

Homework 3

Problem 1:

a) **Solution:**

The histogram of the 1000 generated random variables is shown in Fig 1.

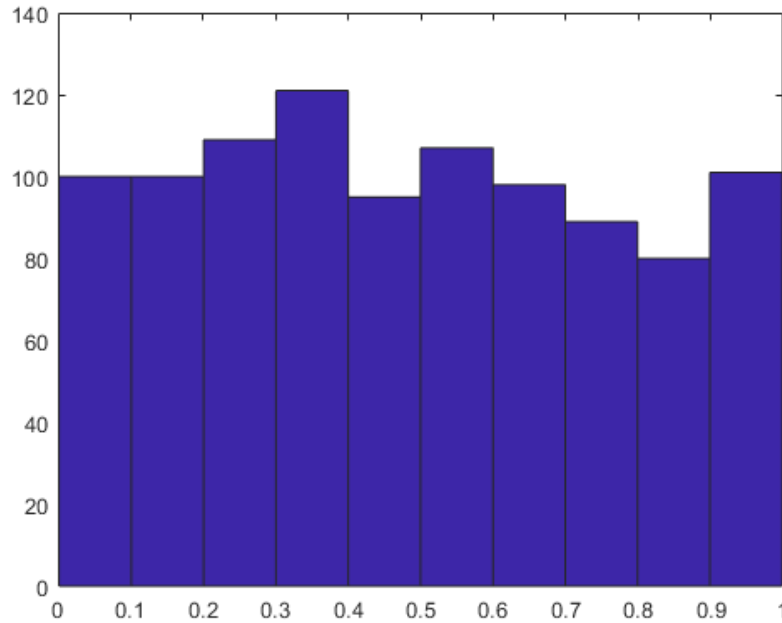


Figure 1: Histogram of 1000 random variables

According to the Law of Large Numbers, $\mathbb{E}[K_h(X - x)] \approx \frac{K_h(X-x_1)+K_h(X-x_2)+\dots+K_h(X-x_n)}{n} \approx f(x)$.

When we approximate the pdf using the Gaussian kernel,

$$K_h(X) = \frac{e^{-\frac{1}{2h}x^2}}{\sqrt{2\pi h}}.$$

The results are shown in Fig 2.

We can see from Fig 2 that when the h is large, the model underfit the pdf, which is not close enough. As the h is decreasing, the support of the random variable becomes closer to $[0, 1]$, and the approximation of pdf becomes more zigzag around 1, starts to overfit.

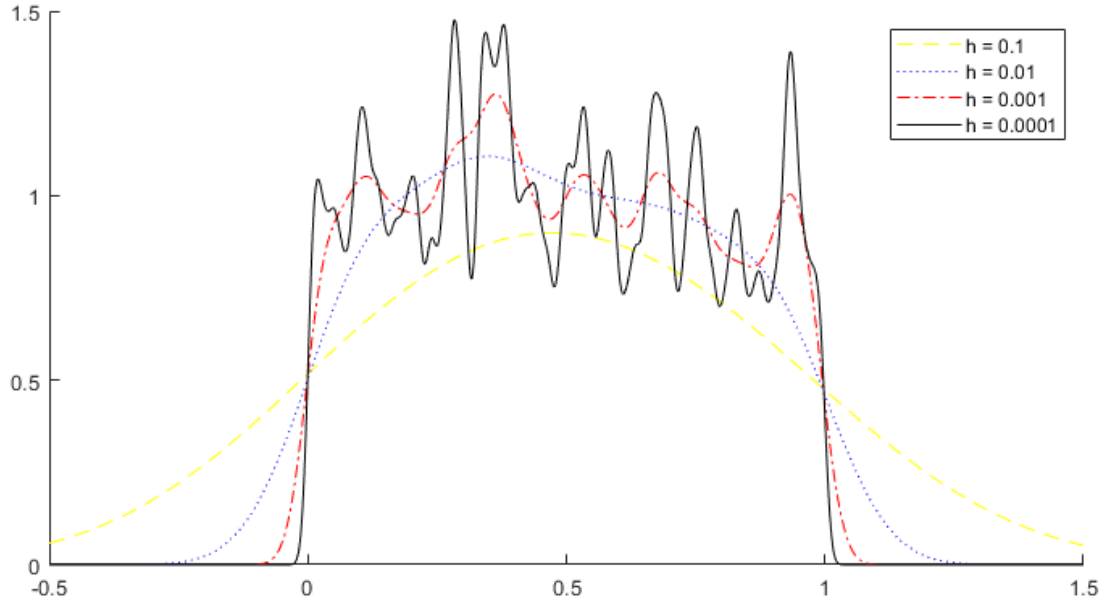


Figure 2: Approximation using Gaussian kernel under different h s.

b) **Solution:**

When we approximate the pdf using the Laplacian kernel,

$$K_h(X) = \frac{e^{-\frac{1}{2h}x^2}}{\sqrt{2\pi h}}.$$

The results are shown in Fig 3.

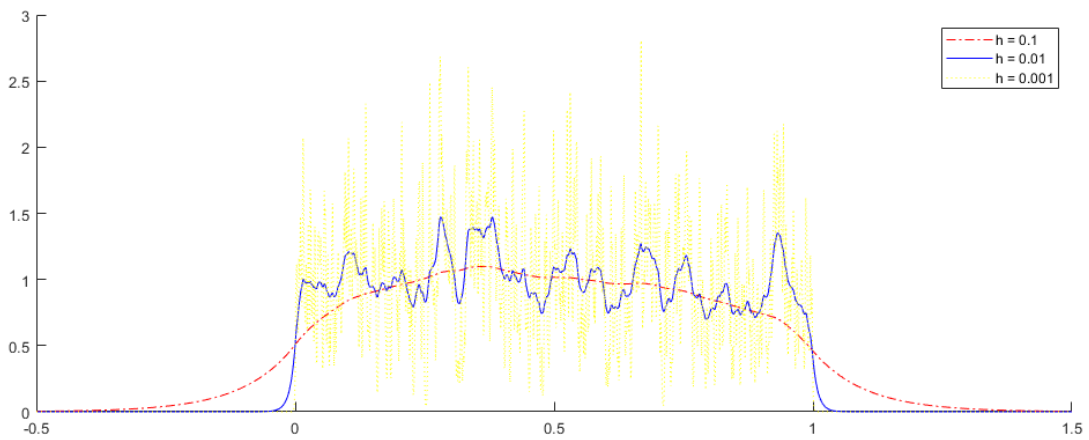


Figure 3: Approximation using Laplacian kernel under different h s.

Problem 2:

a) **Solution:**

By solving this minimization problem:

$$\min_{\phi} \left\{ \sum_{X_i \in \text{stars}} (1 - \phi(X_i))^2 + \sum_{X_j \in \text{circles}} (1 + \phi(X_j))^2 + \lambda \|\phi(X)\|^2 \right\},$$

the idea is to try to minimize the loss.

Define V , the space of functions that generated by the Gaussian kernel

$$K(X, Y) = e^{-\frac{1}{h}\|X-Y\|^2} = e^{-\frac{1}{h}\{(x_1-y_1)^2+(x_2-y_2)^2\}}.$$

Define $\phi(X) : z \rightarrow K(x, z)$, $z \in \mathbb{R}^2$, $K(x, z) \in V$.

In the vector space, $g(x) = \sum_{i=1}^m \alpha_i K(x, z_i)$, $h(x) = \sum_{i=1}^{m'} \beta_i K(x, z'_i)$.

Define the inner product:

$$\langle g(x), h(x) \rangle = \sum_{i=1}^m \sum_{j=1}^{m'} \alpha_i \beta_j K(z_i, z'_j).$$

Then we are trying to solving

$$\min_{g(x) \in V} \left\{ \sum_{X_i \in \text{stars}} (1 - g(X_i))^2 + \sum_{X_j \in \text{circles}} (-1 - g(X_j))^2 + \lambda \|g(X)\|^2 \right\}, \text{ where } g(X) = \langle g(X), g(X) \rangle.$$

According to the theorem, the best choice would be $m = N$, $z_i = x_i$.

Thus, $g(x) = \sum_{i=1}^m \alpha_i K(x, x_i) + \sum_{i=1}^{m'} \beta_i K(x, x'_i)$, where x_i s are data of stars, x'_i s are data of circles.

If we concatenate the input data, the inner product would be:

$$\langle g(x), g(x) \rangle = \sum_{i=1}^m \sum_{j=1}^{m'} \alpha_i \alpha_j K(x_i, x'_j) = \alpha^T K \alpha,$$

and $g(x)$ would be: $g(x) = K\alpha$. K is positive definite and symmetric.

Denote the original class of the input data by b , then the minimization we are trying to solve becomes:

$$\min(\|b - K\alpha\|^2 + \lambda \alpha^T K \alpha).$$

We let the partial derivative equal to 0, we get $\alpha = (\lambda I + K)^{-1}b$.

We hereby successfully turned the question.

b) **Solution:**

Now we can just put the new point X_{new} into $g(x) = \alpha_1 K(x, x_i) + \cdots + \alpha_m K(x, x_m) + \alpha_1 K(x, x'_i) + \cdots + \alpha_{m'} K(x, x'_{m'})$.

If $g(X_{new}) > 0$, we call it a star. Otherwise we call it a circle.

c) **Solution:**

The boundary for the two classes would be lying on where $g(x) = 0$. The results are shown in Fig 4 and Fig 5.

d) **Solution:**

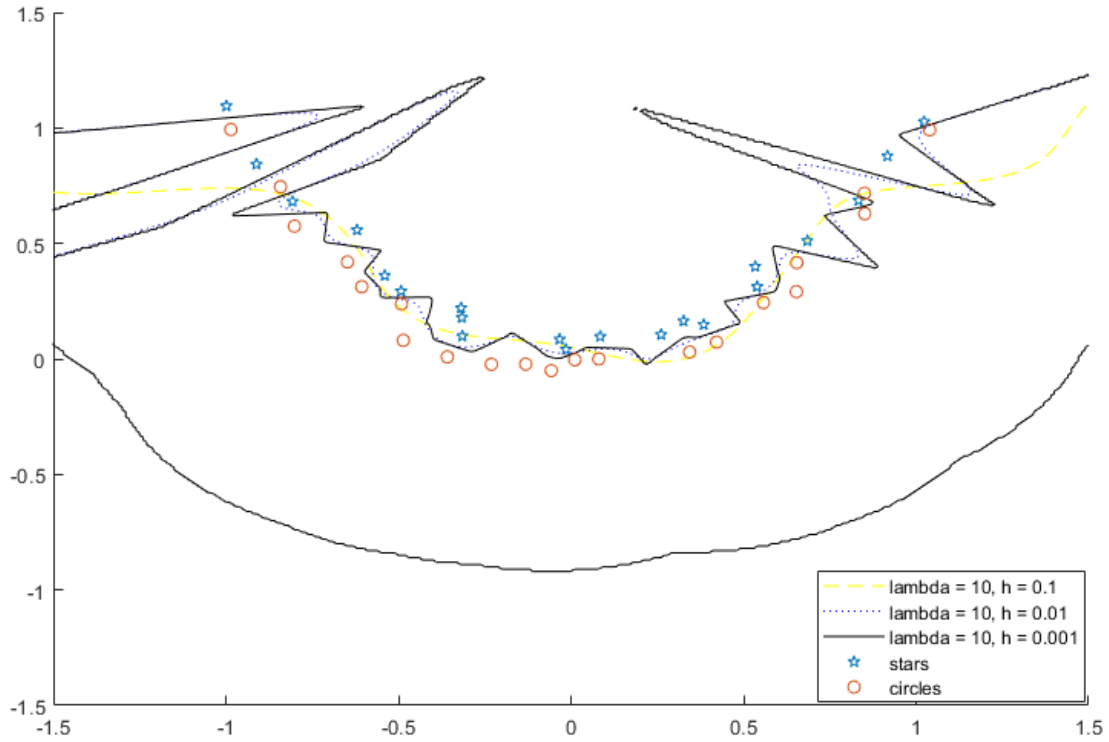


Figure 4: The boundary under different values of h .

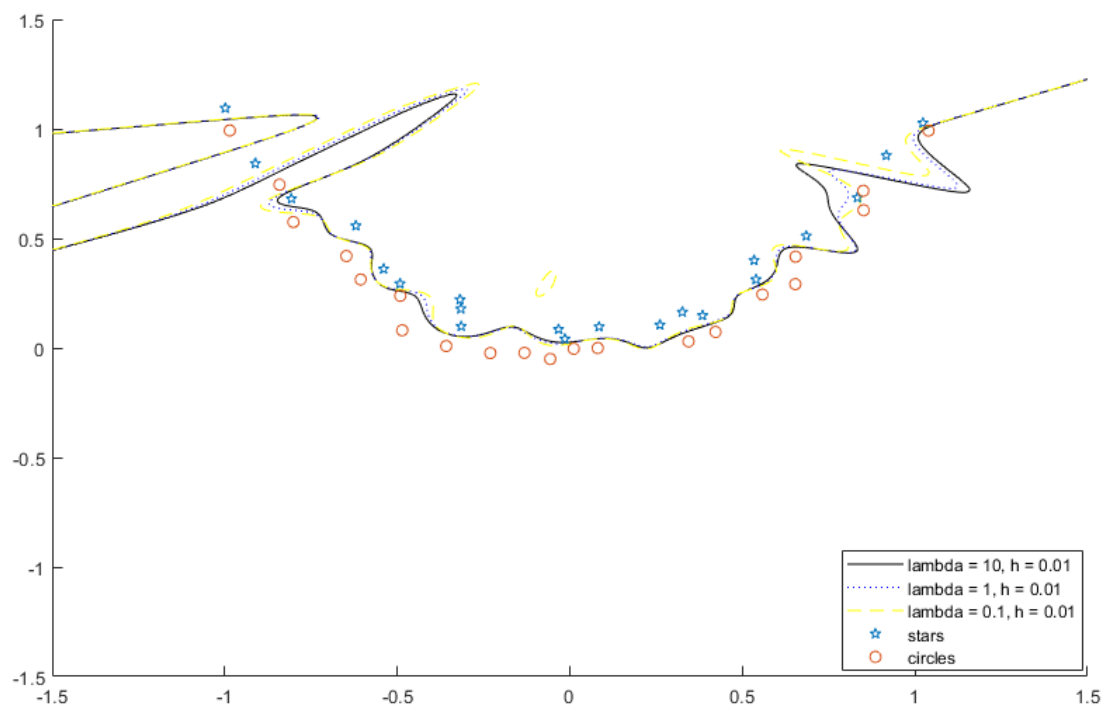


Figure 5: The boundary under different values of λ .

e) **Solution:**