Question 1.

Given $\{(x_1, y_1), \ldots, (x_n, y_n)\}$

1):

Let $\underset{n\times 3}{A} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}$ $\quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$

$\underset{n\times n}{W} = \begin{bmatrix} w_1(x) & & 0 \\ & \ddots & \\ 0 & & w_n(x) \end{bmatrix}$ $\quad w_i(x) = K\left(\frac{x_i - x}{h}\right)$

$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$

$$J(\beta_1, \beta_2, \beta_3) = \sum_{i=1}^{n} w_i(x)\,(y_i - \beta_1 - \beta_2 x_i - \beta_3 x_i^2)^2$$

$$= \underbrace{\begin{bmatrix} \ldots, y_i - \beta_1 - \beta_2 x_i - \beta_3 x_i^2, \ldots \end{bmatrix}_{i=1}^{i=n}}_{1\times n} \underbrace{\begin{bmatrix} w_1(x) & & \\ & \ddots & \\ & & w_n(x) \end{bmatrix}}_{n\times n} \underbrace{\begin{bmatrix} \vdots \\ y_i - \beta_1 - \beta_2 x_i - \beta_3 x_i^2 \\ \vdots \end{bmatrix}_{i=1}^{i=n}}_{n\times 1}$$

In the matrix form:

$$J(\beta) = (y - A\beta)^T W (y - A\beta)$$

2):
$$\frac{\partial J(\beta)}{\partial \beta} = \frac{\partial \{ y^T W y - 2y^T W A\beta + \beta^T A^T W A \beta \}}{\partial \beta}$$

$$= -2(y^T W A)^T + 2A^T W A \beta$$

$$= -2A^T W y + 2A^T W A \beta$$

Taking $\frac{\partial J(\beta)}{\partial \beta} = 0$

$$\Rightarrow \quad A^T W y = (A^T W A)\beta$$

$$\hat{\beta} = (A^T W A)^{-1} A^T W y$$

(require $(A^T W A)^{-1}$ exists.

3): • Non-parametric, since it performs locally second order polynomial fit, therefore the number of parameter scale with data. The solution of $\hat{\beta}$ is depend on $W_i(x)$. Thus, we are fitting the parameter to every point x of intrested. - therefore the number of parameters $\geq n$.

advantage	disadvantage.
No statistical assumption of the underlying distribution or regression function	Computationally expensive and require large size of training set.