SVM

间隔与支持向量

给定训练样本集$D = (x_i, y_1), (x_2, y_2), \ldots, (x_m, y_m), y_i \in +1, -1$

，分类学习最基本的想法就是基于训练集D在样本空间中找到一个划分超平面，将不同类别的样本分开. 但能将训练样本分开的划分超平面可能有很多
在样本空间中，划分超平面可通过如下线性方程来描述:

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

其中$w = (w_1; w_2; ..; w_d)$为法向量决定了超平面的方向; b为位移项，决定了超平面与原点之间的距离. 显然，划分超平面可被法向量w和位移b确定，下面我们将其记为(w,b).样本空间中任意点x到超平面(w,b)的距离可写为

$$r = \frac{|w^T x + b|}{||w||}$$

假设超平面(w,b)能将样本正确分类，即对于
$(x_i, y_i) \in D 若\, y_i = +1, 则有\, w^T x_i + b > 0, 若\, y_i = -1, 则有\, w^T x_i + b < 0$。令

$$当\, w^T x_i + b > 1 时, y_i = 1; 当\, w^T x_i + b < -1 时, y_i = -1$$

距离超平面最近的这几个训练样本点使上式的等号成立，它们被称为“支持向量”(support vector), 两个异类支持向量到超平面的距离之和为

$$r = \frac{2}{||w||}$$

欲找到具有“最大间隔”(maximum margin)的划分超平面，也就是要找到能满足式中约束的参数w和b,使得r最大，即

$$max_{w,b} \frac{2}{||w||}$$

$$s.t. y_i(w^T x_i + b) \geq 1,\ \ i = 1, 2, .., m$$

可重写为

$$min_{w,b} \frac{1}{2} ||w||$$

$$s.t. y_i(w^T x_i + b) \geq 1,\ \ i = 1, 2, .., m$$

对偶问题

上一个式子可以用拉格朗日乘子法，写出拉格朗日函数，对w,b求导后带入可得其对偶问题

$$\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$$s.t. \sum_{j=1}^{m} \alpha_i y_i = 0$$

$$\alpha_i > 0, i = 1, 2, 3..., m$$

核函数

在前面我们假设训练样本是线性可分的，即存在一个划分超平面能将训练样本正确分类.然而在现实任务中，原始样本空间内也许并不存在一个能正确划分两类样本的超平面.对这样的问题，可将样本从原始空间映射到一个更高维的特征空间,使得样本在这个特征空间内线性可分，例如，若将原始的二维空间映射到一个合适的三维空间，就能找到一-个合适的划分超平面.幸运的是，如果原始空间是有限维，即属性数有限，那么一-定存在一个高维特征空间使样本可分.
令Φ表示将x映射后的特征向量,于是,在特征空间中划分超平面所对应的模型可表示为

$$f(x) = w^T \Phi(x) + b = 0$$

其中w,b为模型参数，则有

$$min_{w,b} \frac{1}{2} ||w||$$

$$s.t. y_i(w^T \Phi(x_i) + b) \geq 1，i = 1, 2, .., m$$

其对偶问题是

$$\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \Phi(x_i)^T \Phi(x_j)$$

$$s.t. \sum_{j=1}^{m} \alpha_i y_i = 0$$

$$\alpha_i > 0, i = 1, 2, 3..., m$$

求解$\Phi(x_i)^T \Phi(x_j)$,这是样本x_i, x_j映射到特征空间之后的内积.由于特征空间维数可能很高，甚至可能是无穷维，因此直接计算$\Phi(x_i)^T \Phi(x_j)$通常是困难的.为了避开这个障碍，可以设想这样一个函数:

$$k(x_i, x_j) =< \Phi(x_i), \Phi(x_j) >= \Phi(x_i)^T \Phi(x_j)$$

于是，可写为

$$\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j k(x_i, x_j)$$

$$s.t. \sum_{j=1}^{m} \alpha_i y_i = 0$$

$$\alpha_i > 0, i = 1, 2, 3..., m$$

这里的函数k(;)就是核函数，常用核函数如下
线性核 $k(x_i, x_j) = x_i^T x_j$
多项式核 $k(x_i, x_j) = (x_i^T x_j)^d, d \geq 1 为多项式的次数$
高斯核 $k(x_i, x_j) = exp(\frac{||x_i - x_j||^2}{2\sigma^2}), \sigma > 0$