7.6 参数的有效个数

“参数个数”的概念可以推广，特别是推广到在拟合中使用了正则的模型中。假设我们将输出 $y_1, y_2, \ldots, y_N$ 放进向量 $\mathbf{y}$ 中，类似地对预测值进行同样操作得到 $\hat{\mathbf{y}}$。于是我们可以将线性拟合模型写成：

$$\hat{\mathbf{y}} = \mathbf{S}\mathbf{y} \tag{7.31}$$

其中 $\mathbf{S}$ 为依赖于输入向量 x_i 但不依赖于输出 y_i 的 $N \times N$ 阶矩阵。线性拟合方法包括在原始特征或在导出基的集合中运用的线性回归，以及采用平方收缩的光滑化方法，比如岭回归和三次光滑样条。则 **有效参数个数 (effective number of parameters)** 定义为：

$$\mathrm{df}(\mathbf{S}) = \mathrm{trace}(\mathbf{S}) \tag{7.32}$$

是 $\mathbf{S}$ 对角元之和（也被称作 **有效自由度 (effective degrees-of-freedom)**）。注意到如果 $\mathbf{S}$ 为投影到由 M 个特征张开的 **基础集 (basis set)** 上的正交投影矩阵，则 $\mathrm{trace}(\mathbf{S}) = M$。事实证明 $trace(\mathbf{S})$ 恰巧是 C_p 统计量（式 7.26）替换掉 d 作为参数个数的那个值。

如果 $\mathbf{y}$ 是从加性误差模型 $Y = f(X) + \epsilon$ 中产生的，$\mathrm{Var}(\epsilon) = \sigma_\epsilon^2$，则可以证明 $\sum_{i=1}^N \mathrm{Cov}(\hat{y}_i, y_i) = \mathrm{trace}(\mathbf{S})\sigma_\epsilon^2$，导出了更一般的定义

$$\mathrm{df}(\hat{\mathbf{y}}) = \frac{\sum_{i=1}^N \mathrm{Cov}(\hat{y}_i, y_i)}{\sigma_\epsilon^2} \tag{7.33}$$

（练习 7.4 和 7.5）第 5.4.1 节给出了在光滑样条情形下 $\mathrm{df} = \mathrm{trace}(\mathbf{S})$ 更直观的定义。

对于像神经网络的模型，我们用系数衰减（正则化）$\alpha \sum_m w_m^2$ 来最小化误差函数 $R(w)$，有效参数个数有如下形式：

$$\mathrm{df}(\alpha) = \sum_{i=1}^M \frac{\theta_m}{\theta_m + \alpha} \tag{7.34}$$

其中 θ_m 是 Hessian 矩阵 $\partial^2 R(w)/\partial w \partial w^T$ 的特征值。如果我们对解的误差函数做二次近似便可由式（7.32）导出式（7.34）（Bishop，1995[1]）。

[1] Bishop, C. (1995). Neural Networks for Pattern Recognition, Clarendon Press, Oxford.