Ex 6.5

We can rewrite the log-likelihood (6.19) for a binary response and only fitting the intercept term as:

$$\ell(\beta) = \sum_{i=1}^{N} K_\lambda(x_0, x_i)\left[\sum_{j=1}^{J-1} y_{ij}\,\beta_j(x_0) - \log\left(1 + \sum_{j=1}^{J-1} \exp(\beta_j(x_0))\right)\right]$$

Where y_{ij} is the binary response variable which is 1 if observation i is class j and 0 otherwise. The β_j's correspond to the intercept coefficients.

Now we can maximise $\ell(\beta)$ for a given β_j by solving:

$$\frac{\partial \ell(\beta)}{\partial \beta_j(x_0)} = \sum_{i=1}^{N} K_\lambda(x_0, x_i)\Bigg(y_{ij} - \underbrace{\frac{\exp(\beta_j(x_0))}{1 + \sum_{j=1}^{J-1} \exp(\beta_j(x_0))}}_{P_j}\Bigg) = 0$$

$$\Rightarrow \quad P_j \sum_{i=1}^{N} K_\lambda(x_0, x_i) = \sum_{i=1}^{N} K_\lambda(x_0, x_i)\, y_{ij}$$

$$\Rightarrow \quad P_j = \frac{\sum_{i=1}^{N} K_\lambda(x_0, x_i)\, y_{ij}}{\sum_{i=1}^{N} K_\lambda(x_0, x_i)}$$

and this is just the Nadaraya-Watson Kernel Smoother as required.