

我们通常说的核函数指的是正定和函数，其充要条件是对于任意的 $x$ 属于 $X$ ，要求 $K$ 对应的Gram矩阵要是半正定矩阵。RBF核径向基，这类函数取值依赖于特定点间的距离，所以拉普拉斯核其实也是径向基核。SVM关键是选取核函数的类型，常用核函数主要有线性内核，多项式内核，径向基内核（RBF），sigmoid核。

## 线性核函数

$$\kappa(x, x_i) = x \cdot x_i$$

线性核，主要用于线性可分的情况，我们可以看到特征空间到输入空间的维度是一样的，其参数少速度快，对于线性可分数据，其分类效果很理想，因此我们通常首先尝试用线性核函数来做分类，看看效果如何，如果不行再换别的

## 多项式核函数

$$\kappa(x, x_i) = ((x \cdot x_i) + 1)^d$$

多项式核函数可以实现将低维的输入空间映射到高维的特征空间，但是多项式核函数的参数多，当多项式的阶数比较高的时候，核矩阵的元素值将趋于无穷大或者无穷小，计算复杂度会大到无法计算。

## 高斯（RBF）核函数

$$\kappa(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{\delta^2}\right)$$

高斯径向基函数是一种局部性强的核函数，其可以将一个样本映射到一个更高维的空间内，该核函数是应用最广的一个，无论大样本还是小样

本都有比较好的性能，而且其相对于多项式核函数参数要少，因此大多数情况下在不知道用什么核函数的时候，优先使用高斯核函数。

sigmoid核函数

$$\kappa(x, x_i) = \tanh(\eta \langle x, x_i \rangle + \theta)$$

采用sigmoid核函数，支持向量机实现的就是一种多层神经网络。

因此，在选用核函数的时候，如果我们对我们的数据有一定的先验知识，就利用先验来选择符合数据分布的核函数；如果不知道的话，通常使用交叉验证的方法，来试用不同的核函数，误差最小的即为效果最好的核函数，或者也可以将多个核函数结合起来，形成混合核函数。

在吴恩达的课上，也曾经给出过一系列的选择核函数的方法：

如果特征的数量大到和样本数量差不多，则选用LR或者线性核的SVM；

如果特征的数量小，样本的数量正常，则选用SVM+高斯核函数；

如果特征的数量小，而样本的数量很大，则需要手工添加一些特征从而变成第一种情况。