机器学习课程实验五

2022 年 10 月 6 日 苏博南 202000460020

1 Regularize Linear Regression

考虑解决 overfitting 的问题，将 parameter 引入 Cost Function：

$$J(\theta) = \frac{1}{2m}[\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 + \lambda\sum_{j=1}^{n}\theta_j^2] \tag{1}$$

通过 $\nabla J(\theta) = 0$ 可以解得：

$$\theta^* = (X^TX + \lambda\begin{pmatrix}0 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1\end{pmatrix})^{-1}X^Ty \tag{2}$$

分别取 $\lambda = 0, \lambda = 1, \lambda = 10$ 可以做的如图 1 所示预测结果：

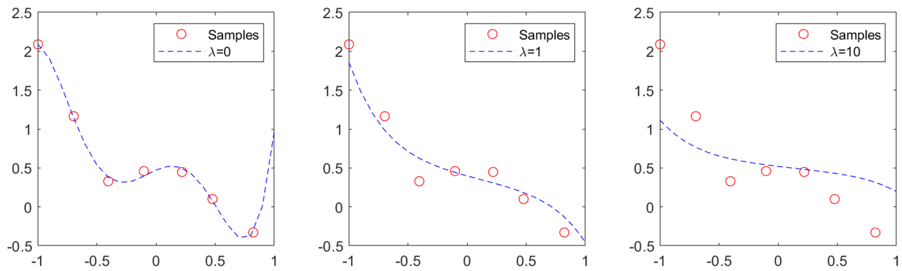

图 1: 拟合结果

代码：

```
x = load('ex5Data/ex5Linx.dat');
y = load('ex5Data/ex5Liny.dat');


[m, n] = size(x);
X = [ones(m, 1), x, x.^2, x.^3, x.^4, x.^5];
[m, n] = size(X);
reg = diag([0, ones(1, n - 1)]);
xs = [-1 : 0.1 : 1]';

subplot(1, 3, 1);
plot(x, y, 'rO');
hold on;
lambda = 0;
theta = inv(X' * X + lambda * reg) * X' * y;
ys = [ones(21, 1), xs, xs.^2, xs.^3, xs.^4, xs.^5] * theta;
plot(xs, ys, 'b--');
legend('Samples', '\lambda=0');
```

```

subplot(1, 3, 2);
plot(x, y, 'rO');
hold on;
lambda = 1;
theta = inv(X' * X + lambda * reg) * X' * y;
ys = [ones(21, 1), xs, xs.^2, xs.^3, xs.^4, xs.^5] * theta;
plot(xs, ys, 'b--');
legend('Samples', '\lambda=1');

subplot(1, 3, 3);
plot(x, y, 'rO');
hold on;
lambda = 10;
theta = inv(X' * X + lambda * reg) * X' * y;
ys = [ones(21, 1), xs, xs.^2, xs.^3, xs.^4, xs.^5] * theta;
plot(xs, ys, 'b--');
legend('Samples', '\lambda=10');
```

2 Regularize Logistic Regression

在预处理完数据后，类似 Linear Regression，可以将 parameters 引入 Cost Function：

$$J(\theta) = -\frac{1}{m}[y^{(i)}ln(h_\theta(x^{(i)})) + (1-y^{(i)})ln(1-h_\theta(x^{(i)}))] + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2 \tag{3}$$

利用牛顿法迭代：

$$\theta^{(t+1)} = \theta^{(t)} - H^{-1}\nabla_\theta J \tag{4}$$

其中，

$$\begin{gathered}\nabla_\theta J = \begin{pmatrix} \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_0^{(i)} \\ \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_1^{(i)} + \frac{\lambda}{m}\theta_1 \\ \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_2^{(i)} + \frac{\lambda}{m}\theta_2 \\ \ldots\ldots \\ \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_n^{(i)} + \frac{\lambda}{m}\theta_n \end{pmatrix} \\ H = \frac{1}{m}[\sum_{i=1}^{m} h_\theta(x^{(i)})(1-h_\theta(x^{(i)}))x^{(i)}(x^{(i)})^T] + \frac{\lambda}{m}\begin{pmatrix} 0 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}\end{gathered} \tag{5}$$

代码如下：

```
x = load('ex5Data/ex5Logx.dat');
y = load('ex5Data/ex5Logy.dat');

pos = find(y);
```

```
neg = find(y == 0);


X = map_feature(x(:, 1), x(:, 2));
[m, n] = size(X);
plot(x(pos, 1), x(pos, 2), '+');
hold on;
plot(x(neg, 1), x(neg, 2), 'o');
% lambda = 1;
% lambda = 0;
lambda = 1;
theta = zeros(n, 1);
for it = 1 : 10
    grad = X' * (sigmond(X * theta) - y) / m;
    reg = lambda / m .* theta;
    reg(1) = 0;
    grad = grad + reg;
    H = X' * diag(sigmond(X * theta) .* sigmond(-X * theta), 0) * X / m + lambda / m .* diag([0, ones(1, n - 1)]);
    theta = theta - H \ grad;
end

u = linspace(-1, 1.5, 200);
v = linspace(-1, 1.5, 200);
z = zeros(length(u), length(v));

for i = 1 : length(u)
    for j = 1 : length(v)
        z(j, i) = map_feature(u(i), v(j)) * theta;
    end
end
contour(u, v, z, [0, 0], 'LineWidth', 2);
legend('y=1', 'y=0', 'decision bound');
```

然后作出 $P(y=1 \mid x;\theta) = 0.5$ 的图像，即 $\theta^T x = 0$:

可以看到，当 λ 越大，分界线越平滑和规则，更接近一个规则图形，不规则的边界减少。

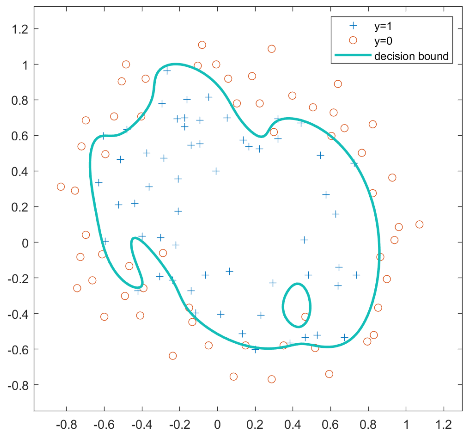

图 2: $\lambda = 0$

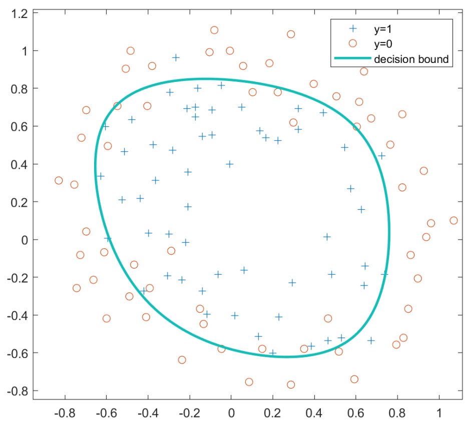

图 3: $\lambda = 1$

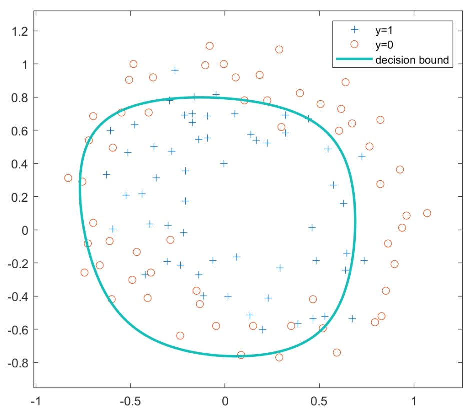

图 4: $\lambda = 10$