Curve fitting by linear smoothing

A) Suppose that we have $y_i = f(x_i) + \varepsilon_i$, and that we have a least-squares estimator $\hat{y}^* := \hat{\beta} x^*$ for a new point x^*. We know, then, that $\hat{\beta} = r \cdot S_y / S_x$, where r is the correlation between the predictors and the responses, S_y is the standard deviation of the responses, and S_x is that of the predictors. Let us rewrite this as follows, keeping in mind that we have assumed that we have subtracted the means already:

$$
\begin{aligned}
\hat{\beta} &= r \frac{S_y}{S_x} \\
&= \left(\frac{1}{n-1} \sum_{i=1}^{n} \frac{x_i y_i}{S_x S_y} \right) \frac{S_y}{S_x} \\
&= \left(\sum_{i=1}^{n} \frac{1}{n-1} \frac{x_i}{S_x^2} \right) y_i \\
&= \left(\sum_{i=1}^{n} \frac{x_i}{\sum_{i=1}^{n} x_i^2} \right) y_i
\end{aligned}
$$

Therefore, if we set

$$w(x_i, x^*) := \frac{x_i x^*}{\sum_{i=1}^{n} x_i^2},$$

we get

$$\hat{y}^* = \hat{\beta} x^* = \sum_{i=1}^{n} w(x_i, x^*) y_i .$$

This smoother determines the value of $\hat{y}^*$ with a weighted sum of the observed values y_i, where the weight of y_i is the associated predictor value scaled by the sum of squares of the predictors and multiplied by the new value x^*. As a result, this smoother is influenced by all observed points (x_i, y_i) for which $x_i \neq 0$, but is most influenced by those points for which the product $x_i y_i$ is large. In particular, this influence is the same regardless of where x^* is relative to the observed predictor values.

On the other hand, the smoother determined by the weight function

$$w_K(x_i, x^*) := \begin{cases} 1/K & \text{if } x_i \text{ is one of the } K \text{ closest sample points to } x^*, \text{ and} \\ 0 & \text{otherwise} \end{cases}$$

gives equal weight to the points whose predictor values are closest to x^*, and no weight to the others.