# Nguyen Huu Thanh

***Assignment 3 - Kernel Density Estimator***

## 1 Show the estimator of $P(x, y)$ by using KERNEL DENSITY ESTIMATOR

$$P(x,y) = \frac{1}{n}\sum_{i=1}^{n} K_{h_x}(x - x_i)K_{h_y}(y - y_i)$$

For Gaussian kernel:

$$K_{h_x}(x - x_i) = \frac{1}{\sqrt{2h_x^2\pi}} e^{-\frac{(x-x_i)^2}{2h_x^2}}$$

$$K_{h_y}(y - y_i) = \frac{1}{\sqrt{2h_y^2\pi}} e^{-\frac{(y-y_i)^2}{2h_y^2}}$$

## 2 Calculate $E(y|x)$

$$E(y|x) = \frac{\int_{-\infty}^{\infty} yP(x,y)dy}{\int_{-\infty}^{\infty} P(x,y)dy}$$

For the numerator:

$$\begin{aligned}\int_{-\infty}^{\infty} yP(x,y)dy &= \int_{-\infty}^{\infty} y\frac{1}{n}\sum_{i=1}^{n} K_{h_x}(x - x_i)K_{h_y}(y - y_i)dy \\ &= \frac{1}{n}\sum_{i=1}^{n}[K_{h_x}(x - x_i)\int_{-\infty}^{\infty} yK_{h_y}(y - y_i)dy]\end{aligned}$$

From Gaussian distribution properties, we have:

$$\begin{aligned}\int_{-\infty}^{\infty} yK_{h_y}(y - y_i)dy &= \int_{-\infty}^{\infty} y\frac{1}{\sqrt{2h_y^2\pi}} e^{-\frac{(y-y_i)^2}{2h_y^2}} dy \\ &= y_i\end{aligned}$$

Hence we can write that

$$\int_{-\infty}^{\infty} yP(x,y)dy = \frac{1}{n}\sum_{i=1}^{n} K_{h_x}(x - x_i)y_i$$

For the denominator:

$$\begin{aligned}\int_{-\infty}^{\infty} P(x,y)dy &= \frac{1}{n}\sum_{i=1}^{n} K_{h_x}(x - x_i)\int_{-\infty}^{\infty} K_{h_y}(y - y_i)dy \\ &= \frac{1}{n}\sum_{i=1}^{n} K_{h_x}(x - x_i)\end{aligned}$$

Therefore:

$$E(y|x) = \frac{\sum_{i=1}^{n} K_{h_x}(x - x_i)y_i}{\sum_{i=1}^{n} K_{h_x}(x - x_i)}$$

# 3 Implement estimator of $E(y|x)$

I have implemented estimator of $E(y|x)$ by using various numbers of $h_x$: $h_x = S_x\frac{k}{M}$ where $S_x$ is standard deviation of X values in *learning_data2.txt*, $M = 1000$ and value of $k$ run from 1 to $M$.
The run time of program is 3 seconds.
Best result for Root Mean Squared Error by Leave-one-out cross validation is **0.02417** when $h_x = S_x * 0.036$ ($S_x$ is is standard deviation of X values).
The following image shows the estimating values of Y given X.

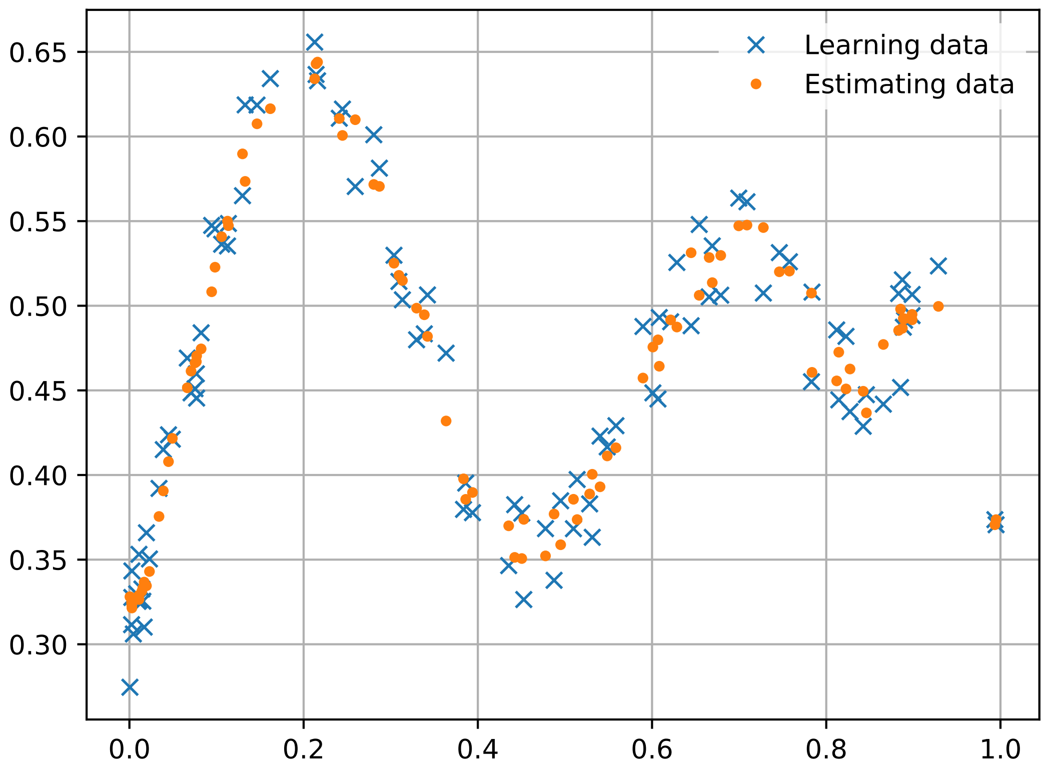


Figure 1: Compare leaning data and estimating data, **RMS = 0.02417**