HW 5

Problem 1.

Answer:

One way to model categorical outcomes is to use the locally constant multinomial logit model, which involves calculating the probability of each class given a set of covariates. The model is defined by the equation:

$p_{k(}x) = \exp(f_k(x)) / 1 + \exp(f_1(x)) + ... + \exp(f_k(x))$ ---- (1)

where $f_k(x)$ represents the log odds of the kth class and K is the number of classes.

To demonstrate that this approach amounts to smoothing the binary response indicators for each class separately using a Nadaraya-Watson kernel smoother, we must first examine the relationship between the log odds and the binary response indicators. Specifically, we can express the binary response indicators for the kth class as y_i, k=1 if y_i=k and 0 otherwise.

By taking log on both sides of equation 1,

$\log(p_{k(}x)) = f_k(x) - \log(1 + \exp(f_1(x)) + ... + \exp(f_k(x)))$

rewriting the above as

$\log(p_{k(}x)) = f_k(x) – \log(\sum_{i=1}^{k} \exp\,(fi(x)\,))$

We can express fk(x) as a linear combination of binary response indicators,

$fk(x) = \sum_{i=1}^{n} K\lambda(x, xi)yi, k$

This smoothing approach is achieved using the Nadaraya-Watson kernel smoother. Therefore, the locally constant multinomial logit model can be viewed as a smoothing process of the binary response indicators for each class independently using the Nadaraya-Watson kernel smoother.

Problem 2:

Answer:

Local polynomial regression involves fitting a polynomial function to a subset of nearby data points, weighted by kernel functions. The leave-one-out cross-validation (LOOCV) is a technique used to evaluate the performance of the model by leaving out one data point at a time and fitting the model on the remaining data points.

the predicted value of y_i at x_i, $f'(x_i)$, can be written as a linear combination of the response values, yi, of the data points within the bandwidth of x_i:

$f'(x_i) = \sum_{j=1}^{n} K_\lambda(x_i-x_j) * \beta j$

where $K_\lambda(\ )$ is the kernel function, λ is the bandwidth parameter, n is the total number of data points, and β_j is the coefficient of the j^{th} data point.

Using this equation, we can derive the LOOCV residual as:

$$e_i(i) = y_i - f'(x_i; -i)$$

$$= y_i - [\sum_{j\neq i} K_\lambda(x_i-x_j) * \beta_j / \sum_{j\neq i} K_\lambda(x_i-x_j)]$$

where the summation excludes the i^{th} data point.

Substituting the above equation for $f'(x_i\,; -i)$, we can expand the LOOCV RSS as:

$$LOOCV\_RSS = \sum_{i=1}^{n} e_i(i)^2$$

$$\Rightarrow \sum_{i=1}^{n} [y_i - [\sum_{j\neq i} K_\lambda(x_i-x_j) * \beta_j / \sum_{j\neq i} K_\lambda(x_i-x_j)]^2$$

$$\Rightarrow \sum_{i=1}^{n} [y_i - [\sum_{j\neq i} K_\lambda(x_i-x_j) * \beta_j / (n-1)\ h(x_i)]^2$$

where $h(x_i)$ is the effective bandwidth at x_i, which can be defined as:

$$h(x_i) = \sum_{j\neq i} K_\lambda(x_i-x_j)/(n-1)$$

Now, we can obtain the coefficients βj using a least-squares method:

$$[\sum_{j\neq i} K_\lambda(x_i-x_j) * x_j^k / (n-1) * h(x_i)] * \beta_k = [\sum_{j\neq i} K_\lambda(x_i-x_j) * y_i / (n-1) * h(x_i)]$$

where k ranges from 0 to m, and m is the degree of the polynomial.

Solving this equation system for βj, we can obtain the expression for LOOCV RSS as:

LOOCV_RSS = $\sum_{i=1}^{n} [y_i - f'(x_i)]^2 / [1 - h(x_i) / n]^2$

where $f'(x_i)$ is the predicted value of y_i at x_i obtained from the local polynomial regression model.

This expression can be used to select the optimal bandwidth and degree of the polynomial that minimize the LOOCV RSS, which can improve the performance of the local polynomial regression model.

Problem 3:

Answer:

Problem 3:

Answer:

let $\hat{f}(x_j) = \hat{y}_j$ then

$$\text{Opt} = \text{Err}_{in} - \mathbb{E}_y \text{err}.$$

$$= n^{-1}\sum_{j=1}^{n} \mathbb{E}_{y^N}\mathbb{E}_y\left(y_j^N - \hat{f}(x_j)\right)^2 - \mathbb{E}_y n^{-1}\sum_{j=1}^{n}\left(y_j - \hat{f}(x_j)\right)^2$$

$$= n^{-1}\sum_{j=1}^{n} \mathbb{E}_{y^N}\mathbb{E}_y\left(y_j^N - \hat{y}_j\right)^2 - \mathbb{E}_y\left(y_j - \hat{y}_j\right)^2$$

$$= n^{-1}\sum_{j=1}^{n} \mathbb{E}_{y^N}\mathbb{E}_y\left(y_j^{N2} - 2y_j^N\hat{y}_j - \hat{y}_j^2\right) - \mathbb{E}_y\left(y_j^2 - 2y_j\hat{y}_j + \hat{y}_j^2\right)$$

$$= n^{-1}\sum_{j=1}^{n} \mathbb{E}_{y^N} y_j^{2N} - 2\mathbb{E}_{y^N}\mathbb{E}_y\left(y_j^N\cdot\hat{y}_j\right) - \mathbb{E}_y y_j^2 + 2\mathbb{E}_y\left(y_j\hat{y}_j\right) - \mathbb{E}_y\hat{y}_j^2$$

$$= n^{-1}\sum_{j=1}^{n} \mathbb{E}_{y^N} y_j^{2N} - 2\mathbb{E}_{y^N}\left(y_j^N\hat{y}_j\right) - \mathbb{E}_y y_j^2 + 2\mathbb{E}_y\left(y_j\hat{y}_j\right)$$

$$= n^{-1}\sum_{j=1}^{n} 2\mathbb{E}_{y^N}\left(y_j\hat{y}_j\right) - 2\mathbb{E}_{y^N}\left(y_j^N\right)\cdot\mathbb{E}_y\left(\hat{y}_j\right)$$

$$= n^{-1}\sum_{j=1}^{n} 2\mathbb{E}_{y^N}\left(y_j\cdot\hat{y}_j\right) - 2\mathbb{E}_y\left(y_j\right)\cdot\mathbb{E}_y\left(\hat{y}_j\right)$$

$$\boxed{\text{Opt} = \frac{2}{n}\sum_{j=1}^{n}\text{Cov}\left(y_j, \hat{y}_j\right)}$$

Problem 4:

Answer:

Let s'_i be the i^{th} row of smoothing matrix S

$$Cov(y_i, y'_i) = Cov(y_i, s'_i * y)$$

$$Cov(y_i, y'_i) = Cov(y_i, \Sigma s_{ij}\, y_j)$$

$$Cov(y_i, y'_i) = s_{ii}\, Cov(y_i , y_i)$$

$$Cov(y_i, y'_i) = s_{ii}\, \sigma^2$$