曲线拟合之概率回访

emacsun

目录

1 回忆 1

2 贝叶斯估计 1

3 概率模型 3

4 我们离真正的贝叶斯估计有多远 4

1 回忆

我们之前在 曲线拟合的过程中 采用的是最小化误差函数的方法来确定拟合的$\mathbf{w}$系数。拟合的多项式为：

$$y(x,\mathbf{w}) = w_0 + w_1 x + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j \tag{1.1}$$

误差函数为：

$$E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n,\mathbf{w}) - t_n\}^2 \tag{1.2}$$

在那里，我们为了解决过度拟合问题还采用了一种叫做正则化的方法。今天，我们从概率的角度来审视多项式曲线拟合问题。通过概率角度，我们可以更深入的理解误差函数和正则化。

2 贝叶斯估计

曲线拟合的目的是对于给定输入x估计输出t。当然，我们有训练数据：对于$\mathbf{x} = (x_1,\ldots,x_N)^T$，对应的值是$\mathbf{t} = (t_1,\ldots,t_N)^T$。对于任意的新的输入值$x$，

我们可以把对t的估计写成一个条件概率估计。什么样的概率密度函数最合适呢？正态分布最合适 !即，对于给定的输入x，我们假设t具有正态分布，均值是$y(x,\mathbf{w})$:

$$p(t|x,\mathbf{w},\beta)=\mathcal{N}(t|y(x,\mathbf{w}),\beta^{-1}) \tag{2.1}$$

其中β是精度参数，等于(2.1)的方差的导数，即$\beta^{-1}=\sigma^2$。

式(2.1)的示意图如1所示。

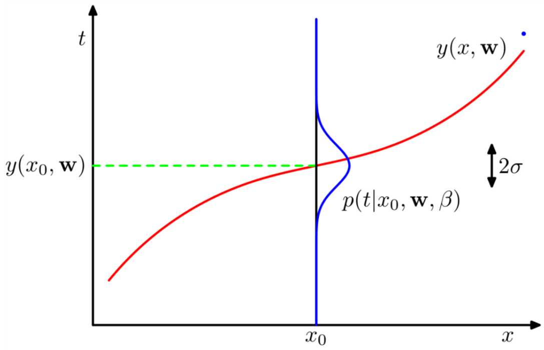

图 1: 式(2.1)的示意图

从图1 中可以看出蓝色曲线就是假设的高斯分布。而精度值β体现了分布的方差。

现在我们用不同于 以往 的方法来求$\mathbf{w},\beta$。如果所有的训练数据都是从(2.1)中独立获得的，也就是说假设$\mathbf{t}$是独立同分布的。那么关于$\mathbf{t},\mathbf{x}$分布的似然函数是：

$$p(\mathbf{t}|\mathbf{x},\mathbf{w},\beta)=\prod_{n=1}^{N}\mathcal{N}(t_n|y(x_n,\mathbf{w})\beta^{-1}) \tag{2.2}$$

其中:

$$\mathcal{N}(y|\mu,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2.3}$$

把(2.3)带入(2.2)，并对(2.2)左右两端取自然对数：

$$\ln p(\mathbf{t}|\mathbf{x},\mathbf{w},\beta)=-\frac{\beta}{2}\sum_{n=1}^{N}(y(x_n,\mathbf{w})-t_n)^2+\frac{N}{2}\ln\beta-\frac{N}{2}\ln(2\pi) \tag{2.4}$$

我们从(2.4)推出曲线拟合系数$\mathbf{w}$的最大似然解。显然，我们可以忽略(2.4)的后两项，因为这两项与$\mathbf{w}$没有关系。另外我们也发现$\mathbf{w}$的最大似然解与等号右边第一项的系数也没有关系，这个系数只是起到缩放作用，我们还可以把$\beta/2$用$1/2$代替。最大化似然函数等效于最小化负的似然函数。最后我们发现最大化(2.4)和

最小化(1.2)是一回事儿。 **因此(1.2)所示的误差函数最小值的解是假定噪声为高斯噪声的最大似然解。**

另外我们还可以使用最大似然准则求得精度值β的最优解。把(2.4)当做β的函数，我们有β的最大似然解满足：

$$\frac{1}{\beta_{ML}} = \frac{1}{N}\sum_{n=1}^{N}\sum_{n=1}^{N}(y(x_n, \mathbf{w}) - t_n)^2 \tag{2.5}$$

所以我们可以先求得$\mathbf{w}$的最大似然解$\mathbf{w_{ML}}$，然后求得$\frac{1}{\beta_{ML}}$。如此，我们便得到了所需高斯分布的两个重要参数，对于任意输入x，我们可以使用这个模型来估计t。

3 概率模型

现在我们有了$\mathbf{w}_{ML}$，$\frac{1}{\beta_{ML}}$，我们就有了一个概率模型：

$$p(t|x, \mathbf{w}_{ML}, \beta_{ML}) = \mathcal{N}(t|y(x, \mathbf{w}_{ML}), \beta_{ML}^{-1}) \tag{3.1}$$

对于给定的x我们用(1.1)来计算其均值$y(x, \mathbf{w}_{ML})$，然后用(3.1)给出t的估计。

现在让我们更深入的理解这个问题。首先，我们引入对(1.1)中系数$\mathbf{w}$的一个先验估计：

$$p(\mathbf{w}|\alpha) = \mathcal{N}(\mathbf{w}|\mathbf{0}, \alpha^{-1}\mathbf{I}) = (\frac{\alpha}{2\pi})^{(M+1)/2}\exp(-\frac{\alpha}{2}\mathbf{w}^T\mathbf{w}) \tag{3.2}$$

其中α是先验概率分布的精度。$M+1$是M阶多项式中的系数个数。α控制着模型的参数（式(1.1)的参数），我们称α为超参数。据贝叶斯理论$\mathbf{w}$的后验分布与先验分布和似然函数成比例，即：

$$p(\mathbf{w}|\mathbf{x}, \mathbf{t}, \alpha, \beta) \propto p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta)p(\mathbf{w}|\alpha) \tag{3.3}$$

利用给定的训练数据，我们通过最大化后验概率来确定$\mathbf{w}$。这个准则叫做最大后验概率准则(maximum posterior, MAP). 结合(3.3)(2.4)(3.2)，我们发现最大后验概率等效于最小化(3.4)：

$$\frac{\beta}{2}\sum_{n=1}^{N}(y(x_n, \mathbf{w}) - t_n)^2 + \frac{\alpha}{2}\mathbf{w}^T\mathbf{w} \tag{3.4}$$

即，最大化后验概率等效于最小化带有正则参数$\lambda = \alpha/\beta$的均方误差函数。

4 我们离真正的贝叶斯估计有多远

截止目前，尽管我们引入了$\mathbf{w}$的一个先验估计$p(\mathbf{w}|\alpha)$，但是我们还是在做$\mathbf{w}$的点估计，算不得真正的贝叶斯方法。因为“纯真血统”的贝叶斯方法需要一直使用概率的和积准则。这个和积准则的实用牵涉到边缘概率的计算。而边缘概率的计算是使用贝叶斯方法进行模式识别的核心内容。

在曲线拟合问题中，给定了训练数据$\mathbf{x}, \mathbf{t}$，还有一个测试点x，我们的目标是估计t。因此，我们希望对$p(t|x, \mathbf{w}, \mathbf{t})$做一个评估。

贝叶斯估计求解$p(t|x, \mathbf{w}, \mathbf{t})$的过程应该是：

$$p(t|x, \mathbf{x}, \mathbf{t}) = \int p(t|x, \mathbf{w}) p(\mathbf{w}|\mathbf{x}, \mathbf{t}) \mathrm{d}\mathbf{w} \tag{4.1}$$

式 (4.1) 中$p(t|x, \mathbf{w})$由 (2.1)给出。此处，我们准备忽略α, β来简化符号表示。$p(\mathbf{w}|\mathbf{x}, \mathbf{t})$是参数$\mathbf{w}$的后验概率，可以对 (3.3)归一化获得。稍后我们会发现，对于曲线拟合问题，这个后验概率分布是高斯分布，进而式 (4.1)可以推演成：

$$p(t|x, \mathbf{x}, \mathbf{t}) = \mathcal{N}(t|m(x), s^2(x)) \tag{4.2}$$

其中均值和方差为：

$$m(x) = \beta \phi(x)^T \mathbf{S} \sum_{n=1}^{N} \phi(x_n) t_n \tag{4.3}$$

$$s^2(x) = \beta^{-1} + \phi(x)^T \mathbf{S} \phi(x) \tag{4.4}$$

矩阵$\mathbf{S}$为：

$$\mathbf{S}^{-1} = \alpha \mathbf{I} + \beta \sum_{n=1}^{N} \phi(x_n) \phi(x)^T \tag{4.5}$$

其中$\mathbf{I}$是单位阵。$\phi(x) = [\phi_0(x), \ldots, \phi_M(x)], \phi_i(x) = x^i$ 我们看到式 (4.3)所示的均值和方差依赖于x。方差的第一项β^{-1}代表了t的不确定度，这个不确定度是由噪声引起的。方差的第二项代表由$\mathbf{w}$带来的不确定度，这个不确定度是由贝叶斯方法带来的。