Calculating Variance ($Var(\hat{f}(x_0))$) for locally weighted regression.

$$y = BW\beta + \varepsilon$$

Where B, W defined on P195, $\beta \in \mathbb{R}^K$, $\varepsilon \sim N(0, \sigma^2)$

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

Now, $\mathbb{E}(\hat{y}(x_0)) = \mathbb{E}(b^T(x_0)\hat{\beta})$

$$= \mathbb{E}[b^T(x_0)(B^T W B)^{-1} B^T W y]$$

$$= \mathbb{E}[b^T(x_0)(B^T W B)^{-1} B^T W B W \beta + b^T(x_0)(B^T W B)^{-1}(B^T W \varepsilon)]$$

$$= \mathbb{E}[b^T(x_0) W \beta] \quad (\text{as } \mathbb{E}(\varepsilon) = 0)$$

$$= b^T(x_0) W \beta$$

$$\Rightarrow \hat{y}(x_0) - \mathbb{E}(\hat{y}(x_0)) = b^T(x_0)\hat{\beta} - b^T(x_0) W \beta$$

$$= b^T(x_0)(B^T W B)^{-1} B^T W y - b^T(x_0) W \beta$$

$$= b^T(x_0)(B^T W B)^{-1} B^T W(B W \beta + \varepsilon) - b^T(x_0) W \beta$$

$$= b^T(x_0)(B^T W B)^{-1} \times B^T W \varepsilon = \hat{y}(x_0) - \mathbb{E}(\hat{y}(x_0)) \quad ①$$

Now,

$$Var(\hat{y}(x_0)) = \mathbb{E}[(\hat{y}(x_0) - \mathbb{E}(\hat{y}(x_0)))(\hat{y}(x_0) - \mathbb{E}(\hat{y}(x_0)))^T]$$

$$= \mathbb{E}\left[\left(b(x_0)^T (B^T W B)^{-1} B^T W \varepsilon\right)\left(\varepsilon^T W^T B (B^T W B)^{-1} b(x_0)\right)\right]$$

$$= \sigma^2\, b(x_0)^T (B^T W B)^{-1} B^T W W^T B (B^T W B)^{-1} b(x_0)$$

$$= \sigma^2 \|\ell(x_0)\|^2$$ (notation from P197)