Ex. 6.8 Suppose that for continuous response Y and predictor X, we model the joint density of X, Y using a multivariate Gaussian kernel estimator. Note that the kernel in this case would be the product kernel $\phi_\lambda(X)\phi_\lambda(Y)$. Show that the conditional mean $E(Y|X)$ derived from this estimate is a Nadaraya–Watson estimator. Extend this result to classification by providing a suitable kernel for the estimation of the joint distribution of a continuous X and discrete Y.

① $\hat{f}_{X,Y}(x,y) = \frac{1}{N\lambda^2}\sum_{i=1}^{N}\phi_\lambda(x-x_i)\phi_\lambda(y-y_i)$

$E(Y|X) \Rightarrow \hat{f}_{Y|X}(y|x)$ 구해야함.

② $\hat{f}_{Y|X}(y|x) = \frac{\hat{f}_{X,Y}(x,y)}{\hat{f}_X(x)}$

$\hat{f}_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \frac{1}{N\lambda}\sum_{j=1}^{N}\phi_\lambda(x-x_j)$

③ $\hat{f}_{Y|X}(y|x) = \frac{\cancel{\frac{1}{N\lambda^2}}\sum_{i=1}^{N}\phi_\lambda(x-x_i)\phi_\lambda(y-y_i)}{\cancel{\frac{1}{N\lambda}}\sum_{j=1}^{N}\phi_\lambda(x-x_j)}$

$= \frac{\sum_{i=1}^{N}\phi_\lambda(x-x_i)\phi_\lambda(y-y_i)}{\lambda\sum_{j=1}^{N}\phi_\lambda(x-x_j)}$

④ $E(Y|X=x) = \int_{-\infty}^{\infty} y\cdot f_{Y|X}(y|x)\,dy = \frac{1}{\lambda}\int_{-\infty}^{\infty} y\cdot\frac{\sum_{i=1}^{N}\phi_\lambda(x-x_i)\phi_\lambda(y-y_i)}{\sum_{j=1}^{N}\phi_\lambda(x-x_j)}\,dy$

$= \frac{1}{\lambda\sum_{j=1}^{N}\phi_\lambda(x-x_j)} \times \sum_{i=1}^{N}\phi_\lambda(x-x_i) \times \int_{-\infty}^{\infty} y\,\phi_\lambda(y-y_i)\,dy$

gaussian dencity with mean 0 and sd λ

$\underbrace{\int_{-\infty}^{\infty}(y-y_i)\phi_\lambda(y-y_i)\,dy}_{0} + \underbrace{\int_{-\infty}^{\infty} y_i\,\phi_\lambda(y-y_i)\,dy}_{\lambda y_i} = \lambda y_i$

5

$$E(Y|X=x) = \frac{\sum_{i=1}^{N} \phi_\lambda(x-x_i)}{\cancel{\lambda}\sum_{j=1}^{N} \phi_\lambda(x-x_j)} \times \cancel{\lambda} y_i = \frac{\sum_{i=1}^{N} \phi_\lambda(x-x_i)\, y_i}{\sum_{j=1}^{N} \phi_\lambda(x-x_j)}$$

$K_\lambda(x_0, x_i) = \phi_\lambda(x-x_i)$

$\therefore$ fits into the definition of a NW-estimator.