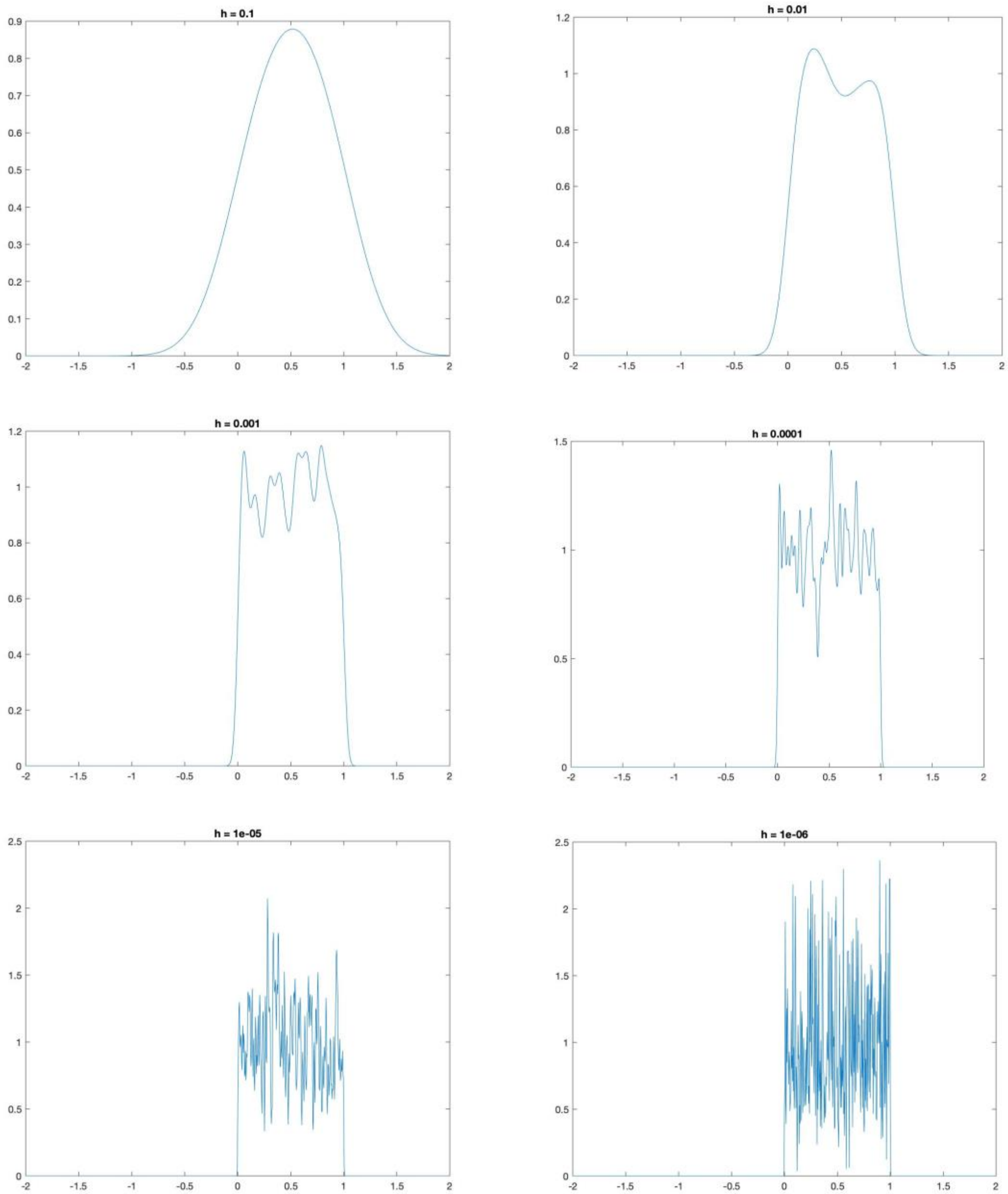
596 Assignment4 / Yujia Fan

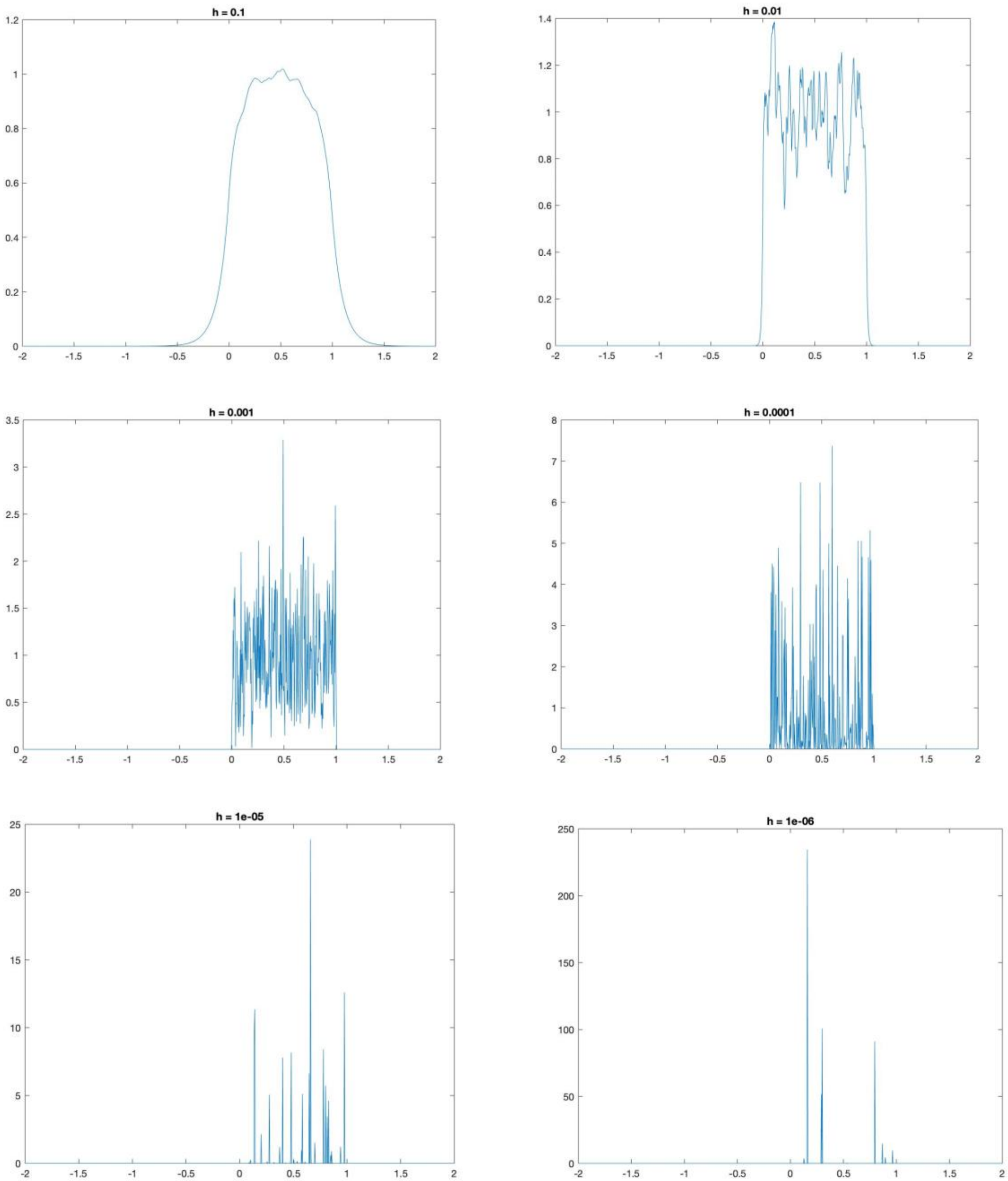
Problem 1

a)



As h becomes smaller, the approximate lines become more dramatic. When h equals to 0.1, the line becomes very smoothly.

b)



As h becomes smaller, the approximate lines become more dramatic. When h equals to 0.1, the line becomes very smoothly.

Problem 1

1a.

```
r = rand(1000, 1);
x = [-2:0.004:1.996];
h = 0.1;
y = [];
for i = 1:1000
    temp = 0;
    for j = 1:1000
        temp = temp + 1/sqrt(2*pi*h)*exp(-1 / (2 * h) * ((x(i) - r(j)) ^2));
    end
    temp = temp / 1000;
    y(i) = temp;
end
plot(x, y);
title(['h = ',num2str(h)]);
```

1b.

```
r = rand(1000, 1);
r = transpose(r);
x = [-2:0.004:1.996];
h = 0.001;
y = [];
for i = 1:1000
    temp = 0;
    for j = 1:1000
        temp = temp + 1 / (2 * h) * exp(-1 / h * abs(x(i) - r(j)));
    end
    temp = temp / 1000;
    y(i) = temp;
end
plot(x, y);
title(['h = ',num2str(h)]);
```

Problem 2

(a) We want to find optimum $\phi(X)$

$$\min_{\phi \in V} \left\{ \sum_{X_i \in stars} (1-\phi(X_i))^2 + \sum_{X_j \in circles} (1+\phi(X_j))^2 + \lambda \|\phi(x)\|^2 \right\} \quad (*)$$

$$\phi = \sum_{i=1}^{n} a_i \hat{\phi}(X_i) + U, \quad \langle U, \hat{\phi}(X_i) \rangle = 0$$

$$\hat{\phi}(X_j) = \langle \sum_{i=1}^{n} a_i \hat{\phi}(X_i) + U, \hat{\phi}(X_j) \rangle = \sum_{i=1}^{n} a_i \langle \hat{\phi}(X_i), \hat{\phi}(X_j) \rangle$$

so we can convert (*) to :

$$\min_{a_1, \cdots, a_n} \left\{ \sum_{X_i \in stars} (1 - \sum_{t=1}^{n} a_i k(X_i, X_t))^2 + \sum_{X_j \in circles} (1 + \sum_{t=1}^{n} a_j k(X_j, X_t))^2 + \lambda \|\hat{\phi}(x)\|^2 \right\}$$

(b)
$$\begin{aligned}
\|\phi(x)\|^2 &= \|\phi(x) - \hat{\phi}(x) + \hat{\phi}(x)\|^2 \\
&= \|\phi(x) - \hat{\phi}(x)\|^2 + \|\hat{\phi}(x)\|^2 + 2 \cdot \underbrace{\|\phi(x) - \hat{\phi}(x)\| \cdot \|\hat{\phi}(x)\|}_{=0} \\
&= \|\phi(x) - \hat{\phi}(x)\|^2 + \|\hat{\phi}(x)\|^2 \\
&\geq \|\hat{\phi}(x)\|^2
\end{aligned}$$

Yes. If we want to minimize $\|\phi(x)\|^2$, that means we need to minimize $\|\hat{\phi}(x)\|^2$.

(c) First, we construct a matrix $X = [x_1^s, \cdots x_n^s, x_1^c, \cdots, x_n^c]$ and x_i is the ith column of X.
Then construct a kernel matrix K which the entry of (i,j) is $k(x_i, x_j)$.
Obviously, the kernel matrix K is symmetric.
For the coeffients α_i, β_i, we construct a vector $A = [\alpha_1, \cdots, \alpha_n, \beta_1, \cdots, \beta_n]$.
And we can rewrite

$$\|\hat{\phi}(x)\| = <\hat{\phi}(x), \hat{\phi}(x)> = AKA^T$$

Assumed that K_i is the ith row of kernel matrix K, we could transform the minimum problem into matrix form as:

$$\min_{\hat{\phi}} \left\{ \sum_{i=1}^{n} (1 - AK_i^T)^2 + \sum_{i=n+1}^{2n} (1 + AK_j^T)^2 + \lambda AKA^T \right\}$$

Take derivative with respect to A and set it to zero vector:

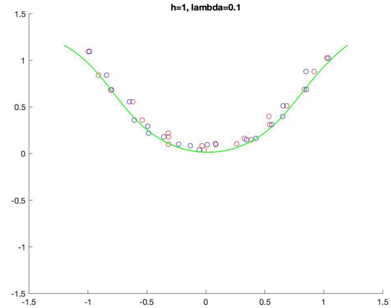
$$\sum_{i=1}^{n} -2K_i^T(1 - AK_i^T) + \sum_{j=n+1}^{2n} 2K_j^T(1 + AK_j^T) + 2\lambda KA^T = 0$$

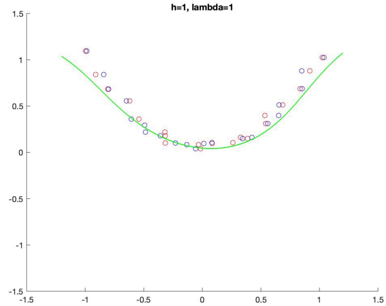
so we have $\left(\sum_{i=1}^{n} K_i^T K_i + \sum_{j=n+1}^{2n} K_j^T K_j + \lambda K \right) A^T = \sum_{i=1}^{n} K_i^T - \sum_{j=n+1}^{2n} K_j^T$

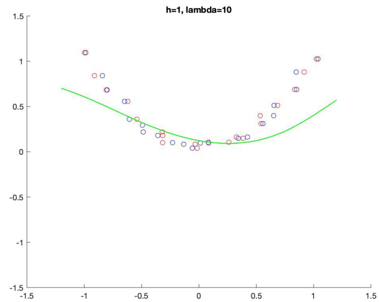
and $A = [\alpha_1, \cdots, \alpha_n, \beta_1, \cdots, \beta_n]$.
In this way, we can find all optimal α_i and β_i.

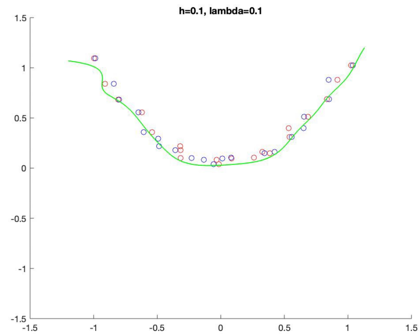
(d) After we identify $\hat{\phi}(x)$, if given a new point X_{new}, we just need to calculate $\hat{\phi}(X_{new})$. If $\hat{\phi}(X_{new})$ is positive, it can be considered as a star; otherwise, it is a circle.

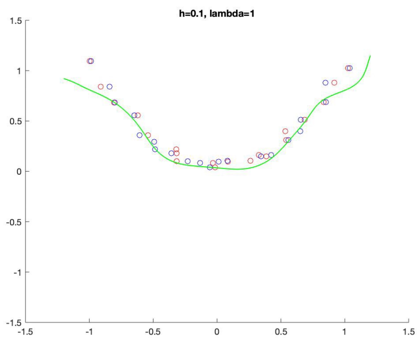
(e)

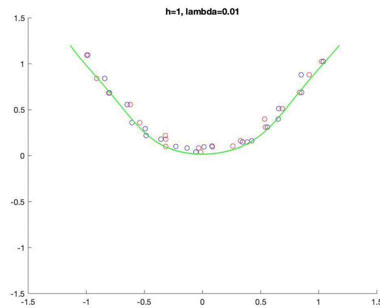

h=1, lambda=0.1

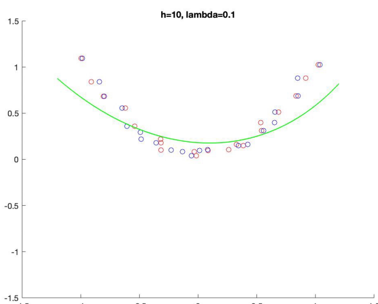

h=1, lambda=1

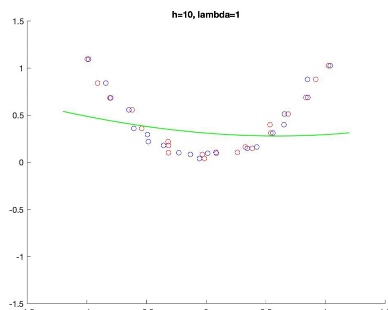

h=1, lambda=10


h=0.1, lambda=0.1


h=0.1, lambda=1


h=1, lambda=0.01


h=10, lambda=0.1


h=10, lambda=1

Problem 2

```
lamda = 10;
h = 1;
format long
S = stars;
C = circles;
scatter(S(:,1), S(:,2),'r')
hold on
scatter(C(:,1), S(:,2),'b')
hold on
T = [S;C];
K = zeros(42,42);
for i=1:42
    for j=1:42
        K(i,j) = exp(-((T( i , 1 )-T(j,1))^2+(T(i,2)-T(j,2))^2)/h);
    end
end
b=zeros(42,1);
A=zeros(42,42);
for i=1:21
    b = b+K(:,i);
    A = A+K(:,i)*K(i,:);
end
for i=22:42
    b=b-K(:,i);
    A=A+K(:,i)*K(i,:);
end
A=A+lamda*K;
X=A\b;
x=linspace(-1,1,100);
y=zeros(100);
syms x y
F=0;
for i=1:42
    F = F+X(i)*exp(-((x-T(i,1))^2 + (y-T(i,2))^2)/h);
end
fp = fimplicit(F,[-1.2,1.2]);
fp.LineWidth=1;
fp.Color='g';
title('h=1, lambda=10')
```

Problem 3

(a) $\frac{\partial E[y^2 - 2y\theta^T x + (\theta^T x)^2]}{\partial \theta} = 0 \Rightarrow \frac{E[\partial y \theta^T x]}{\partial \theta} - \frac{E[\partial (\theta^T x)^2]}{\partial \theta} = 0 \Rightarrow E[x(y - \theta^T x)] = 0$

so the optimal θ_{opt} satisfies: $E(y) = \theta_{opt}^T E[x]$

And $y = \theta_*^T X + w$ and $E(w) = 0$

so $E(y) = \theta_*^T E(x)$

The optimal θ is equal to θ_*.

(b) we want to $\min_\theta E[(\theta^T X - y)^2]$

Let $J(\theta) = (\theta^T x - y)^2$

Consider the gradient descent algorithm: $\theta_{j+1} := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$

And $\frac{\partial}{\partial \theta_j} J(\theta) = \frac{\partial}{\partial \theta_j}(h_\theta(x) - y)^2$

$= 2 \cdot (h_\theta(x) - y) \cdot \frac{\partial}{\partial \theta_j}(h_\theta(x) - y)$

$= 2 \cdot (h_\theta(x) - y) \cdot \frac{\partial}{\partial \theta_j}\left(\sum_{i=0}^{n} \theta_i X_i - y\right)$

$= 2(\theta^T x - y) \cdot X_j$

$= -2(y - \theta^T x) \cdot x_j$

For the Least Mean Squares algorithm:

$y_t = \theta_t^T X_t$

$e_t = y_t - \theta_t^T x_t$

$G_t = -\nabla_\theta(e_t^2) = 2\alpha e_t X_t$

so $\theta_{t+1} = \theta_t + 2\alpha e_t X_t$

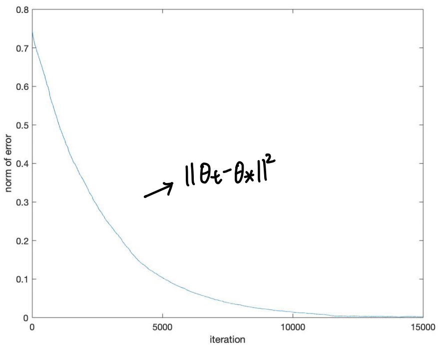
(c) Plot the squared norm $\|\theta_t - \theta_*\|^2$

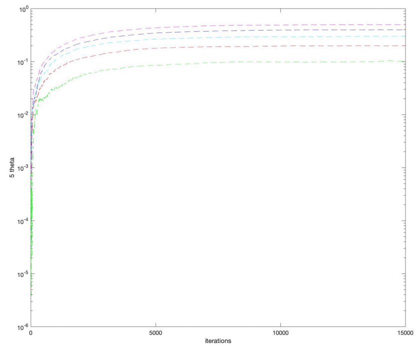
As the iterations increase, the $\|\theta_t - \theta_*\|^2$ will converge to 0.

We set $\theta^* = (0.2, 0.4, 0.5, 0.1, 0.3)$ and $\theta_0 = (0, 0, 0, 0, 0)$, study rate $\alpha = 2e\text{-}4$.

As the iterations increase, θ_s converges to θ^*.

All θs have almost the same converging trend.



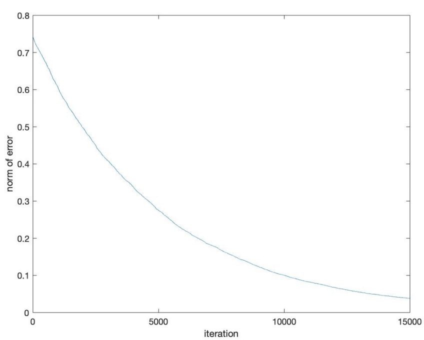


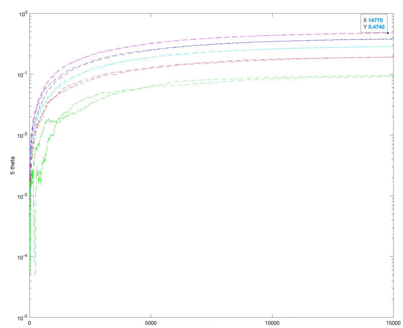
(d) If we choose study rate half of the previous experiment,

And the convergence rate is lower than the previous experiment.

The steady state error is larger than the previous one.

So the performance is worse if we halve the convergence rate.





Problem 3

```
num_samples=15000;
x=randn(5,num_samples);
theta_true=[0.2,0.4,0.5,0.1,0.3];
w=normrnd(0,0.1,1,num_samples);
y=theta_true*x+w;
diff=zeros(1,num_samples);
x_axis=(1:num_samples);
theta1=zeros(1,num_samples);
theta2=zeros(1,num_samples);
theta3=zeros(1,num_samples);
theta4=zeros(1,num_samples);
theta5=zeros(1,num_samples);

epochs=15000;
eta=2e-4/2;
theta=[0,0,0,0,0];

for i=1:epochs
    diff(i)=norm(theta_true-theta);
    theta1(i)=theta(1);
    theta2(i)=theta(2);
    theta3(i)=theta(3);
    theta4(i)=theta(4);
    theta5(i)=theta(5);
    ym=theta*x(:,i);
    e=y(i)-ym;
    theta=theta + 2*eta*x(:,i)'.*e;
    disp(theta);
end
```