第一章

笔记本:	PRML读书笔记		
创建时间:	2018/7/22 23:02	**更新时间:**	2018/7/31 11:13
作者:	王		
标签:	简介		

1.1
模式识别所领域所关心的问题：通过计算机算法自动发现数据之间的规则，并应用这些规则（比如分类，回归预测等）

在做多项式拟合问题的时候，增加样本数据也是减少过拟合的方法，但是我们不能根据数据量的多少来决定模型参数多少，而是应该根据要解决的问题的复杂度来决定参数量

1.2
var[x] = E[x^2] - E[x]^2
conv(x,y)=E[(x-E[x])(y-E[y])]=E[xy]-E[x]E[y]

频率学派与贝叶斯学派:
简单来说，当我们在解决一个问题的时候，一般是构建一个问题模型来求解，模型是由许多参数组成的，我们的目的就是找到这些参数，频率学派与贝叶斯学派的最大不同之处就在于对于参数空间的认知上面。频率学派认为一个模型的参数是固定的，求解问题时要做的就是要找到在参数空间中最有可能的参数，所以有极大似然估计和置信区间。而贝叶斯学派则认为参数空间中的值都是可能的，每个值都拥有各自的概率，所以有先验概率和后验概率。

贝叶斯学派的优点：采用先验概率更加自然，显得不那么极端。比如抛硬币，连续三次出现正面，频率学派会将出现正面的概率视为1，而贝叶斯学派却不会。
贝叶斯学派的缺点：依赖于先验知识的选择，先验选择的不好，结果就可能会很差

单变量高斯分布函数：x为单变量

$$\mathcal{N}\left(x|\mu,\sigma^2\right)=\frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \qquad (1.46)$$

多变量高斯分布函数：x为D维向量， Σ为D*D的协方差矩阵， |Σ|为协方差矩阵的行列式

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})=\frac{1}{(2\pi)^{D/2}}\frac{1}{|\boldsymbol{\Sigma}|^{1/2}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\} \qquad (1.52)$$

极大似然估计

极大似然估计存在问题：样本方差的计算是一个有偏估计，因为它有系统性误差，无论怎样进行抽样，样本方差值总是小于理论方差值。之所以会出现这种系统性误差，是由于在计算均值的时候采用的是样本均值而不是总体均值，样本均值比理论上的总体均值更加靠近这一组样本中心。

求解过程:

$$\mathcal{N}\left(x|\mu,\sigma^2\right)=\frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \qquad (1.46)$$

$$p(\mathbf{x}|\mu,\sigma^2)=\prod_{n=1}^{N}\mathcal{N}\left(x_n|\mu,\sigma^2\right). \qquad (1.53)$$

根据上述两个式子，我们可以得到求最大似然估计的对数概率函数：

$$\ln p\left(\mathbf{x}|\mu,\sigma^2\right)=-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2-\frac{N}{2}\ln\sigma^2-\frac{N}{2}\ln(2\pi). \qquad (1.54)$$

分别进行求导计算之后得到优化目标函数：

$$\mu_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N} x_n \tag{1.55}$$

$$\sigma^2_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu_{\mathrm{ML}})^2 \tag{1.56}$$