Shengjie Li
RUID: 188008047

Homework 3

Problem 1:

a) **Solution:**
The histogram of the 1000 generated random variables is shown in Fig 1.

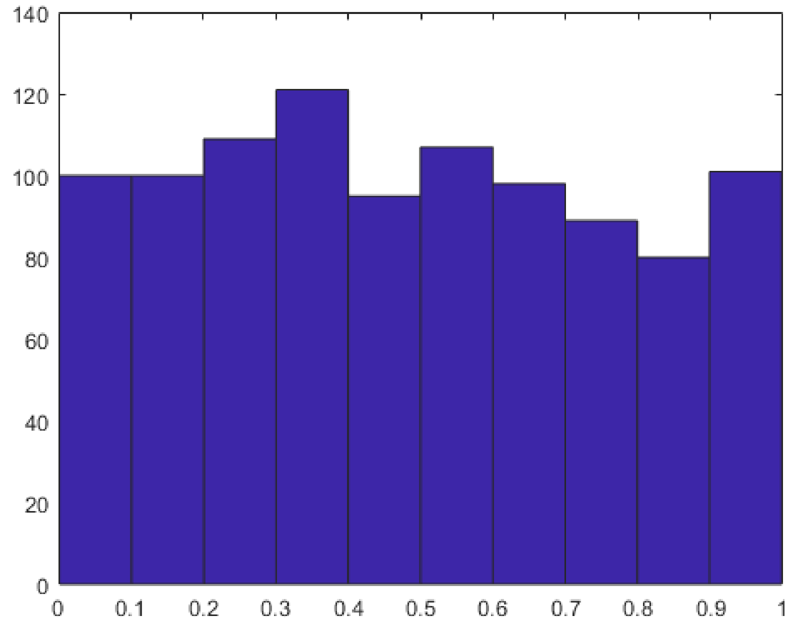

Figure 1: Histogram of 1000 random variables

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

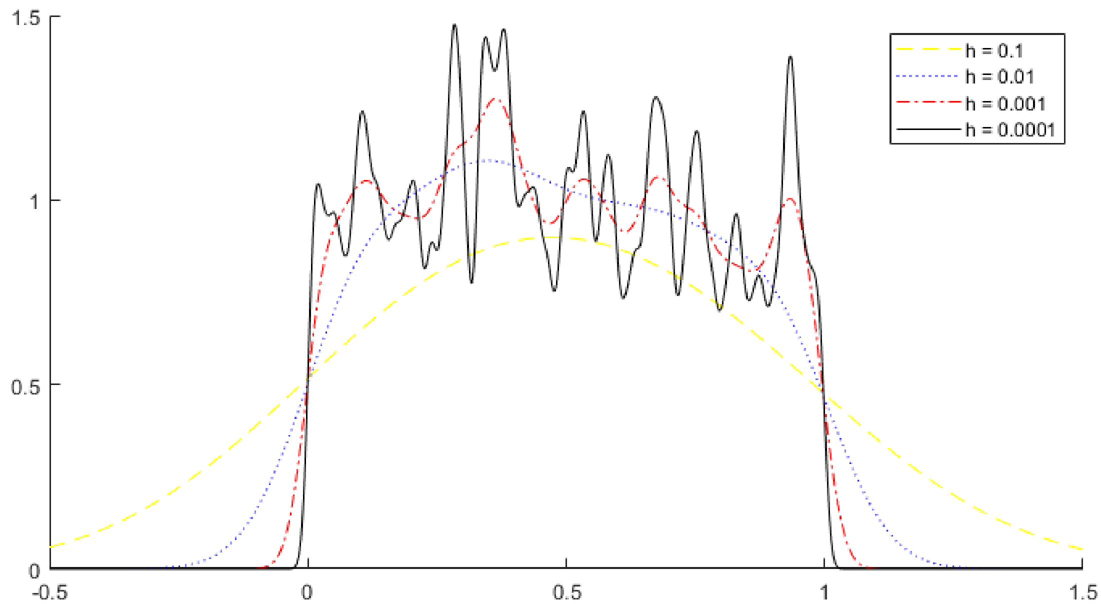

Figure 2: Approximation using Gaussian kernel under different hs.

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.

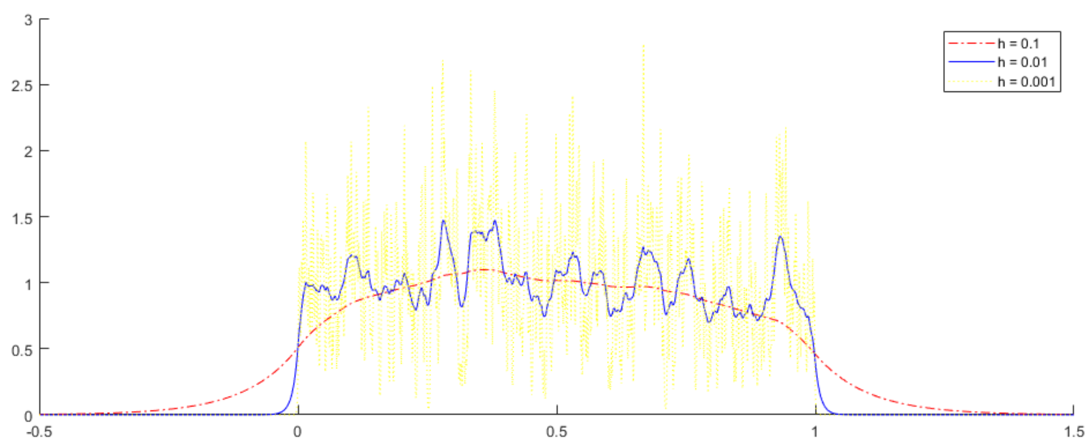

Figure 3: Approximation using Laplacian kernel under different hs.

Problem 2:

a) **Solution:**

By solving this minimization problem:

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

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

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

Then we are trying to solving

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

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_is are data of stars, x'_is are data of circles.
If we concatenate the input data, the inner product would be:

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

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 transformed 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:**

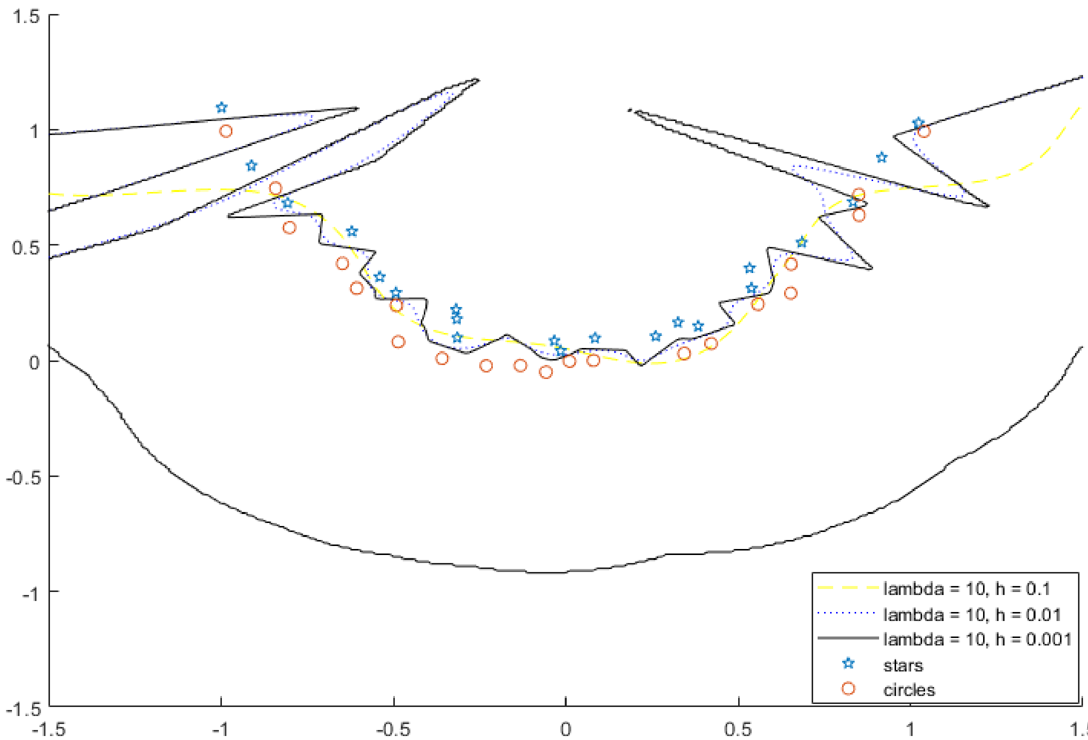

Figure 4: The boundary under different values of h.

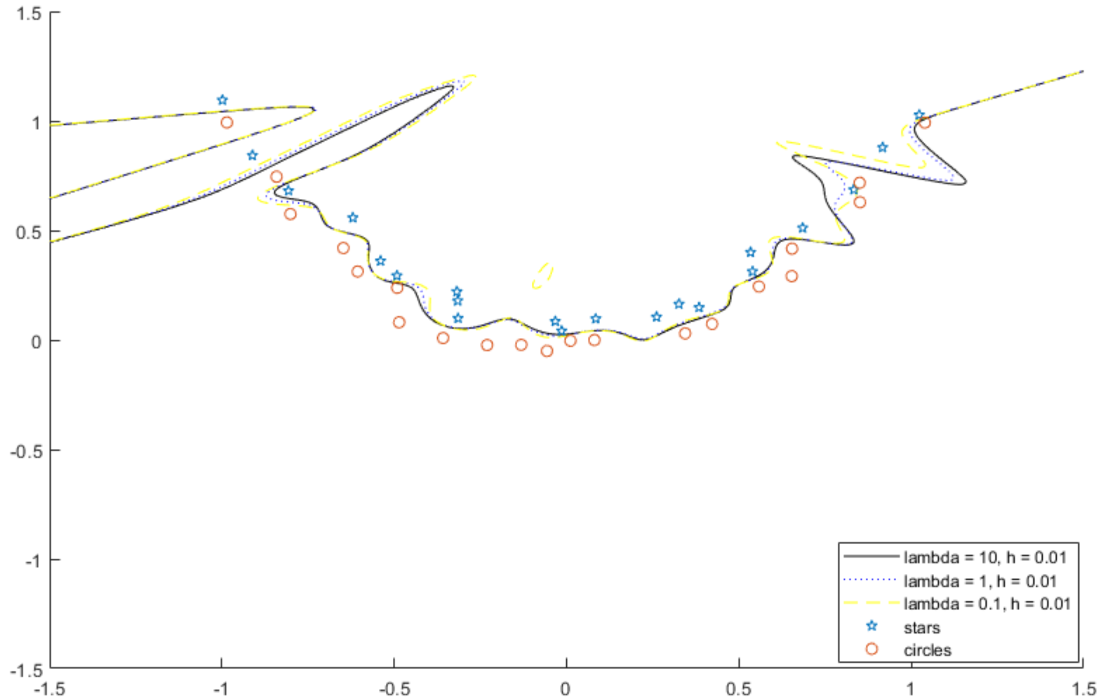

Figure 5: The boundary under different values of λ.

e) **Solution:**

When using the simpler kernel function

$$K(X, Y) = (1 + x_1y_1 + x_2y_2)^2,$$

we can still use the theorem.

We can actually use the same $g(x)$, so that if given a new point X_{new}, if $g(X_{new}) > 0$, we call it a star. Otherwise we call it a circle.

The results for different values of λ are shown in Fig 6.

Some examples of functions $\phi(X)$ are:

- $$g(x) = C_0 + x_1y_1 + x_2y_2 + \cdots + x^2y^2$$
- $$g(x) = C_0 + C_1x_1 + C_1x_2 + C_2x_1x_2 + d_1x_1^2 + d_2x_2^2$$

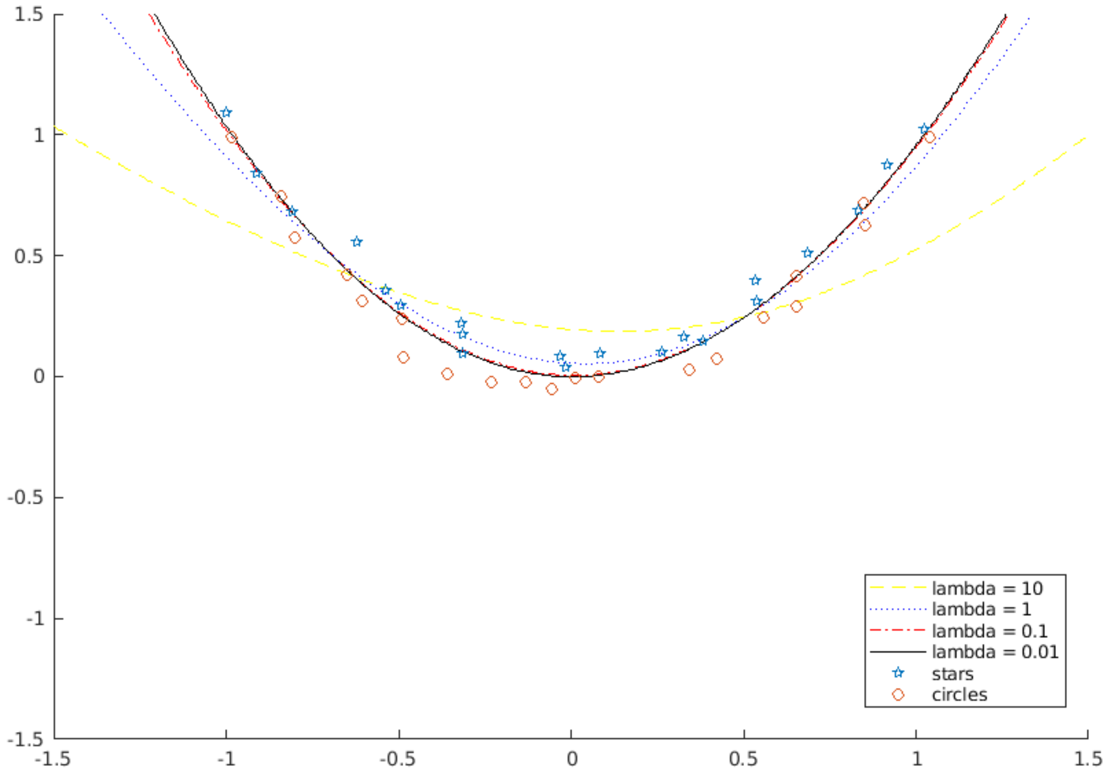

Figure 6: The boundary under different values of λ.