之前介绍了算法，其基本思想就是**就是要找出一个分类面使得离它最近的点到它的距离最大。**并由此衍生不几种不同的情况，1）数据线性可分；2）数据在当前特征空间中线性不可分，但是通过非线性的特征变换使得在变换后的特征空间中数据线性可分；3）更好的鲁棒性的 。为方便计算引入核函数后，判决函数如下：



之前讲过，对于判决函数而言有两种形式，一种是直接形式，如上式，还有一种是写成概率的形式：，而就是通过引入贝叶斯框架，在先验参数的结构下基于主动相关决策理论（ARD），来移除不相关的点，从而获得稀疏化的模型。

**回归模型**

给定一组训练数据对，假设我们的训练数据是从下面的回归模型中采样获得的：，其中是采样过程中的噪声项，假设服从高斯分布，则：



也就是说，假设我们的训练数据是从一个服从高斯分布的概率模型中采样的，其均值为，方差为，其中与的模型一致。对于采样到的数据对，最有可能的模型参数可以通过的方法求得，可以得到似然函数：



其中，是大小为的’design’ matrix：，其中，是核函数，与中的一致。

按常规方法，对似然函数求导并令导数为零可求出参数，而关键点在于给中的每一个分量都分配一个服从高斯分布的先验：



其中，，是个超参数，这个超参数之间是相互独立的。

定义完超参数，下面就需要求出各种参数的值，根据贝叶斯推断：



参数求出来后，新的测试点，有：



显然，（2）式的计算是比较困难的，因为归一化因子这个积分形式比较难求出来（如下）：



实际上，我们在求后验概率的时候，可以分解为先通过训练数据求出超参数，再通过参数求出：



（4）式右边第一部分可以由式求得：



其中， ，后验概率的均值和方差分别为：









其中，，。

1. 式右边第二部可以通过最大化下式求得：



**超参数优化：**

方法一：

对（5）式求导，并令其为0：



其中，。

方法二：

EM算法迭代求参数，具体请参考给出的参考文献。

**实际上，最优结果显示，超参数的大部分都会趋于一个特别大的值（原则上式无穷大），因此，对应于这些超参数的权参数的后验概率。因此这些参数以及对应的基函数被从模型中去掉，对于新输入的预测没有作用，对应与剩下的非零权值的输入被称为相关向量。**

找到最优的超参数后，我们可以预测出新的测试数据的分布：



其中：，。

**分类模型**

将相关向量机的框架推广的分类问题，此时模型变为：



大体思路与回归模型一样，不过对参数的求解上有所不同，具体参考Tipping（2001）

的方法。

参考文献：

1. Tipping,《Sparse Bayesian Learning and the Relevance Vector Machine》；
2. TODD K.MOON,《The EM algorithm》；
3. Dimitris G.Tzikas,《the variational approximation for bayesian inference》。



2017.08.13.