

Pattern Recognition and Machine Learning: Homework 5

Qingru Hu 2020012996

March 27, 2023

Problem 1

The posterior distribution of $\omega_i, i = 1, 2$ is that:

$$P(\omega_1|(x, y)) = \frac{P(x, y|\omega_1)P(\omega_1)}{P(x, y)} = C \frac{1}{3\pi} e^{-\frac{x^2+y^2}{2}}$$
$$P(\omega_2|(x, y)) = \frac{P(x, y|\omega_2)P(\omega_2)}{P(x, y)} = C \frac{1}{6\pi} e^{-\frac{(x-2)^2+(y-2)^2}{2}}$$

Given that the probability $P(x, y) = 1/C$ is a constant, according to the minimum prediction error principle, we can have the optimal classifier:

$$f(x, y) = \omega^* = \begin{cases} \omega_1 & P(\omega_1|(x, y)) > P(\omega_2|(x, y)) \\ \omega_2 & P(\omega_1|(x, y)) \leq P(\omega_2|(x, y)) \end{cases}$$

Let $P(\omega_1|(x, y)) = P(\omega_2|(x, y))$ and we can get the decision boundary as shown in Fig.1:

$$x + y - 2 - \frac{1}{2} \ln 2 = 0$$

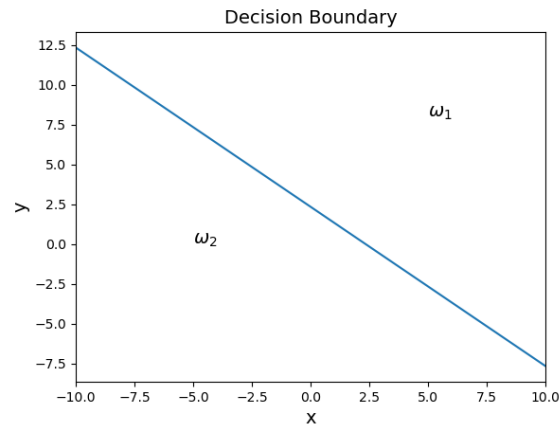


Figure 1: The decision boundary for (x, y)

Problem 2

If the final classifier outputs $f^*(x) = \omega_1$, the expected cost of $f(x) = \omega_1$ is smaller than or equal to that of $f(x) = \omega_2$:

$$\begin{aligned}\int_x \lambda_2 P(\omega_2, x) dx &\leq \int_x \lambda_1 P(\omega_1, x) dx \\ \lambda_2 P(\omega_2, x) &\leq \lambda_1 P(\omega_1, x) \\ \frac{P(\omega_1, x)}{P(\omega_2, x)} &\geq \frac{\lambda_2}{\lambda_1}\end{aligned}$$

Vice versa, if $\frac{P(\omega_1, x)}{P(\omega_2, x)} \geq \frac{\lambda_2}{\lambda_1}$, then:

$$\begin{aligned}\frac{P(\omega_1, x)}{P(\omega_2, x)} &\geq \frac{\lambda_2}{\lambda_1} \\ \int_x \lambda_2 P(\omega_2, x) dx &\leq \int_x \lambda_1 P(\omega_1, x) dx\end{aligned}$$

the expected cost of $f(x) = \omega_1$ is smaller than or equal to that of $f(x) = \omega_2$, so the final classifier outputs $f^*(x) = \omega_1$.

Specially, when $\lambda_1 = \lambda_2$, this corresponds to our original max-posterior-solution.

Problem 3

The hypothesis class is:

$$\mathcal{H} = \{h(x) = (-1)^{k+\mathbb{I}[a \leq x \leq b]} : k \in \{0, 1\}, a, b \in \mathbb{R}, a \leq b\}$$

Since k and $\mathbb{I}[a \leq x \leq b]$ can only have values of 0 or 1, the class labels are 1 and -1 .

Suppose we have three points $x_1 < x_2 < x_3$, there are 8 possible labelings and all the possible labelings can be classified by a specific function (a specific set of (k, a, b)) in \mathcal{H} :

x_1	x_2	x_3	k	a, b satisfy
1	1	1	1	$a < x_1 < x_2 < x_3 < b$
-1	1	1	1	$x_1 < a < x_2 < x_3 < b$
1	-1	1	0	$x_1 < a < x_2 < b < x_3$
1	1	-1	1	$a < x_1 < x_2 < b < x_3$
-1	-1	1	1	$x_1 < x_2 < a < x_3 < b$
-1	1	-1	1	$x_1 < a < x_2 < b < x_3$
1	-1	-1	1	$a < x_1 < b < x_2 < x_3$
-1	-1	-1	0	$a < x_1 < x_2 < x_3 < b$

However, for four points $x_1 < x_2 < x_3 < x_4$, the label of $(1, -1, 1, -1)$ can not be classified by any function in \mathcal{H} . Therefore, the VC dimension of \mathcal{H} is 3.

Problem 4

(1)

I use Parzen window of width $h = 1$ to estimate the density function of $P(x, y)$. The the density function is shown in Fig.2.

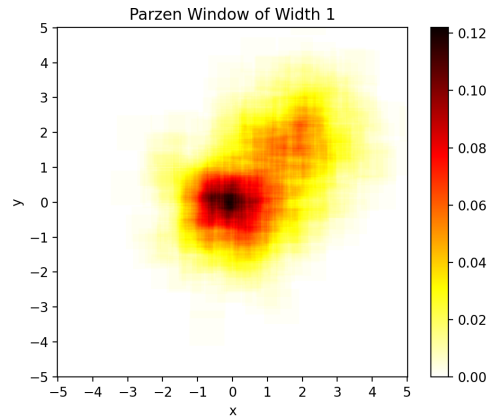


Figure 2: The heatmap of desity function from parzen window of width $h = 2$

(2)

Change the window width in $[0.3, 0.5, 1]$ and kernel function in $['Parzen', 'Gaussian', 'Exp']$, plot the heatmap, and calculate the mean density error. All the relative information are labelled in figures. According the mean density error, the kernel function 'Parzen' with window width of 1 is the best composition.

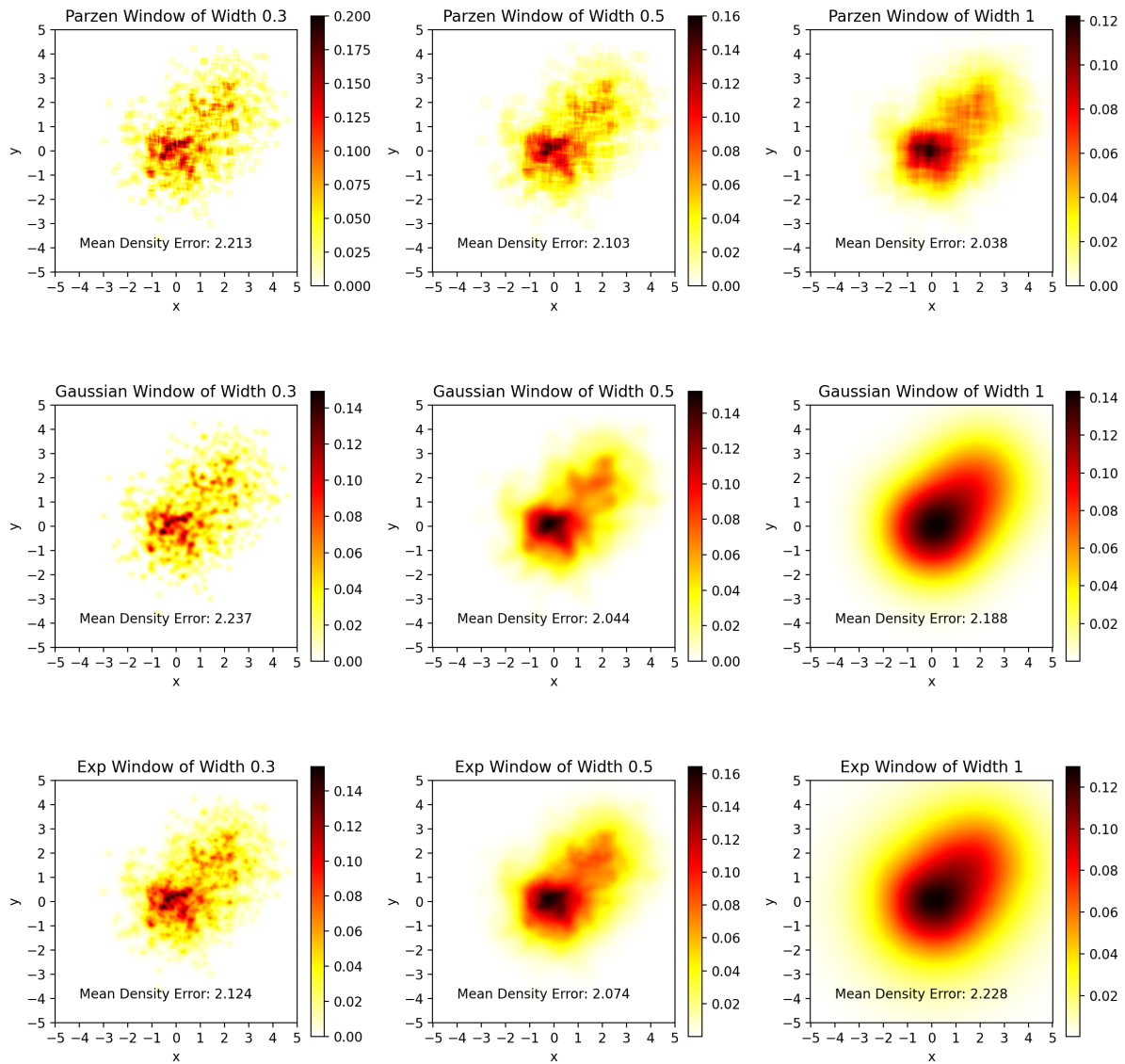


Figure 3: The heatmap of desity function of different kernels and window widths