✎ 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.

By the hint, $\hat{f}_{X,Y}(x_0, y_0) = \frac{1}{N\lambda^2}\Sigma\phi_\lambda(x_i - x_0)\,\phi_\lambda(y_i - y_0)$, $\hat{f}_X(x_0) = \frac{1}{N\lambda}\Sigma\phi_\lambda(x_i - x_0)$

$$\therefore \frac{\hat{f}_{X,Y}(x_0, y_0)}{\hat{f}_X(x_0)} = \frac{\Sigma\phi_\lambda(x_i - x_0)\,\phi_\lambda(y_i - y_0)}{\lambda\,\Sigma\phi_\lambda(x_i - x_0)}$$

$$E[Y|X = x_0] = \int y\,\frac{\hat{f}_{X,Y}(x_0, y)}{\lambda\,\hat{f}_X(x_0)}\,dy = \frac{1}{\lambda\Sigma\phi_\lambda(x_i - x_0)}\,\Sigma\phi_\lambda(x_i - x_0)\int y\,\phi_\lambda(y - y_i)\,dy$$

Since ϕ_λ is a gaussian kernel, $\int y\,\phi_\lambda(y - y_i)\,dy = \lambda y_i$.

$$\therefore E[Y|X = x_0] = \frac{\Sigma\phi_\lambda(x_i - x_0)\,y_i}{\Sigma\phi_\lambda(x_i - x_0)}$$