Penalized Regression

Prof. Rozenn Dahyot
Trinity College Dublin, Ireland

2017

Consider the regression problem:

$$s(t) = \mu(t) + \epsilon(t) \quad \text{with} \quad \mu(t) = \sum_{k=0}^{K} \theta_k \, \phi_k(t) \tag{1}$$

with observations $\{(s^{(i)}, t^{(i)}\})_{i=1,\cdots,n}$ collected on the interval of time $[0,1]$ (i.e. $0 \leq t^{(i)} \leq 1$, $\forall i = 1,\cdots,n$). The (Fourier) basis of functions defined on $[0,1]$ is used: $\phi_0(t) = 1$, $\phi_1(t) = \cos(\omega t)$, $\phi_2(t) = \sin(\omega t)$, $\phi_3(t) = \cos(2\omega t)$, $\phi_4(t) = \sin(2\omega t)$, with $\omega = 2\pi$.

1. Show that[1]

$$\hat{\vec{\theta}} = (\Phi^T \Phi)^{-1} \Phi^T \vec{s} \tag{2}$$

minimises the Sum of Square Errors

$$SSE = \sum_{i=1}^{n} \left(s^{(i)} - \sum_{k=0}^{K} \theta_k \, \phi_k(t^{(i)}) \right)^2 = (\vec{s} - \Phi\vec{\theta})^T (\vec{s} - \Phi\vec{\theta})$$

Identify explicitly $\hat{\vec{\theta}}$, Φ and $\vec{s}$.

Ans.

$$SSE = (\vec{s} - \Phi\vec{\theta})^T (\vec{s} - \Phi\vec{\theta}) = \vec{s}^T \vec{s} - (\Phi\vec{\theta})^T \vec{s} - \vec{s}^T \Phi\vec{\theta} + (\Phi\vec{\theta})^T \Phi\vec{\theta}$$

Linear Algebra results: $(\Phi\vec{\theta})^T = \vec{\theta}^T \Phi^T$ so

$$SSE = \vec{s}^T \vec{s} - \vec{\theta}^T \Phi^T \vec{s} - \vec{s}^T \Phi\vec{\theta} + \vec{\theta}^T \Phi^T \Phi\vec{\theta}$$

and differentiation of SSE by $\vec{\theta}$ gives:

$$\frac{\partial SSE}{\partial \vec{\theta}} = \underbrace{\frac{\partial(\vec{s}^T \vec{s})}{\partial \vec{\theta}}}_{=0} - \underbrace{\frac{\partial(\vec{\theta}^T \Phi^T \vec{s})}{\partial \vec{\theta}}}_{=\Phi^T \vec{s}} - \underbrace{\frac{\partial(\vec{s}^T \Phi\vec{\theta})}{\partial \vec{\theta}}}_{=(\vec{s}^T \Phi)^T} + \underbrace{\frac{\partial(\vec{\theta}^T \Phi^T \Phi\vec{\theta})}{\partial \vec{\theta}}}_{=(\Phi^T\Phi + (\Phi^T\Phi)^T)\vec{\theta}}$$

or

$$\frac{\partial SSE}{\partial \vec{\theta}} = -2\Phi^T \vec{s} + 2(\Phi^T \Phi)\vec{\theta}$$

The Least Squares solution is such that $\frac{\partial SSE}{\partial \vec{\theta}} = 0$ so

$$\hat{\vec{\theta}} = (\Phi^T \Phi)^{-1} \Phi^T \vec{s}$$

[1] Rules to differentiate w.r.t. a vector https://en.wikipedia.org/wiki/Matrix_calculus#Vector-by-vector_identities

with $n \times 5$ matrix

$$\Phi = \begin{bmatrix} \phi_0(t^{(1)}) & \phi_1(t^{(1)}) & \cdots & \phi_4(t^{(1)}) \\ \vdots & & & \vdots \\ \phi_0(t^{(n)}) & \phi_1(t^{(n)}) & \cdots & \phi_4(t^{(n)}) \end{bmatrix}$$

with $n \times 1$ vector

$$\vec{s} = \begin{bmatrix} s^{(1)} \\ \vdots \\ s^{(n)} \end{bmatrix}$$

and 5×1 vector of parameters:

$$\vec{\theta} = \begin{bmatrix} \theta_0 \\ \vdots \\ \theta_4 \end{bmatrix}$$

---

2. The smoothness constraint is added to the cost function SSE as follow:

$$PENSSE = (\vec{s} - \Phi\vec{\theta})^T (\vec{s} - \Phi\vec{\theta}) + \gamma \int \left(\frac{D^2\mu(t)}{D^2 t} \right)^2 dt$$

   (a) Compute $\langle \phi_i | \phi_j \rangle = \int_0^1 \phi_i(t)\ \phi_j(t)\ dt, \forall i, j = 0, \cdots, 4.$

   (b) Compute $\left\langle \frac{D^2\phi_i}{Dt^2} | \frac{D^2\phi_j}{Dt^2} \right\rangle, \forall i, j = 0, \cdots, 4.$

   (c) Compute

$$\int_{0,1} \left(\frac{D^2\mu(t)}{D^2 t} \right)^2 dt$$

   (d) Show

$$PENSSE = (\vec{s} - \Phi\vec{\theta})^T (\vec{s} - \Phi\vec{\theta}) + \gamma\, \vec{\theta}^T \mathrm{R} \vec{\theta}$$

   Give an explicit expression of R.

   (e) Compute the solution $\vec{\theta}$ minimising PENSSE.

3. Consider the new penalty term

$$J_L = \int_0^1 [L\mu(t)]^2 dt$$

with operator L on function μ defined as $L\mu = \alpha\mu + D^2\mu$ (notation: $D^2\mu$ indicating $\frac{D^2\mu(t)}{D^2 t}$) and compute the solution $\vec{\theta}$ that minimises $SSE + \gamma J_L$.