Problem 3

Consider the truncated power series representation for cubic splines with K interior knots

$$f(x) = \sum_{j=0}^{3} \beta_j X^j + \sum_{k=1}^{K} \theta_k (x - \varepsilon_k)^3_+$$

Prove that the natural boundary conditions for natural cubic spline imply the following linear constraints on the coefficients:

$$\beta_2 = 0, \quad \sum_{k=1}^{K} \theta_k = 0$$

$$\beta_3 = 0, \quad \sum_{k=1}^{K} \varepsilon_k \theta_k = 0$$

As per the above given constraints, we get the following eqn.
β_2 & β_3

$$f(x) = \beta_0 X^0 + \beta_1 x^1 + \cancel{\beta_2 x^2}^{\,0} + \cancel{\beta_3 x^3}^{\,0} + \sum_{k=1}^{K} \theta_k (x - \varepsilon_k)^3_+$$

$$= \beta_0 + \beta_1 x^1 + \sum_{k=1}^{K} \theta_k (x - \varepsilon_k)^3_+$$

Here $\underbrace{\beta_0 \cdot 1}_{N_1(X) = 1}$ and $\underbrace{\beta_1 x^1}_{N_2(x) = X}$ [Given in the question]

we here constructed a new basis with the first two basis function.

Let us now consider the θ constraints, let the write down that

$$\sum_{k=1}^{K-2} \theta_k = -\theta_{K-1} - \theta_K \quad \text{and} \quad \sum_{k=1}^{K-2} \varepsilon_k \theta_k = -\varepsilon_{K-1}\theta_{K-1} - \varepsilon_K \theta_K$$

Now consider the truncated function and use the last two terms like done above

i.e consider $\sum_{k=1}^{K} \theta_k (x - \varepsilon_k)^3_+ = \sum_{k=1}^{K-2} \theta_k (x - \varepsilon_k)^3_+ +$

$$\underbrace{\theta_{K-1} (x - \varepsilon_{K-1})^3_+}_{(i)} + \underbrace{\theta_K (x - \varepsilon_K)^3_+}_{(ii)} \rightarrow (1)$$

To the above equation, apply the θ constraints to show last two terms of the above equation can be written as the sum of the N-2 first terms we should now consider the last two terms for the computation from the equation 1

Let us now consider the term (i) from equation 1

$$\theta_{k-1}(x-\varepsilon_{k-1})_+^3 = \frac{(x-\varepsilon_{k-1})_+^3}{(\varepsilon_k-\varepsilon_{k-1})}\left(\theta_{k-1}\varepsilon_k-\theta_{k-1}\varepsilon_{k-1}\right)$$

we got the above equation by multiplying and dividing $(\varepsilon_k-\varepsilon_{k-1})$ on the LHS..

$$= \frac{(x-\varepsilon_{k-1})_+^3}{(\varepsilon_k-\varepsilon_{k-1})}\left(\theta_{k-1}\varepsilon_k-\theta_{k-1}\varepsilon_{k-1}\right)$$

$$= \frac{(x-\varepsilon_{k-1})_+^3}{(\varepsilon_k-\varepsilon_{k-1})}\left(\theta_{k-1}\varepsilon_k-\theta_{k-1}\varepsilon_{k-1}+\theta_k\varepsilon_k-\theta_k\varepsilon_k\right)$$

(Add and subract $\theta_k\varepsilon_k$)

$$= \frac{(x-\varepsilon_{k-1})_+^3}{(\varepsilon_k-\varepsilon_{k-1})}\left(\varepsilon_k(\theta_{k-1}+\theta_k)-\varepsilon_{k-1}\theta_{k-1}-\theta_k\varepsilon_k\right)$$

$$= \frac{(x-\varepsilon_{k-1})_+^3}{(\varepsilon_k-\varepsilon_{k-1})}\left(-\varepsilon_K\sum_{k=1}^{K-2}\theta_k+\sum_{k=1}^{K-2}\theta_k\varepsilon_k\right)$$

$$= -\frac{(x-\varepsilon_{k-1})_+^3}{(\varepsilon_k-\varepsilon_{k-1})}\sum_{k=1}^{K-2}\theta_k\,(\underset{\text{big } K}{\underset{\downarrow}{\varepsilon_K}}-\underset{\text{small } k}{\underset{\downarrow}{\varepsilon_k}})$$

$$= -\sum_{k=1}^{K-2}\theta_k(\varepsilon_K-\varepsilon_k)\left(\frac{x-\varepsilon_{K-1}}{\varepsilon_K-\varepsilon_{K-1}}\right)_+^3 \rightarrow \text{(A)}$$

Let us now consider the term (ii) from equation 1

$$\theta_k(x-\varepsilon_k)_+^3 = \frac{(x-\varepsilon_K)_+}{(\varepsilon_k-\varepsilon_{k-1})}\left(\theta_k\varepsilon_k-\theta_k\varepsilon_{k-1}\right)$$

∵ Multiply and didide by $(\varepsilon_k-\varepsilon_{k-1})$

$$= \frac{(x-\varepsilon_k)_+^3}{(\varepsilon_k - \varepsilon_{k-1})} \left(\theta_k \varepsilon_k - \theta_k \varepsilon_{k-1} + \theta_{k-1}\varepsilon_{k-1} - \theta_{k-1}\varepsilon_{k-1} \right)$$

Add and subtract $(\theta_{k-1}\varepsilon_{k-1})$

$$= \frac{(x-\varepsilon_k)_+^3}{(\varepsilon_k - \varepsilon_{k-1})} \left(-\varepsilon_{k-1}(\theta_{k-1} + \theta_k) + \varepsilon_{k-1}\theta_{k-1} + \varepsilon_k \theta_k \right)$$

$$= \frac{(x-\varepsilon_k)_+^3}{(\varepsilon_k - \varepsilon_{k-1})} \left(\varepsilon_{k-1} \sum_{k=1}^{k-2} \theta_k - \sum_{k=1}^{k-2} \theta_k \varepsilon_k \right)$$

$$= (x-\varepsilon_k)_+^3 \sum_{k=1}^{k-2} \theta_k \frac{(\varepsilon_{k-1} - \varepsilon_k)}{(\varepsilon_k - \varepsilon_{k-1})}$$

Multiple and divide by $(\varepsilon_k - \varepsilon_k)$

$$= (x-\varepsilon_k)_+^3 \sum_{k=1}^{k-2} \theta_k \frac{(\varepsilon_{k-1} - \varepsilon_k)}{(\varepsilon_k - \varepsilon_{k-1})} \times \frac{(\varepsilon_k - \varepsilon_k)}{(\varepsilon_k - \varepsilon_k)}$$

Add and Subtract ε_k from the above equation and rearrange terms

$$= (x-\varepsilon_k)_+^3 \sum_{k=1}^{k-2} \theta_k \underbrace{\frac{(\varepsilon_k - \varepsilon_k)(\varepsilon_{k-1} - \varepsilon_k + \varepsilon_k - \varepsilon_k)}{(\varepsilon_k - \varepsilon_k)(\varepsilon_k - \varepsilon_{k-1})}}$$

Representing the above equation in a different form

$$= (x-\varepsilon_k)_+^3 \sum_{k=1}^{k-2} \theta_k (\varepsilon_k - \varepsilon_k) \left[\frac{1}{(\varepsilon_k - \varepsilon_{k-1})} - \frac{1}{(\varepsilon_k - \varepsilon_k)} \right] \longrightarrow \text{(B)}$$

Taking LCM and solving will give the above equation representation

We have equation (1)

$$\sum_{k=1}^{k} \theta_k (x-\varepsilon_k)_+^3 = \sum_{k=1}^{k-2} \theta_k (x-\varepsilon_k)_+^3 + \theta_{k-1}(x-\varepsilon_{k-1})_+^3 + \theta_k (x-\varepsilon_k)_+^3$$

Substitute the values (A) and (B) in (1)

$$= \sum_{k=1}^{K-2} \theta_k (x-\varepsilon_k)_+^3 - \sum_{k=1}^{K-2} \theta_k (\theta_k - \varepsilon_k) \frac{(x-\varepsilon_{K-1})_+^3}{(\varepsilon_K - \varepsilon_{K-1})}$$

$$+ (x-\varepsilon_K)_+^3 \sum_{k=1}^{K-2} \theta_k (\varepsilon_K - \varepsilon_k) \left(\frac{1}{\varepsilon_K - \varepsilon_{K-1}} - \frac{1}{\varepsilon_K - \varepsilon_k} \right)$$

multiply and divide by $(\varepsilon_K - \varepsilon_k)$ from the first term too.

$$= \sum_{k=1}^{K-2} \theta_k (\varepsilon_K - \varepsilon_k) \frac{(x-\varepsilon_k)_+^3}{(\varepsilon_K - \varepsilon_k)} - \sum_{k=1}^{K-2} \theta_k (\varepsilon_K - \varepsilon_k) \frac{(x-\varepsilon_{K-1})_+^3}{(\varepsilon_K - \varepsilon_{K-1})}$$

$$+ \sum_{k=1}^{K-2} \theta_k (\varepsilon_K - \varepsilon_k) \left(\frac{(x-\varepsilon_K)_+^3}{\varepsilon_K - \varepsilon_{K-1}} - \frac{(x-\varepsilon_K)_+^3}{\varepsilon_K - \varepsilon_k} \right)$$

$$= \sum_{k=1}^{K-2} \theta_k (\varepsilon_K - \varepsilon_k) \left(\frac{(x-\varepsilon_k)_+^3}{\varepsilon_K - \varepsilon_k} - \frac{(x-\varepsilon_{K-1})_+^3}{\varepsilon_K - \varepsilon_{K-1}} + \frac{(x-\varepsilon_K)_+^3}{\varepsilon_K - \varepsilon_{K-1}} - \frac{(x-\varepsilon_K)_+^3}{\varepsilon_K - \varepsilon_k} \right)$$

$$= \sum_{k=1}^{K-2} \theta_k (\varepsilon_K - \varepsilon_k) \left(\frac{(x-\varepsilon_k)_+^3 - (x-\varepsilon_K)_+^3}{\varepsilon_K - \varepsilon_k} - \frac{(x-\varepsilon_{K-1})_+^3 - (x-\varepsilon_K)_+^3}{\varepsilon_K - \varepsilon_{K-1}} \right)$$

$$\Rightarrow \sum_{k=1}^{K} \theta_k (x-\varepsilon_k)_+^3 = \sum_{k=1}^{K-2} \theta_k (\varepsilon_K - \varepsilon_k) (d_k(x) - d_{K-1}(x))$$

where, $d_k(x) = \dfrac{(x-\varepsilon_k)_+^3 - (x-\varepsilon_K)_+^3}{\varepsilon_K - \varepsilon_k}$

$$d_{K-1}(x) = \frac{(x-\varepsilon_{K-1})_+^3 - (x-\varepsilon_K)_+^3}{\varepsilon_K - \varepsilon_{K-1}}$$

Hence, the value $d_k(x)$ is as below i.e

$$\boxed{d_k(x) = \frac{(x-\varepsilon_k)_+^3 - (x-\varepsilon_K)_+^3}{\varepsilon_K - \varepsilon_k}}$$

Problem 1

Principle Component Analysis

Show first principle component minimizes the residual sum of Squares

We consider the below case of a 1D projection of the datapoints. The p dimensional vectors are projected on a line through the origin.

→ Let the line be represented by vector $\vec{w}$ = unit vector

→ Let the project of data points be represented by vector $\vec{x}_i$

→ This projection on the line will give $\vec{w}\vec{x}_i$ and this is the distance of that data point from the origin.

→ The coordinate in p dimensional space is $(\vec{x}_i \cdot \vec{w})\vec{w}$

We know that the mean of projected datatypes is 0 ~~bease~~ mean of x_i is 0

ie $\sum_{i=1}^{n} \vec{x}_i = 0$

when number of observations are n, we have below equation

$$\Rightarrow \frac{1}{n} \sum_{i=1}^{n} (\vec{x}_i \cdot \vec{w})\vec{w}$$

The residual of the projection is given by $\|\vec{x}_i - (\vec{w} \cdot \vec{x}_i)\vec{w}\|^2$

$$\|\vec{x}_i - (\vec{w} \cdot \vec{x}_i)\vec{w}\|^2 = \|\vec{x}_i\|^2 - 2(\vec{w} \cdot \vec{x}_i)(\vec{w} \cdot \vec{x}_i) + \|\vec{w}\|^2$$

$$= \|\vec{x}_i\|^2 - 2(\vec{w} \cdot \vec{x}_i)^2 + 1 \text{ (because w is a unit vector)}$$

We will add the residuals of all the vectors

$$\therefore RSS(\vec{w}) = \sum_{i=1}^{n} \|\vec{x}_i\|^2 - \sum_{i=1}^{n} 2(\vec{w}\vec{x}_i)^2$$

In the above equation the first term does not depend on $\vec{w}$ and hence does not contribute to minimize the RSS.

To make RSS small we should maximize the term obtained

$$\sum_{i=1}^{n} (\vec{w} \cdot \vec{x}_i)^2$$

In this case n is not depending on $\vec{w}$, so we should maximize

$$\frac{1}{n} \sum_{i=1}^{n} (\vec{w} \vec{x}_i)^2$$

i.e mean of $(\vec{w} \vec{x}_i)^2$

We know that the mean of a square is always equal to the square of the mean added to some variance.

i.e $$\frac{1}{n} \sum_{i=1}^{n} (\vec{w} \cdot \vec{x}_i)^2 = \left(\frac{1}{n} \vec{x}_i \vec{w}\right)^2 + \mathrm{Var}[\vec{w} \cdot \vec{x}_i]$$

We initially saw that the mean of the projections is always zero. So minizing RSS will be equal to maximizing the variance of the projections

$$\frac{1}{n} \sum_{i=1}^{n} (\vec{w} \cdot \vec{x}_i)^2 = \mathrm{Var}[\vec{w} \cdot \vec{x}_i]$$

$$\therefore \boxed{\sigma_{\vec{w}}^2 = \frac{1}{n} \sum_{i=1}^{n} (\vec{w}_i \cdot \vec{x}_i)^2}$$