method) 进行非线性分类, 是常见的核学习 (kernel learning) 方法之一。



SVM 方法是通过一个非线性映射 ϕ , 把样本空间映射到一个高维乃至无穷维的特征空间中 (Hilbert 空间), 使得在原来的样本空间中非线性可分的问题转化为在特征空间中的线性可分的问题. 简单地说, 就是升维和线性化. 升维, 就是把样本向高维空间做映射, 一般情况下这会增加计算的复杂性, 甚至会引起"维数灾难", 因而人们很少问津. 但是作为分类、回归等问题来说, 很可能在低维样本空间无法线性处理的样本集, 在高维特征空间中却可以通过一个线性超平面实现线性划分(或回归). 一般的升维都会带来计算的复杂化, SVM 方法巧妙地解决了这个难题: 应用核函数的展开定理, 就不需要知道非线性映射的显式表达式; 由于是在高维特征空间中建立线性学习机, 所以与线性模型相比, 不但几乎不增加计算的复杂性, 而且在某种程度上避免了"维数灾难". 这一切要归功于核函数的展开和计算理论.