

$$RSS(f, \lambda) = \sum_{i=1}^N (y_i - f(x_i))^2 + \lambda \int_{-\infty}^{\infty} (f''(t))^2 dt$$

We have:

$$f(x_i) = \sum_{j=1}^N \theta_j N_j(x) = N(x_i)^T \theta$$

$$N(x_i) = \begin{pmatrix} N_1(x_i) \\ \vdots \\ N_N(x_i) \end{pmatrix} \quad \theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_N \end{pmatrix}$$

$$N = \begin{pmatrix} N(x_1)^T \\ \vdots \\ N(x_N)^T \end{pmatrix}$$

$$\rightarrow \sum_{i=1}^N (y_i - f(x_i))^2 = \sum_{i=1}^N (y_i - N(x_i)^T \theta)^2 = (y - N\theta)^T (y - N\theta)$$

Further,

$$f(t) = \sum_{j=1}^N \theta_j N_j(t)$$

$$\rightarrow f'(t) = \sum_{j=1}^N \theta_j N_j'(t)$$

$$\rightarrow f''(t) = \sum_{j=1}^N \theta_j N_j''(t) = (N''(t))^T \theta, \quad N''(t) = \begin{pmatrix} N_1''(t) \\ \vdots \\ N_N''(t) \end{pmatrix}$$

This gives

$$\int_{-\infty}^{\infty} (f''(t))^2 dt = \int_{-\infty}^{\infty} (N''(t)^T \theta)^2 dt$$

$$(N''(t)^T \theta)^2 = \theta^T N''(t) (N''(t))^T \theta$$

$$N''(t) (N''(t))^T = \begin{pmatrix} N_1''(t) \\ \vdots \\ N_N''(t) \end{pmatrix} (N_1''(t) \dots N_N''(t))$$

$$= \begin{pmatrix} N_1''(t)^2 & N_1''(t) N_2''(t) & \dots & N_1''(t) N_N''(t) \\ N_1''(t) N_2''(t) & (N_2''(t))^2 & & \\ \vdots & & \ddots & \\ N_1''(t) N_N''(t) & & & (N_N''(t))^2 \end{pmatrix}$$

$$\int_{-\infty}^{\infty} (f''(t))^2 dt = \int_{-\infty}^{\infty} (\theta^T N''(t) \theta) dt = \theta^T \underbrace{\begin{pmatrix} \int_{-\infty}^{\infty} N_1''(t)^2 dt & \dots & \int_{-\infty}^{\infty} N_1''(t) N_2''(t) dt \\ \int_{-\infty}^{\infty} N_1''(t) N_2''(t) dt & \dots & \int_{-\infty}^{\infty} N_1''(t) N_N''(t) dt \\ \vdots & & \ddots & \int_{-\infty}^{\infty} N_N''(t)^2 dt \end{pmatrix}}_{= R} \theta$$

$$= \theta^T R \theta$$

$$\rightarrow RSS(f, \lambda) = (y - N\theta)^T (y - N\theta) + \lambda \theta^T R \theta$$

$$= y^T y - y^T N \theta - \underbrace{\theta^T N y}_{= y^T N \theta} + \theta^T N^T N \theta + \lambda \theta^T R \theta$$

$$= y^T y - 2 y^T N \theta + \theta^T N^T N \theta + \lambda \theta^T R \theta$$

Recall that $\frac{\partial a^T b}{\partial b} = a$ and $\frac{\partial b^T A b}{\partial b} = (A + A^T) b$

$$\rightarrow \frac{\partial y^T N \theta}{\partial \theta} = N^T y, \quad \frac{\partial \theta^T N^T N \theta}{\partial \theta} = (N^T N + \underbrace{(N^T N)^T}_{N^T N}) \theta = 2 N^T N \theta$$

$$\frac{\partial \theta^T R \theta}{\partial \theta} = (R + \underbrace{R^T}_{=R}) \theta = 2 R \theta$$

$$\rightarrow \frac{\partial RSS(f, \lambda)}{\partial \theta} = -2 N^T y + 2 N^T N \theta + 2 \lambda R \theta = 2 (-N^T y + (N^T N + \lambda R) \theta) = 0$$

$$\rightarrow (N^T N + \lambda R) \theta = N^T y$$

$$\rightarrow \theta = (N^T N + \lambda R)^{-1} N^T y$$

We have:

$$\ell(f, \lambda) = \sum_{i=1}^N \left(y_i \underbrace{\sum_{k=1}^K \theta_k N_k(x_i)}_{f(x_i)} - \log \left(1 + \exp \left(\sum_{k=1}^K \theta_k N_k(x_i) \right) \right) \right. \\ \left. - \frac{\lambda}{2} \int \sum_{j=1}^K \underbrace{(N_j''(t) \theta_j)^2}_{f''(t)} dt \right)$$

$$= \sum_{i=1}^N \left(y_i N(x_i)^T \theta - \log \left(1 + \exp \left(N(x_i)^T \theta \right) \right) - \frac{\lambda}{2} \theta^T \Omega \theta \right)$$

We get:

$$\frac{\partial \ell(f, \lambda)}{\partial \theta} = \sum_{i=1}^N \left(y_i N(x_i) - \frac{\exp(N(x_i)^T \theta)}{1 + \exp(N(x_i)^T \theta)} N(x_i) \right) - \frac{\lambda}{2} \Omega \theta \\ = \sum_{i=1}^N \left(y_i - p(x_i) \right) N(x_i) - \lambda \Omega \theta$$

$p(x_i) = P(Y_i = 1 | x_i)$

We have:

$$\frac{\partial p(x_i)}{\partial \theta} = \frac{\partial}{\partial \theta} \frac{e^{N(x_i)^T \theta}}{1 + e^{N(x_i)^T \theta}} = \frac{e^{N(x_i)^T \theta} \cdot N(x_i) (1 + e^{N(x_i)^T \theta}) - e^{N(x_i)^T \theta} \cdot e^{N(x_i)^T \theta}}{(1 + e^{N(x_i)^T \theta})^2} \\ = \frac{e^{N(x_i)^T \theta} N(x_i)}{(1 + e^{N(x_i)^T \theta})^2} \\ = \frac{e^{N(x_i)^T \theta}}{1 + e^{N(x_i)^T \theta}} \cdot \frac{1}{1 + e^{N(x_i)^T \theta}} N(x_i) \\ \underbrace{\hspace{1cm}}_{p(x_i)} \quad \underbrace{\hspace{1cm}}_{1-p(x_i)} \\ = p(x_i) (1 - p(x_i)) N(x_i)$$

$$\frac{\partial^2 \ell(f, \lambda)}{\partial \theta \partial \theta^T} = - \sum_{i=1}^N p(x_i) (1 - p(x_i)) N(x_i)^T - \lambda \Omega \\ = -N^T W N - \lambda \Omega$$

W is the diagonal matrix with diagonal entries $p(x_i)(1-p(x_i))$