Tutorial & Practical 7: Solutions

Question 1:

1. The linear spline model for f is given by

$$f(x_i) = \beta_0 + \beta_1 x_i + \sum_{k=1}^{K} u_k (x_i - k_k)_+$$

   $\boldsymbol{\beta}^T = [\beta_0 \ \beta_1]$ and $\mathbf{u}^T = [u_1, \ldots, u_k]$ define the coefficients of the polynomial functions and truncated functions respectively.

2. Define

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

$$\mathbf{Z} = \begin{pmatrix} (x_1 - k_1)_+ & \ldots & (x_1 - k_K)_+ \\ \vdots & \ddots & \vdots \\ (x_n - k_1)_+ & \ldots & (x_n - k_K)_+ \end{pmatrix}$$

   The penalized spline fitting criterion when divided by $\boldsymbol{\sigma}_\epsilon^2$ can be written as

$$\frac{1}{\boldsymbol{\sigma}_\epsilon^2} ||\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}||^2 + \frac{\lambda}{\sigma_\epsilon^2} ||\mathbf{u}||^2$$

3. Assuming $\text{cov}(\mathbf{u}) = \mathbf{G} = \boldsymbol{\sigma}_u^2 \mathbf{I}$ and $\sigma_u^2 = \sigma_\epsilon^2 / \lambda$ and $\text{cov}(\boldsymbol{\epsilon}) = \mathbf{R} = \sigma_\epsilon^2 \mathbf{I}$

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}) + \mathbf{u}^T \mathbf{G}^{-1} \mathbf{u}$$

   which corresponds to the negative log-likelihood of probability density of $(\mathbf{y}, \mathbf{u})$ with $\mathbf{y}|\mathbf{u} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \mathbf{R})$ and $\mathbf{u} \sim \mathcal{N}(0, \mathbf{G})$

4. $\mathbf{C} = [\mathbf{X} \ \mathbf{Z}]$, $\mathbf{D} = \text{diag}(0, 0, 1, 1, ..., 1)$, and $\alpha = \sigma_\epsilon^2 / \sigma_\mathbf{u}^2$,

$$\boldsymbol{\theta} = [\boldsymbol{\beta}^T \ \mathbf{u}^T]^T = (\mathbf{C}^T \mathbf{C} + \boldsymbol{\alpha} \mathbf{D})^{-1} \mathbf{C}^T \mathbf{y}$$

5. The fitted values can be written as

$$\tilde{\mathbf{f}} = \mathbf{C}(\mathbf{C}^T \mathbf{C} + \alpha \mathbf{D})^{-1} \mathbf{C}^T \mathbf{y}$$

6. $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}^*$, where $\boldsymbol{\epsilon}^* = \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}$

$$\text{cov}(\boldsymbol{\epsilon}^*) = \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R}$$

7. $\tilde{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}^T)^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}$, and $\text{cov}(\tilde{\boldsymbol{\beta}}) = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1}$

Question 2:

1. The Nadaraya-Watson kernel estimation at x_0 is given by

$$\hat{f}(x_0) = \frac{\sum_{i=1}^{N} K_h(x_0, x_i) y_i}{\sum_{i=1}^{N} K_h(x_0, x_i)} I\left[\sum_{i=1}^{N} K_h(x_0, x_i) \neq 0\right]$$

2. For a Gaussian kernel we have

$$K_h(x_0, x) = \frac{1}{\sqrt{2\pi} h} \exp\left\{-\frac{(x - x_0)^2}{2h^2}\right\}$$

Since $K_h(x_0, x) \neq 0$ for all x_0 and x_i we don't have a singularity in the denominator of $\hat{f}(x_0)$.
The Gaussian kernel is also everywhere differentiable in x_0 and therefore the Nadaraya-Watson estimator is differentiable as a function of x_0.

3. For the Epanechnikov Kernel we have

$$K_h(x_0, x) = K\left(\frac{|x - x_0|}{h}\right)$$

$$K(z) = \begin{cases} \frac{3}{4}(1 - z^2) & \text{if } |z| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

We can observe from the expression that for $|x - x_0| \to h^-$, $\frac{|x-x_0|}{h} \to 1^-$

$$\frac{\partial K_h(x_0, x)}{\partial x_0} = \frac{\partial K(z)}{\partial z} . \frac{\partial z}{\partial x_0} = \left(\frac{-3}{2} z\right)\left(\frac{-1}{h}\right) = \frac{3z}{2h} = \frac{3|x - x_0|}{2h^2} \neq 0$$

However, when $|x - x_0| \to h^+$, $\frac{|x-x_0|}{h} \to 1^+$ and

$$\frac{\partial K_h(x_0, x)}{\partial x_0} = 0$$

because the function is zero in this domain. Since the two limits are different this kernel is not differentiable everywhere.

Question 3:

1. N plays the role of the smoothing parameter similar to h for the kernel smoothing.

2.
$$\hat{f}_{nN}(x) = \sum_{j=1}^{N} \hat{\theta}_j \rho_j(x) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{N} y_i \rho_j(x_i) \rho_j(x) = \sum_{i=1}^{n} y_i W_{ni}(x)$$

where $W_{ni}(x) = \frac{1}{n} \sum_{j=1}^{N} \rho_j(x_i) \rho_j(x)$ and $\hat{f}_{nN}$ is therefore a linear estimator.

3.
$$\hat{\theta}_j = \frac{1}{n}\sum_{i=1}^{n} y_i \rho_j(x_i) = \frac{1}{n}\left(\sum_{i=1}^{n} f(x_i)\rho_j(x_i) + \sum_{i=1}^{n} \epsilon_i \rho_j(x_i)\right)$$

$$\mathbb{E}(\hat{\theta}_j) = \frac{1}{n}\sum_{i=1}^{n} f(x_i)\rho_j(x_i) = \theta_j + r_j$$

$$r_j = \frac{1}{n}\sum_{i=1}^{n} f(x_i)\rho_j(x_i) - \int_0^1 f(x)\rho_j(x)dx = \frac{1}{n}\sum_{i=1}^{n} f(x_i)\rho_j(\mathbf{x}_i) - \theta_j$$

4.
$$\mathbb{E}\left[\left(\hat{\theta}_j - \theta_j\right)^2\right] = \left(E(\hat{\theta}_j) - \theta_j\right)^2 + \mathbb{E}\left[\left(\hat{\theta}_j - \mathbb{E}(\hat{\theta}_j)\right)^2\right]$$
$$= r_j^2 + \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} \epsilon_i \rho_j(x_i)\right)^2\right]$$
$$= r_j^2 + \frac{1}{n^2}\sum_{i=1}^{n} \sigma_\epsilon^2 \rho_j^2(x) = r_j^2 + \frac{\sigma_\epsilon^2}{n}$$

5.
$$f(x) = \sum_{j=1}^{\infty} \theta_j \rho_j(x)$$

$$\hat{f}_{nN}(x) = \sum_{j=1}^{N} \hat{\theta}_j \rho_j(x)$$

$$\left\|\hat{f}_{nN} - f\right\|_2^2 = \sum_{j=1}^{N} (\hat{\theta}_j - \theta_j)^2 + \sum_{j=N+1}^{\infty} \theta_j^2 \quad \text{since}$$

$$\hat{f}_{nN} - f = \sum_{j=1}^{N} (\hat{\theta}_j - \theta_j)\rho_j(x) + \sum_{j=N+1}^{\infty} \theta_j \rho_j(x)$$

and $\{\rho_j\}$ is an orthonormal basis

$$\mathbb{E}\|\hat{f}_{nN} - f\|_2^2 = \sum_{j=1}^{N} \mathbb{E}\left[(\hat{\theta}_j - \theta_j)^2\right] + \sum_{j=N+1}^{\infty} \theta_j^2$$
$$= \frac{N\sigma_\epsilon^2}{n} + \sum_{j=1}^{N} r_j^2 + \sum_{j=N+1}^{\infty} \theta_j^2 = C$$

6. As shown in the figure

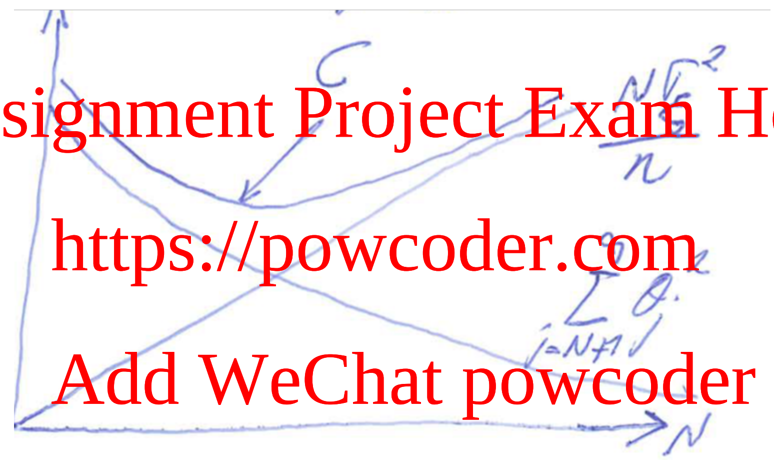

Figure 1: Solution to 3.6