1. Identify the smoother matrix, \( A \), of the Nadaraya-Watson estimator. Show that we have
\[
\text{tr}(A) \approx \frac{K(0)(b-a)}{h^2}
\]
where \( a \) and \( b \) are the end points of the support, \( [a, b] \), of the design time points \( X_i, i = 1, 2, \ldots, n \), \( K \) and \( h \) are the associated kernel and bandwidth used.
Assume that the degrees of freedom, \( \text{df} = \text{tr}(A) \), of the Nadaraya-Watson estimator has
a range \([\text{df}_{\text{min}}, \text{df}_{\text{max}}]\). Find the associated approximate range for the bandwidth \( h \). This range can be used for bandwidth selection.

2. Do a similar problem as Problem 1, but now for the local linear estimator. You may consult the asymptotic result of Problem 3 of Tutorial 7. Assume that the local linear smoother matrix \( A \) is a symmetric matrix (This is generally not true for the local linear smoother matrix). Then it has a singular value decomposition: \( A = UDU^T \) where \( U \) is an orthonormal matrix, containing all the eigenvectors of \( A \) while \( D = \text{diag}(d_1, d_2, \ldots, d_n) \) is a diagonal matrix containing all the eigenvalues of \( A \). Show that \( \sum_{r=1}^{n} d_r \approx \frac{K(0)(b-a)}{h^2} \) and \( \text{tr}(A^l) = \sum_{r=1}^{n} d_r^l \) for \( l = 1, 2, \ldots \).

3. Recall that the local linear estimator \( \hat{m}(x) \) can be expressed as
\[
\hat{m}(x) = \sum_{j=1}^{n} w_j(x)Y_j
\]
where \( w_j(x) = \frac{s_{nl}(x)}{s_{nl}(x) - s_{nl}(x)^2} \) with \( s_{nl}(x) = \sum_{j=1}^{n} K_h(X_j - x)(X_j - x)^l \) for \( l = 0, 1, \ldots \). Denote \( s_{nl}^{(-i)}(x) \), \( w_j^{(-i)}(x) \) and \( \hat{m}^{(-i)}(x) \) be the associated \( s_{nl}(x), w_j(x) \) and \( \hat{m}(x) \) but using all the design time points except \( X_i \).

(a) Show that \( s_{nl}^{(-i)}(X_i) = s_{nl}(X_i) - K(0)/h \) and \( s_{nl}^{(-i)}(X_i) = s_{nl}(X_i) \) when \( l > 0 \).
(b) Show that \( w_j^{(-i)}(X_i) = w_j(X_i)/(1 - w_i(X_i)) \) for \( j = 1, 2, \ldots, n \).
(c) Show that \( Y_i - \hat{m}^{(-i)}(X_i) = (Y_i - \hat{m}(X_i))/(1 - w_i(X_i)) \) so that \( CV(h) = n^{-1} \sum_{i=1}^{n} \left[ Y_i - \hat{m}^{(-i)}(X_i) \right]^2 = n^{-1} \sum_{i=1}^{n} \left[ \frac{Y_i - \hat{m}(X_i)}{1 - w_i(X_i)} \right]^2 \).

4. Explain why in the following situations, a large bandwidth is preferred.

(a) The noise variances are large.
(b) The regression function \( m(x) \) is smooth.
(c) The sample size is small.