* 单变量正态密度函数的均值学习

设一个模式样本集，其类概率密度函数是单变量正态分布N(θ,σ2)，均值θ待求，即



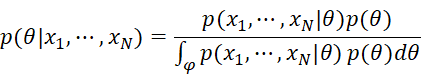
给出N个训练样本{x1, x2,…, xN}，用贝叶斯学习计算其均值估计量。

设最初的先验概率密度p(θ)为，这里θ0是凭先验知识对未知量θ的“最好”推测，表示上述推测的不确定性度量。这里可以假定p(θ)是正态的，因为均值的估计量是样本的线性函数，因样本x是正态分布的，因此p(θ)取为正态分布是合理的，这样计算起来可比较简单。

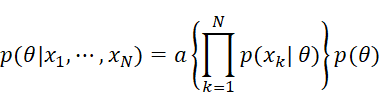
初始条件已知，即p(θ)为，p( x1|θ)为N(θ,σ2)，由贝叶斯公式p(θ| x1)=a p( x1|θ) p(θ)，可得：



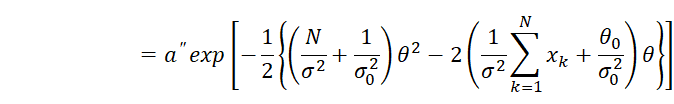
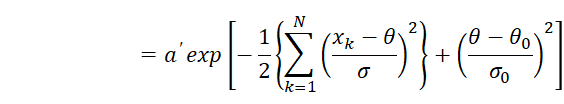
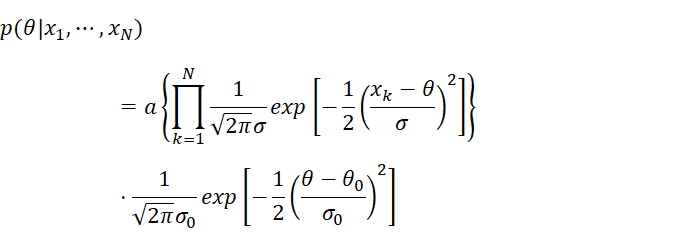
其中a是一定值。由贝叶斯法则有：



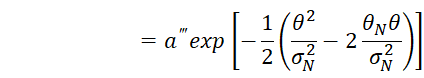
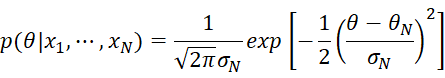
这里φ表示整个模式空间。由于每一次迭代是从样本子集中逐个抽取一个变量，所以N次运算是独立地抽取N个变量，因此上式可写成：



代入p( xk|θ) 和p(θ)的值，得：



上式每一步中与θ无关的项都并入常数项和，这样p(θ| x1, …, xN)是θ平方函数的指数集合，仍是一正态密度函数。将它写成的形式，即：



将上述两式相比较，得：



解出θN和σN，得：



即根据对训练样本集{xi}i=1,2,…,N的观察，求得均值θ的后验概率密度p(θ| xi)为，其中θN是经过N个样本观察之后对均值的最好估计，它是先验信息（即θ0，和σ2）与训练样本所给信息（即N和）适当结合的结果，是用N个训练样本对均值的先验估计θ0的补充；是对这个估计的不确定性的度量，因随N的增加而减小，因此当时，趋于零。由于θN是和θ0的线性组合，两者的系数都非负且其和为1，因此只要，当时，θN趋于样本均值的估计量。

图中所示为一正态密度的均值学习过程，每增加一次对样本的预测，都可减小对θ估计的不确定性，所以p(θ| x1, …, xN)变得越来越峰形突起，且其均值与估计量之间的偏差的绝对值亦越来越小。

