

Chapter 3

Linear Models for Regression

- 首先 Linear Model 的含义是模型对于未知参数是线性的，而不要求对 input 是否线性，事实上为了通常 input 都是经过非线性处理的。未知参数的线性形式为分析提供了方便，当然同时也一定程度上限制了模型的能力。最简单的一个线性模型就是 Linear Regression，它对于 input 和未知参数都是线性的。

- Linear Model:**

给定一个含有 N 个数据的训练数据集 $\{\mathbf{x}_n\}$ (其中 $n = 1, \dots, N$)，同时给定了训练数据的标签值 $\{t_n\}$ 。

✓ **问题定义:**

$$y(\mathbf{x}, \mathbf{w}) = w_0 + \sum_{j=1}^{M-1} w_j \phi_j(\mathbf{x}) = \sum_{j=0}^{M-1} w_j \phi_j(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x})$$

其中 M 定义了模型的尺寸或者复杂度； $\phi_j(\mathbf{x})$ 称为 *basis functions*，一般是非线性的，可以看作是原始 input 的一种变换或者特征提取。

Linear Regression 中 basis function 选择的 $\phi_j(\mathbf{x}) = x_j$ ；Polynomial Regression 中选择的是 $\phi_j(\mathbf{x}) = x^j$ 。

还有一些常用的 basis functions 有：

$$\text{Gaussian Basis Function : } \phi_j(\mathbf{x}) = \exp\left\{-\frac{(x - \mu_j)^2}{2s^2}\right\}$$

$$\text{Sigmodial Basis Function : } \phi_j(\mathbf{x}) = \sigma\left(\frac{x - \mu_j}{s}\right)$$

此外，还有 Fourier basis、wavelets 正交基等。

✓ **解决方案:**

- Least Squares 最小二乘法：就是最小化 sum-of-squares error function (均方误差)，目标函数为：

$$E_D(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{t_n - \mathbf{w}^T \phi(\mathbf{x}_n)\}^2$$

- Maximum Likelihood Estimation (MLE) 最大似然估计：假设一个 Gaussian 噪声 ϵ ，即 $t = y(\mathbf{x}, \mathbf{w}) + \epsilon$ ，则在训练数据样本 i.i.d. 的前提下，对数似然为：

$$\ln p(\mathbf{t}|\mathbf{X}, \mathbf{w}, \beta) = \sum_{n=1}^N \ln \mathcal{N}(t_n | \mathbf{w}^T \phi(\mathbf{x}_n), \beta^{-1}) = \frac{N}{2} \ln \beta - \frac{N}{2} \ln(2\pi) - \beta E_D(\mathbf{w})$$

其中的 $E_D(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{t_n - \mathbf{w}^T \phi(\mathbf{x}_n)\}^2$ 就是最小二乘法的目标函数，二者最终殊途同归，即：**最小二乘法 \iff 高斯噪声假设下的 MLE。**

通过令梯度为 0 可以得出参数 \mathbf{w} 的解：

$$\mathbf{w}_{ML} = (\Phi^T \Phi)^{-1} \Phi^T \mathbf{t}$$

，其中 Φ 称之为 Design Matrix。

$$\Phi = \begin{pmatrix} \phi_0(\mathbf{x}_1) & \phi_1(\mathbf{x}_1) & \dots & \phi_{M-1}(\mathbf{x}_1) \\ \phi_0(\mathbf{x}_2) & \phi_1(\mathbf{x}_2) & \dots & \phi_{M-1}(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0(\mathbf{x}_N) & \phi_1(\mathbf{x}_N) & \dots & \phi_{M-1}(\mathbf{x}_N) \end{pmatrix}$$

✓ **Regularized Least Squares 正则化最小二乘法**：最小二乘法或者 MLE 的一个最大的问题是过拟合，正则化是通常的一个解决方案，用来抑制估计的参数过大产生的过拟合，即修改损失函数为：

$$E_D(\mathbf{w}) + \lambda E_W(\mathbf{w})$$

λ 和 $E_W(\mathbf{w})$ 分别就是正则化因子和正则项。关于正则项的选择，一个比较通用的正则项形式为：

$$\frac{1}{2} \sum_{n=1}^N \{t_n - \mathbf{w}^T \phi(\mathbf{x}_n)\}^2 + \frac{\lambda}{2} \sum_{j=1}^M |w_j|^q$$

其中当 $q = 2$ 时，就是最常用的一种 sum-of-squares 正则项，此时的解为：

$$\mathbf{w} = (\lambda \mathbf{I} + \Phi^T \Phi)^{-1} \Phi^T \mathbf{t}$$